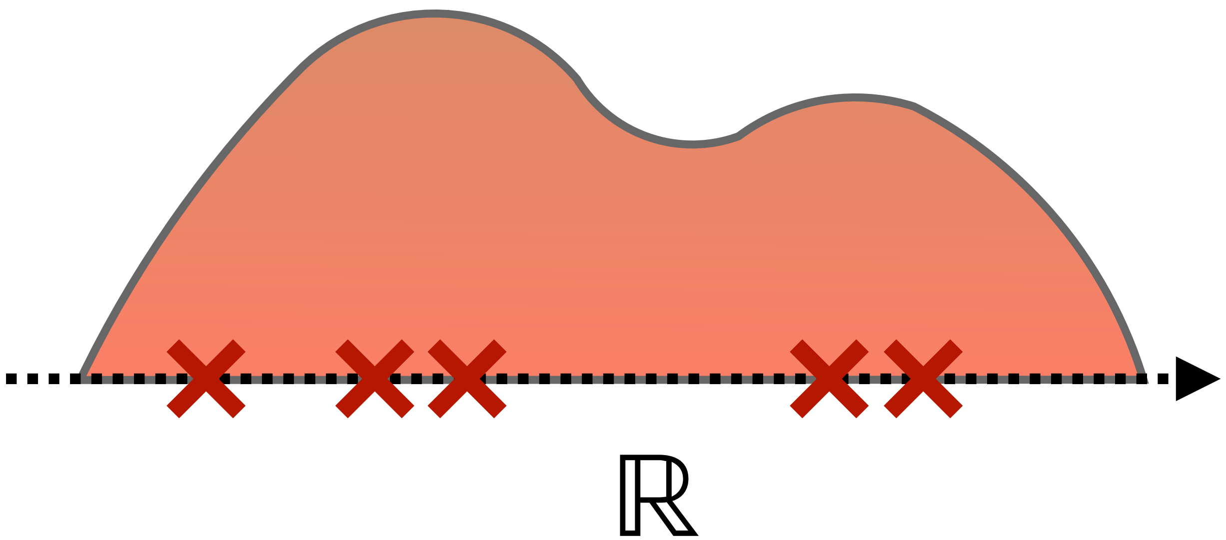
A Guided Tour through Optimal Transport

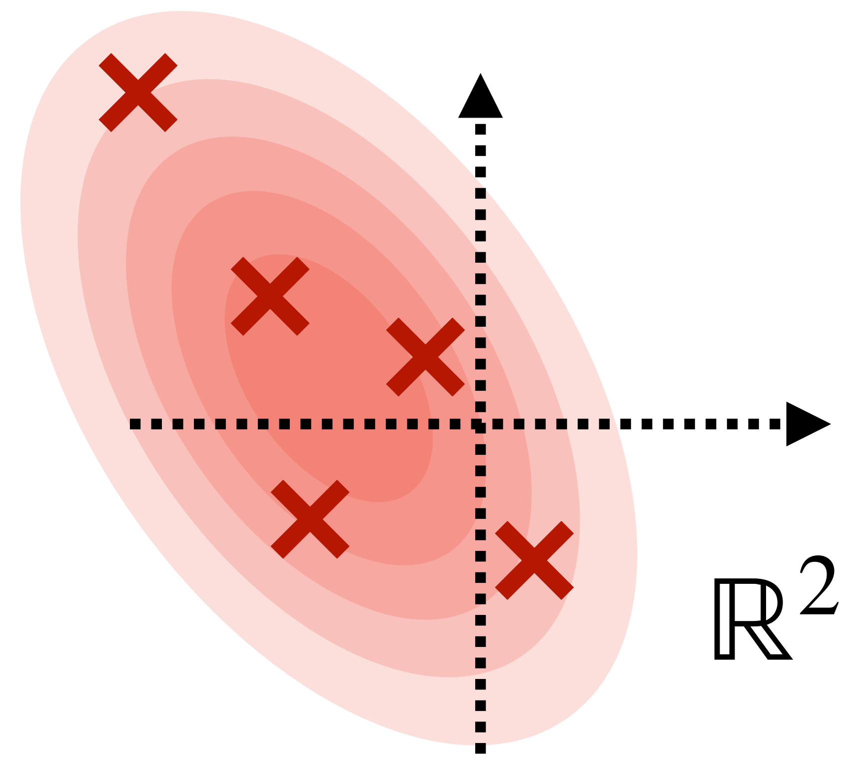
Marcel Klatt
Institute for Mathematical Stochastics
Georg-August University Göttingen

University of Twente, 6th April 2022

Probability Measures

Let $\mathscr{X} \neq \varnothing$ and $\mathscr{A} \subset 2^{\mathscr{X}}$ be a σ-Algebra. A σ-additive* set function $\mu\colon \mathscr{A} \to [0,\infty]$ with $\mu(\mathscr{X}) = 1$ is a probability measure.





$$\mu = \sum_{i=1}^{K} \mu_i \delta_{x_i}$$

$$\mu = \sum_{i=1}^{K} \mu_i \delta_{(x_i, y_i)}$$

$$\mu = f(x)\,\mathrm{d}x$$

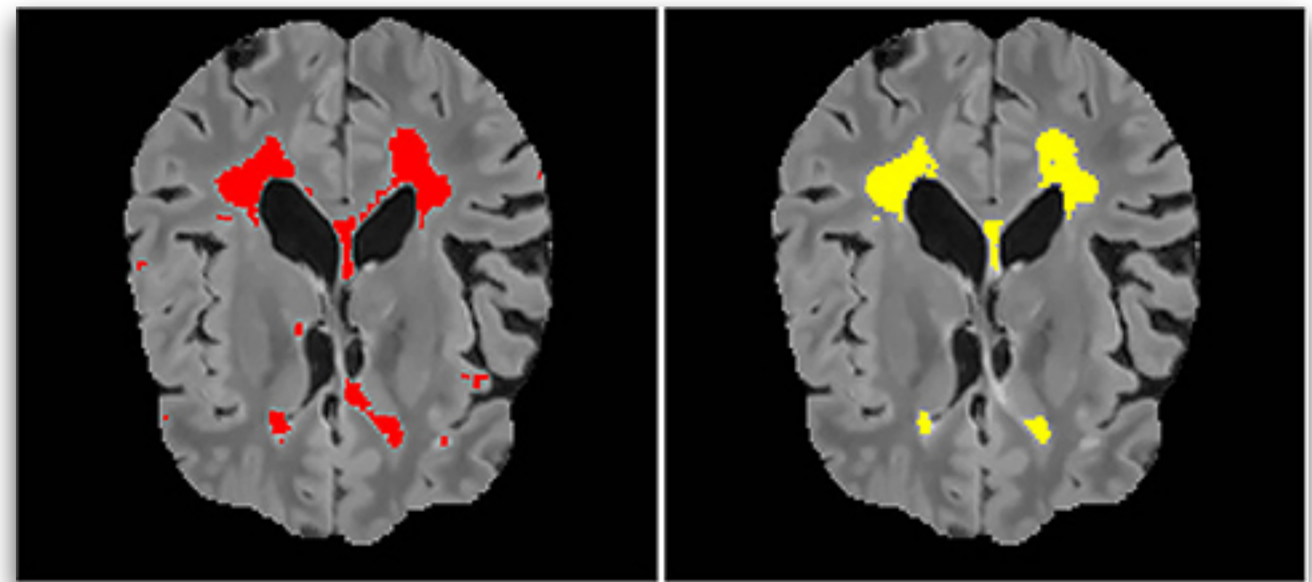
$$\mu = f(x,y)\,\mathrm{d}(x,y)$$

* $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ for any choice of countably many mutually disjoint sets $A_1, A_2, \ldots \in \mathscr{A}$

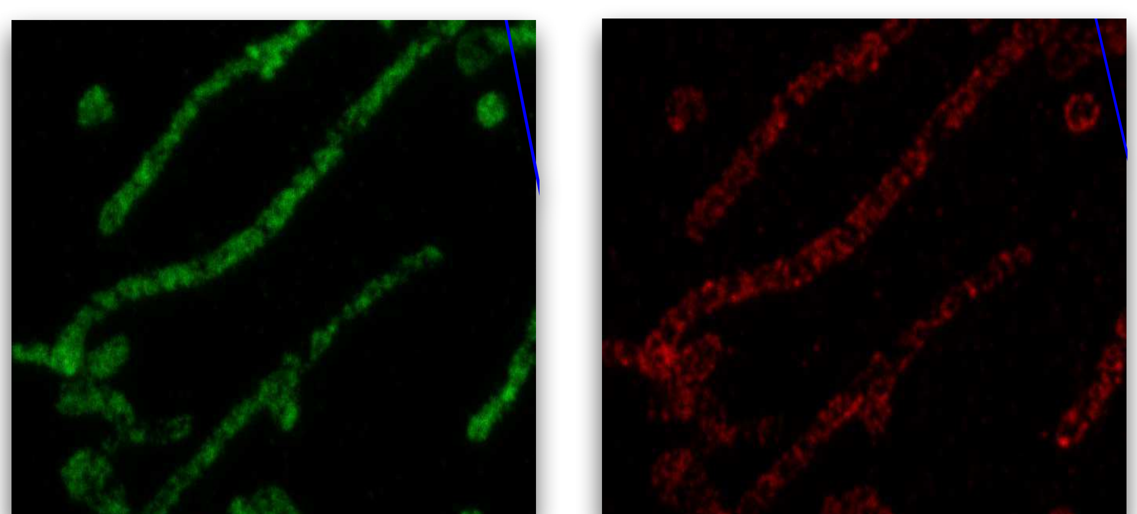
Probability Measures

Let $\mathscr{X} \neq \emptyset$ and $\mathscr{A} \subset 2^{\mathscr{X}}$ be a σ-Algebra. A σ-additive* set function $\mu : \mathscr{A} \to [0,\infty]$ with $\mu(\mathscr{X}) = 1$ is a probability measure.

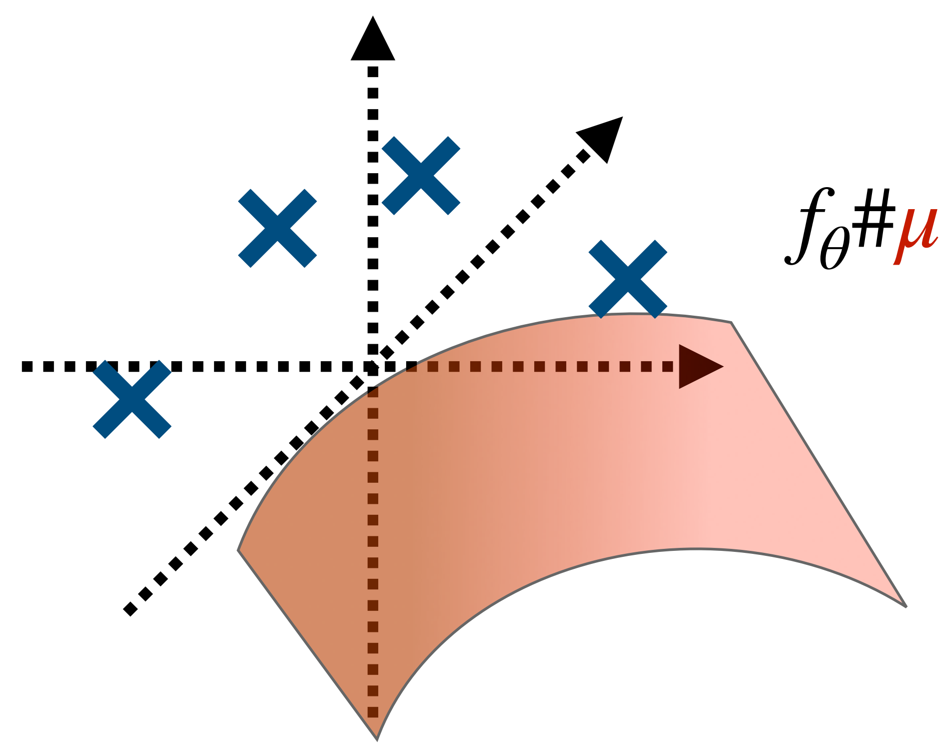
Basic Statistics



Neuroimaging Data



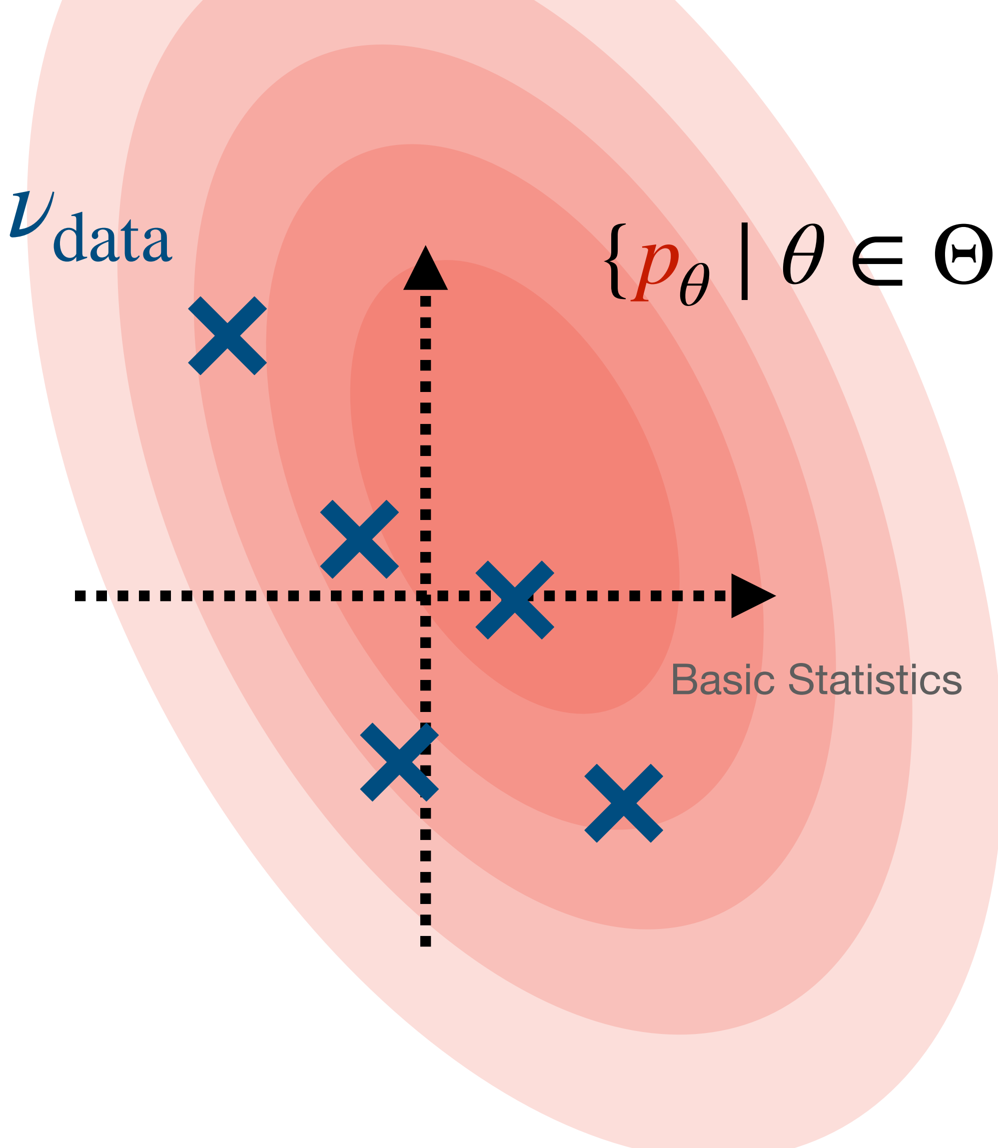
Colocalization Analysis



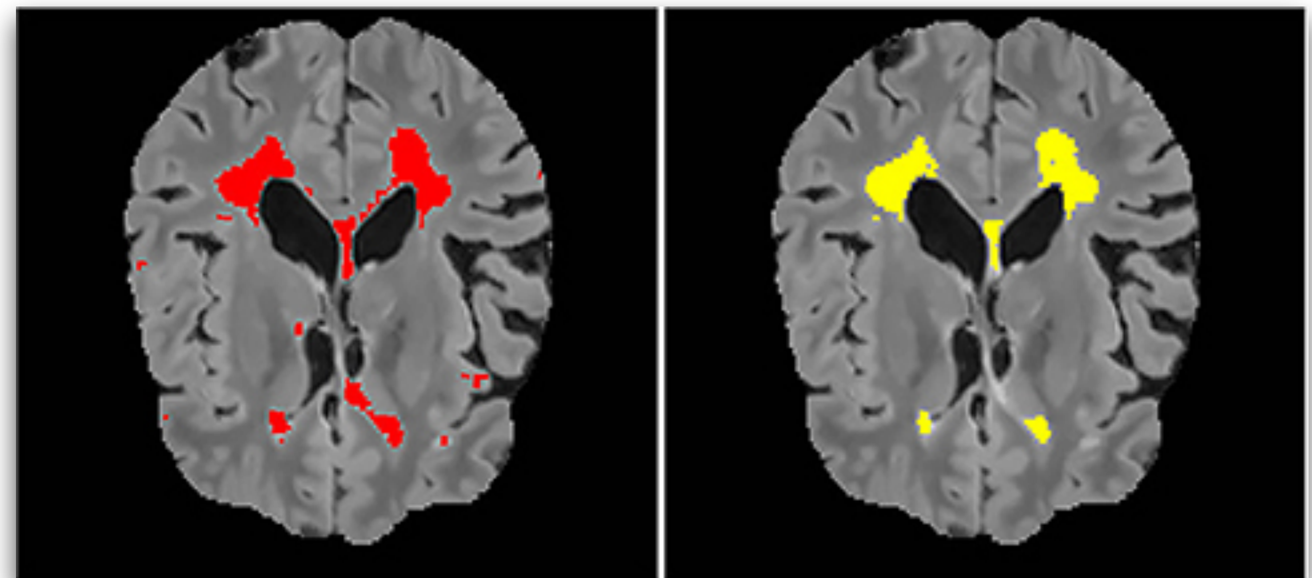
Machine Learning: GAN

Probability Measures

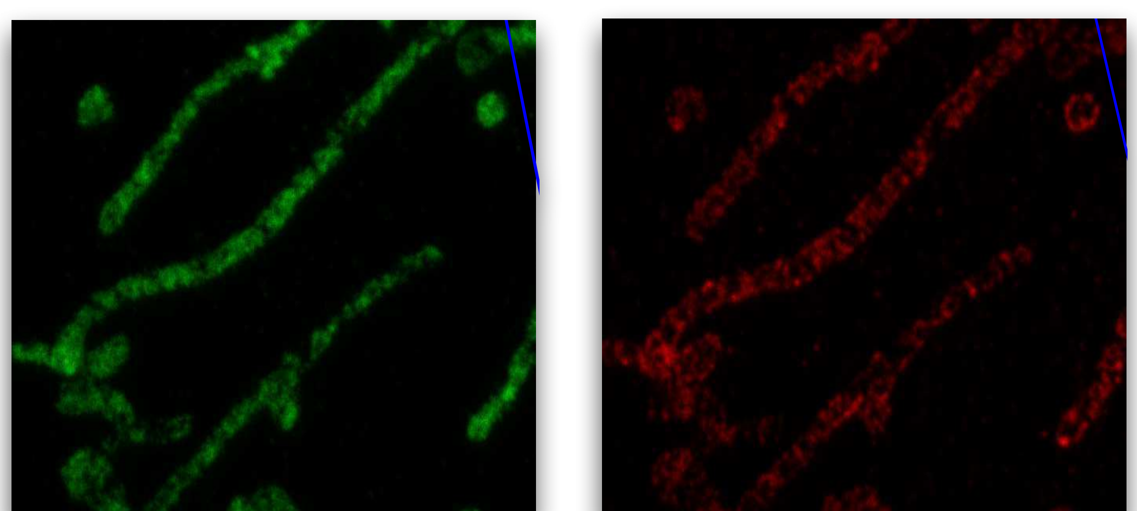
Let $\mathscr{X} \neq \varnothing$ and $\mathscr{A} \subset 2^{\mathscr{X}}$ be a σ-Algebra. A σ-additive* set function $\mu : \mathscr{A} \to [0,\infty]$ with $\mu(\mathscr{X}) = 1$ is a probability measure.



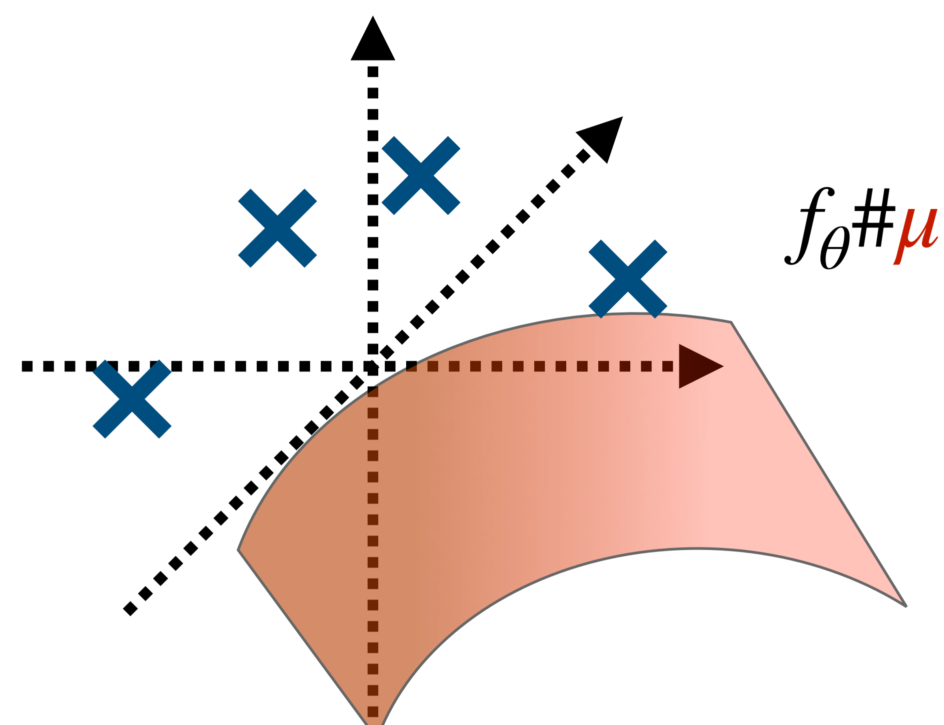
Basic Statistics



Neuroimaging Data



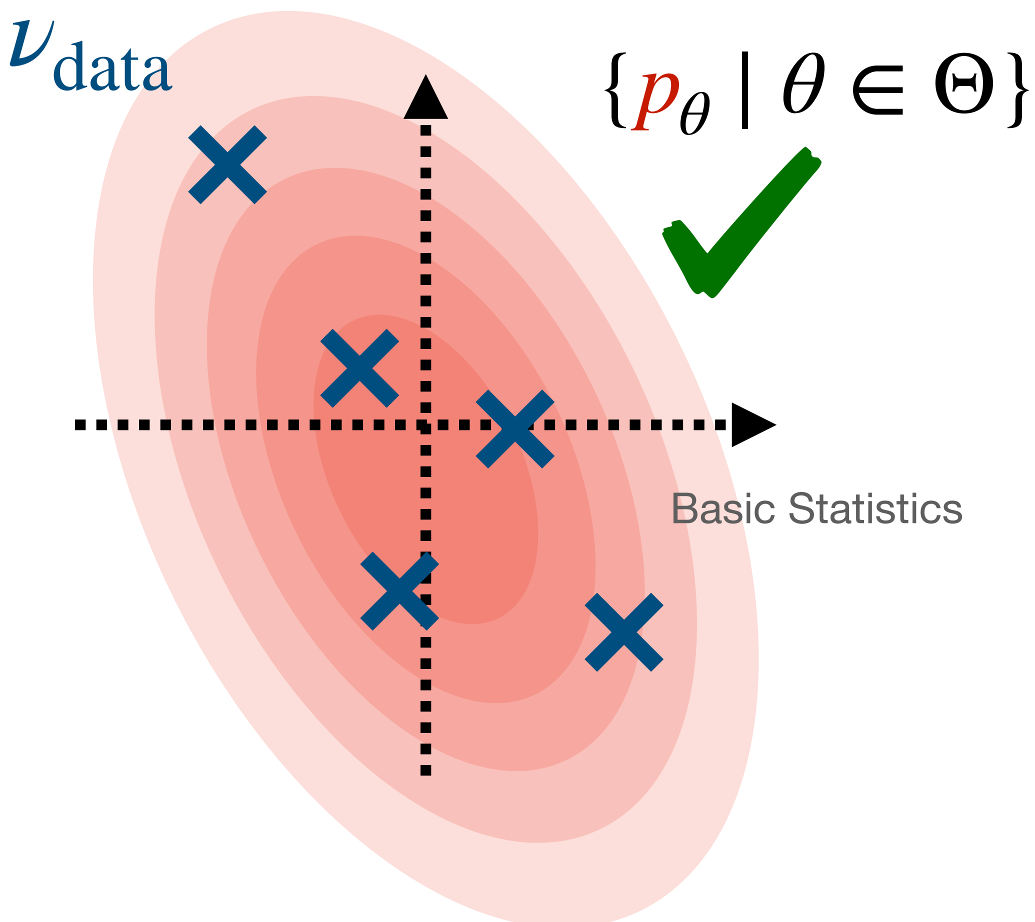
Colocalization Analysis



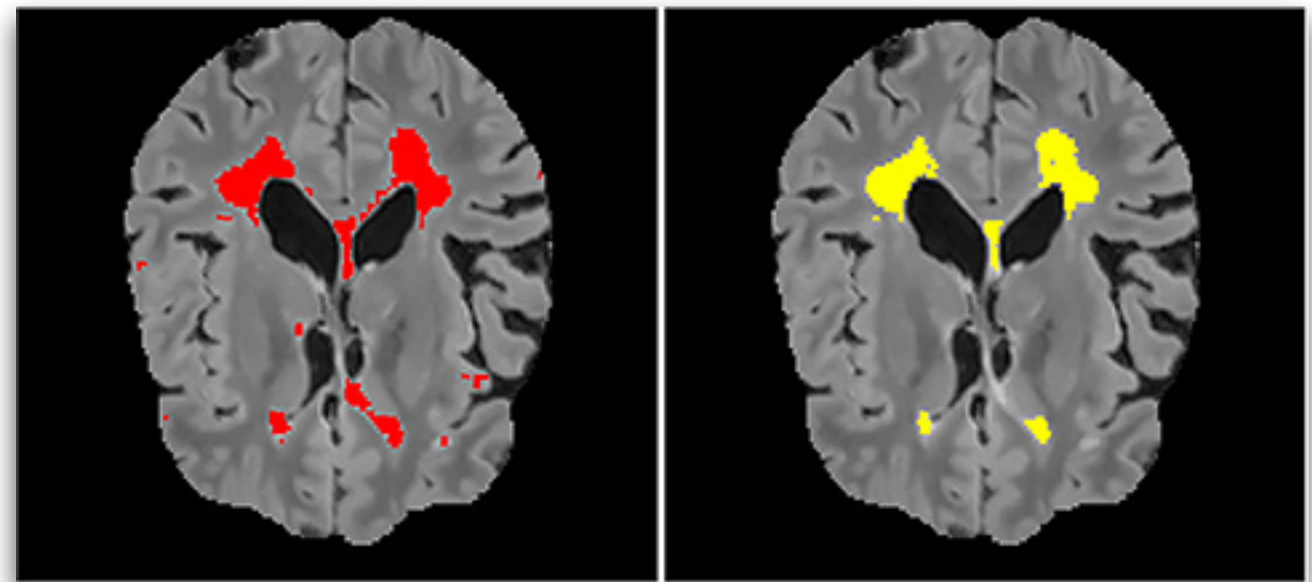
Machine Learning: GAN

Probability Measures

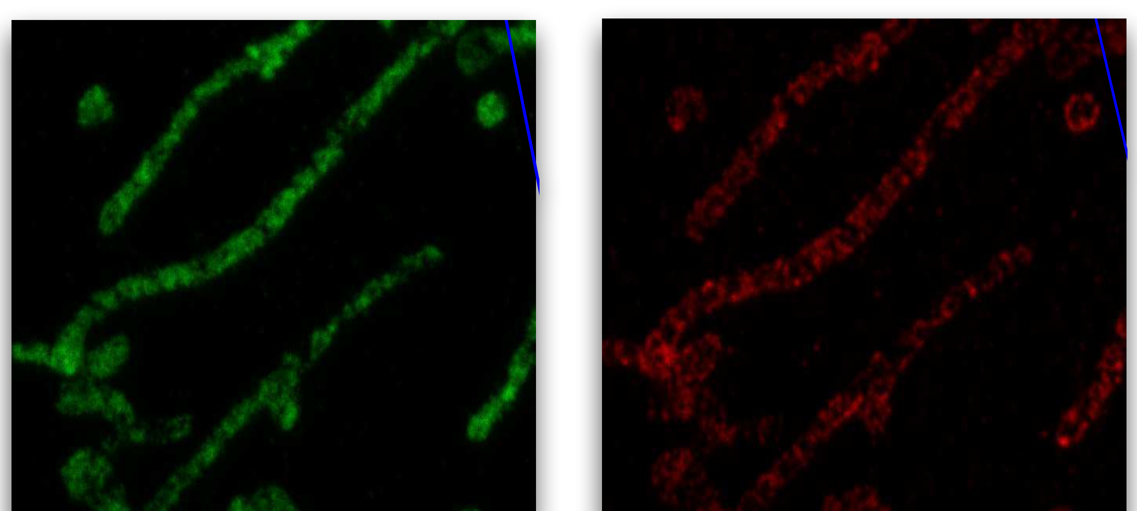
Let $\mathscr{X} \neq \emptyset$ and $\mathscr{A} \subset 2^{\mathscr{X}}$ be a σ-Algebra. A σ-additive* set function $\mu : \mathscr{A} \to [0,\infty]$ with $\mu(\mathscr{X}) = 1$ is a probability measure.



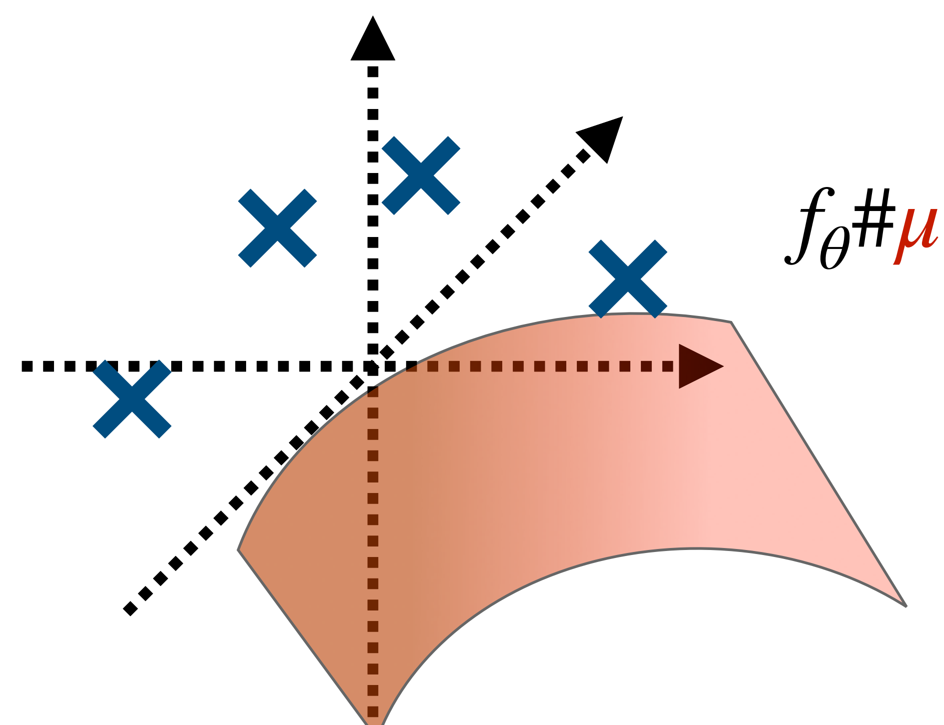
Basic Statistics



Neuroimaging Data



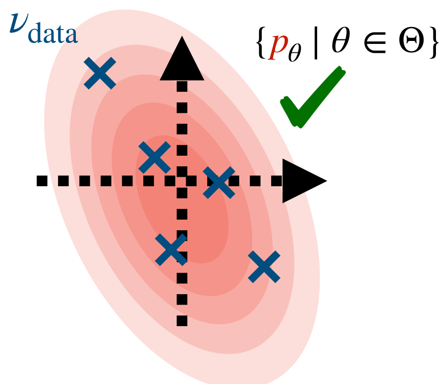
Colocalization Analysis

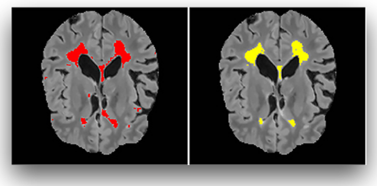


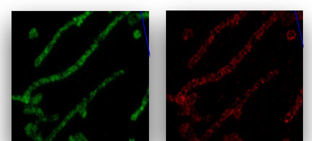
Machine Learning: GAN

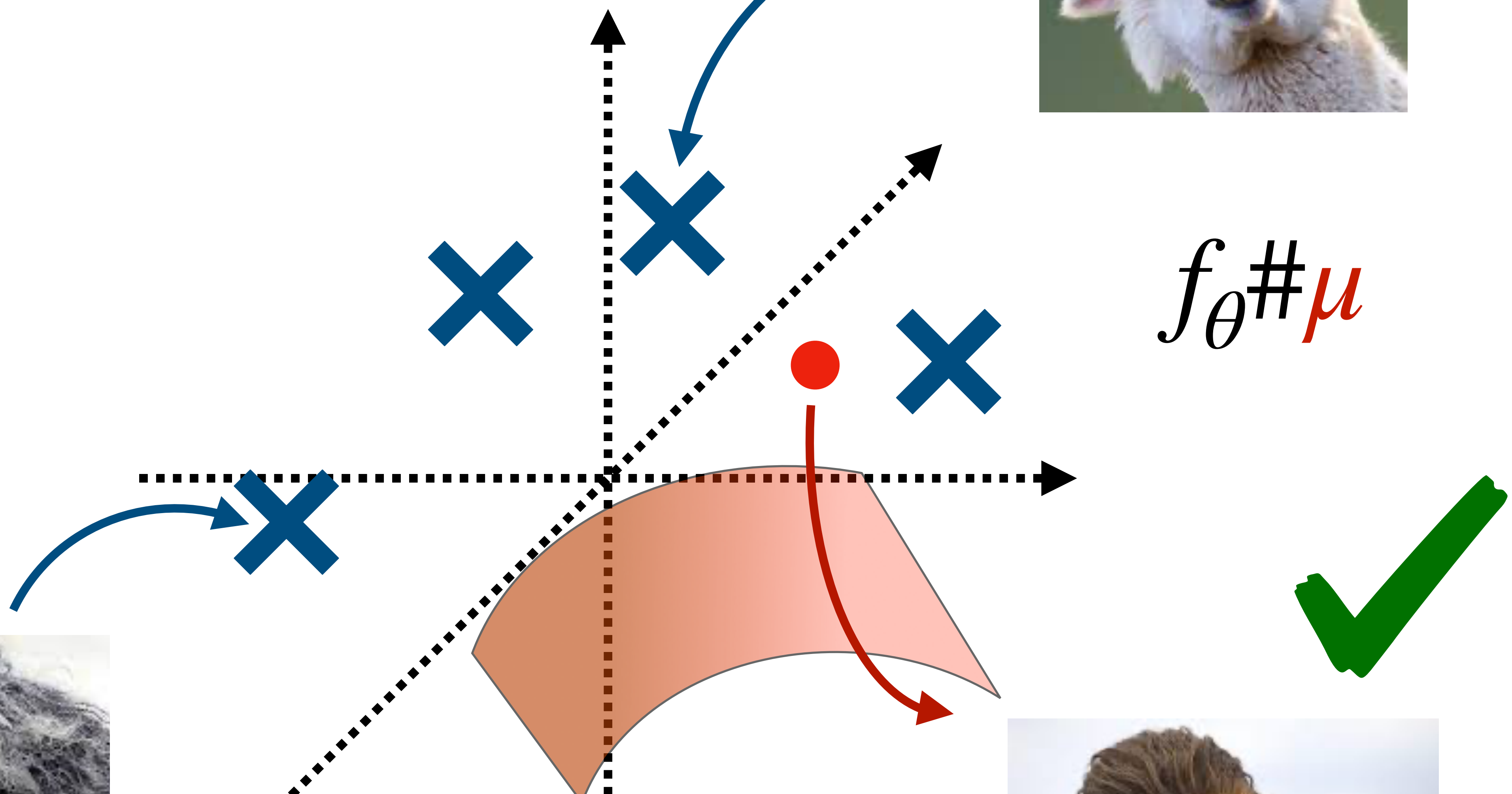
Probability Measures

Let $\mathscr{X} \neq \emptyset$ and $\mathscr{A} \subset 2^{\mathscr{X}}$ be a σ-Algebra. A σ-additive* set function $\mu : \mathscr{A} \to [0,\infty]$ with $\mu(\mathscr{X}) = 1$ is a probability measure.

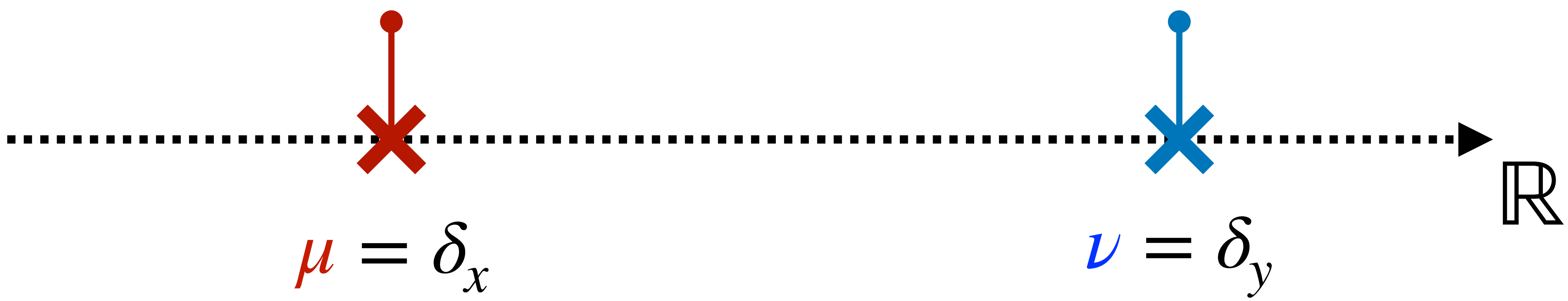






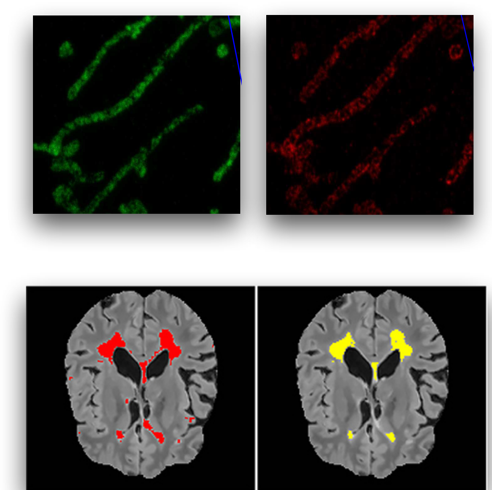


(Dis)Similarities between Probability Measures



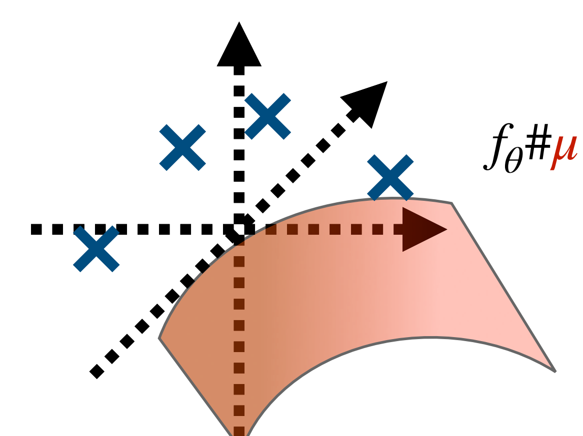
Kullback-Leibler (KL) divergence

$$\mathrm{KL}(\mu \,||\, \nu) = \int \log\left(\frac{P_\mu(x)}{P_\nu(x)}\right) P_\mu(x)\,\mathrm{d}\tau(x) = \infty$$

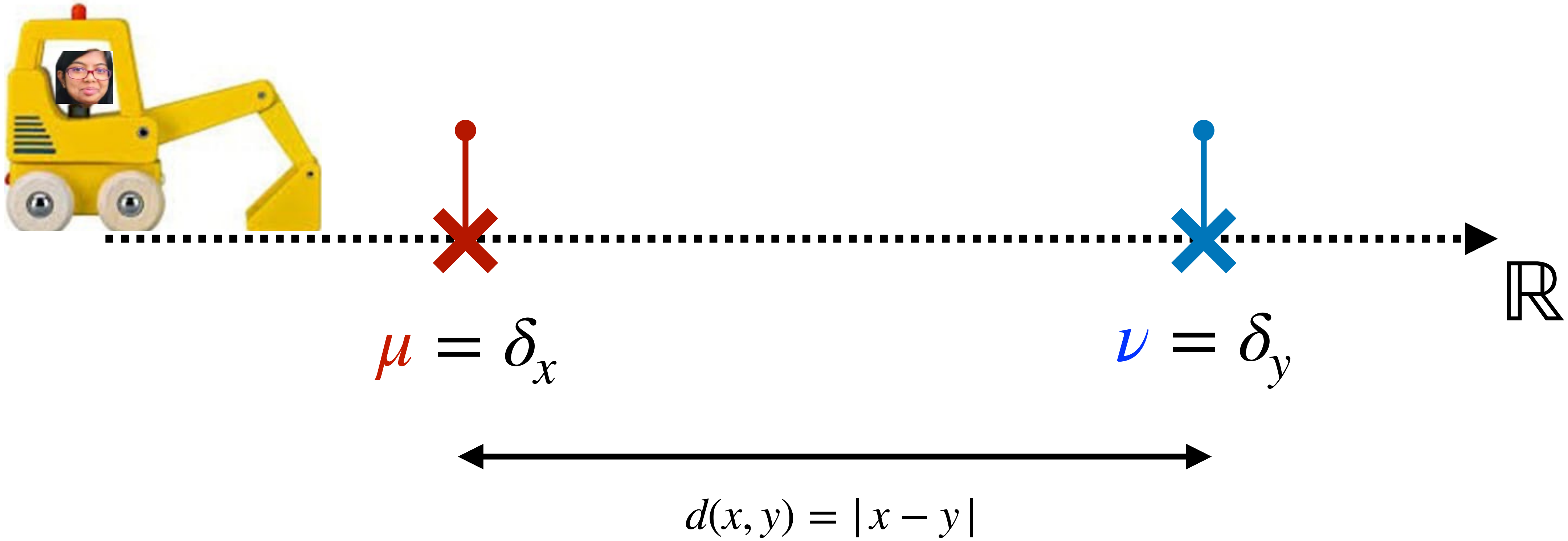


Total Variation (TV) distance

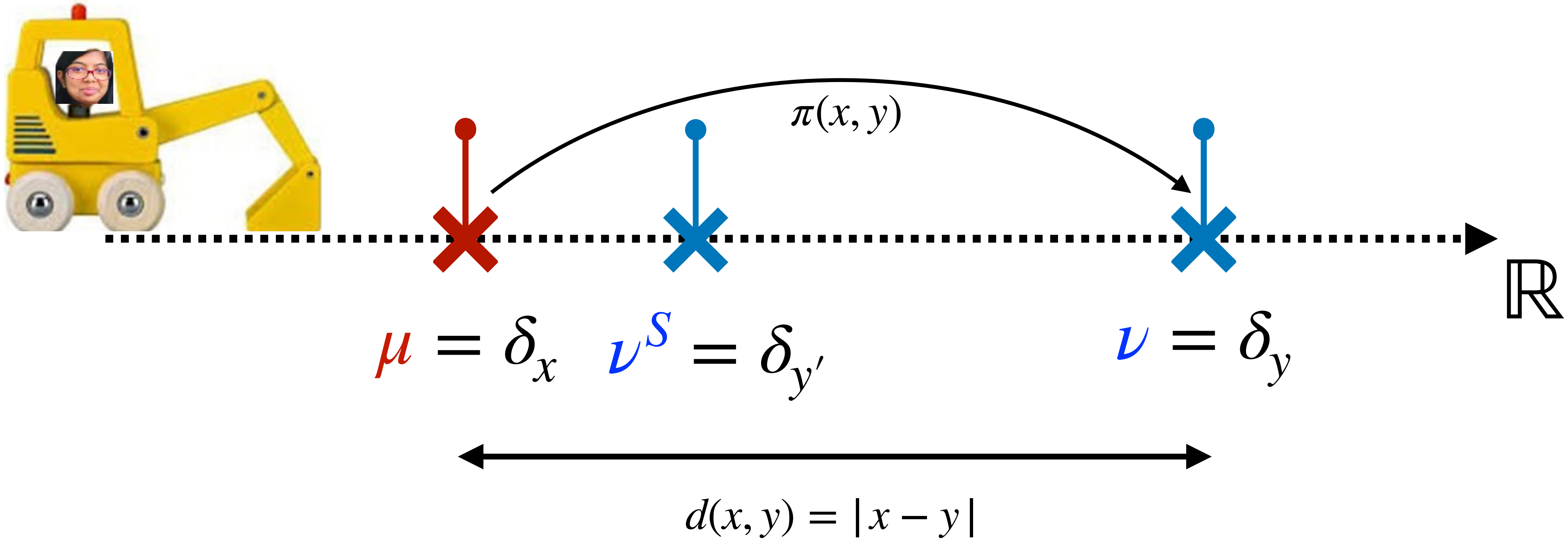
$$\mathrm{TV}(\mu, \nu) = \sup_{A \in \mathscr{A}} |\mu(A) - \nu(A)| = 1$$



Optimal Transport based (Dis)Similarity



Optimal Transport based (Dis)Similarity

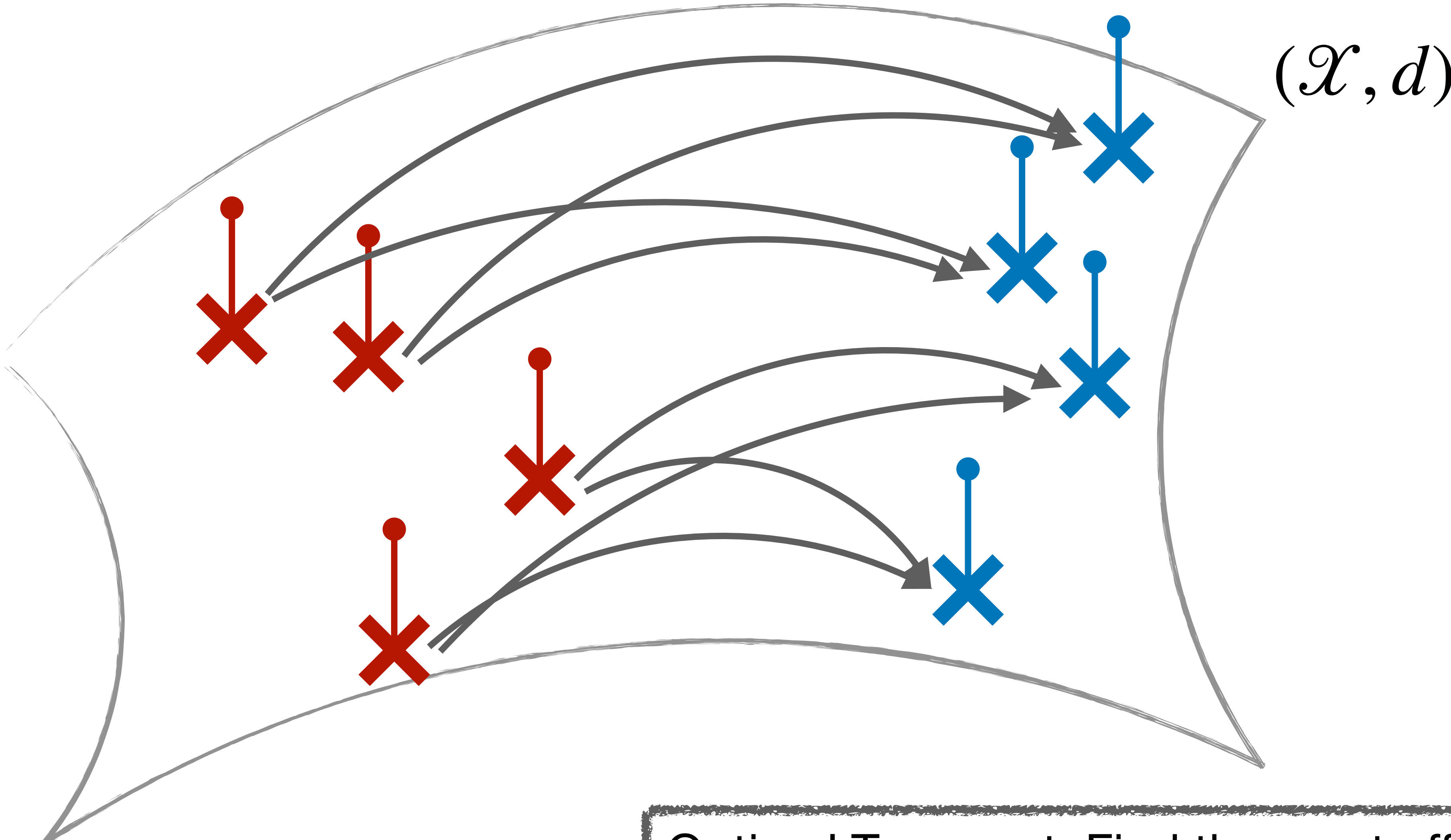


$$\mathrm{OT}_{|\cdot|}(\mu, \nu) = \pi(x, y) \cdot |x - y| = |x - y|$$

$$>$$

$$\mathrm{OT}_{|\cdot|}(\mu, \nu^S) = \pi(x, y') \cdot |x - y'| = |x - y'|$$

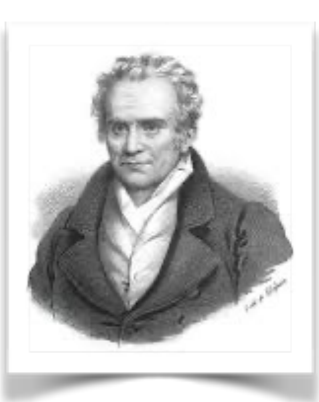
Optimal Transport based (Dis)Similarity



$$\mu = \sum_{i=1}^{K} \mu_i \delta_{x_i}$$

$$\nu = \sum_{j=1}^{K} \nu_j \delta_{x_j}$$

Optimal Transport: Find the most efficient way to transport μ to ν.



...

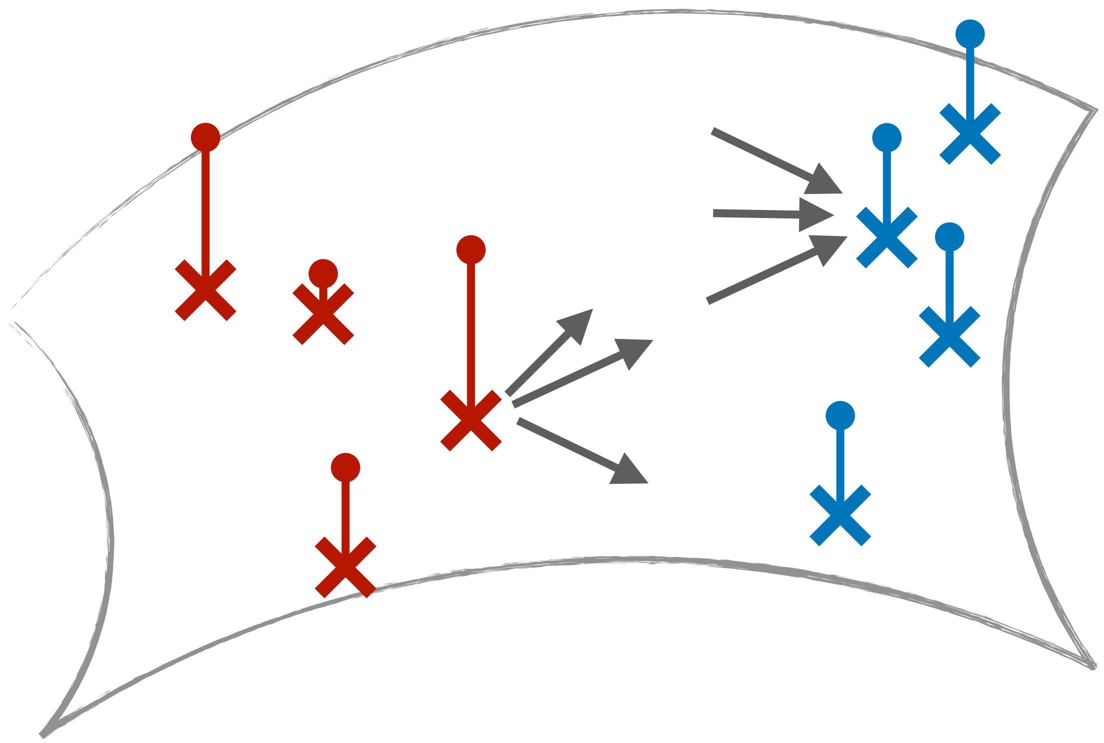
...

G. Monge (1781)

L. Kantorovich & T.C. Koopmans (1975)

C.Villani (2010) & A. Figalli (2018)

Optimal Transport based (Dis)Similarity



$$\mu = \sum_{i=1}^{K} \mu_i \delta_{x_i} \qquad \nu = \sum_{j=1}^{K} \nu_j \delta_{x_j}$$

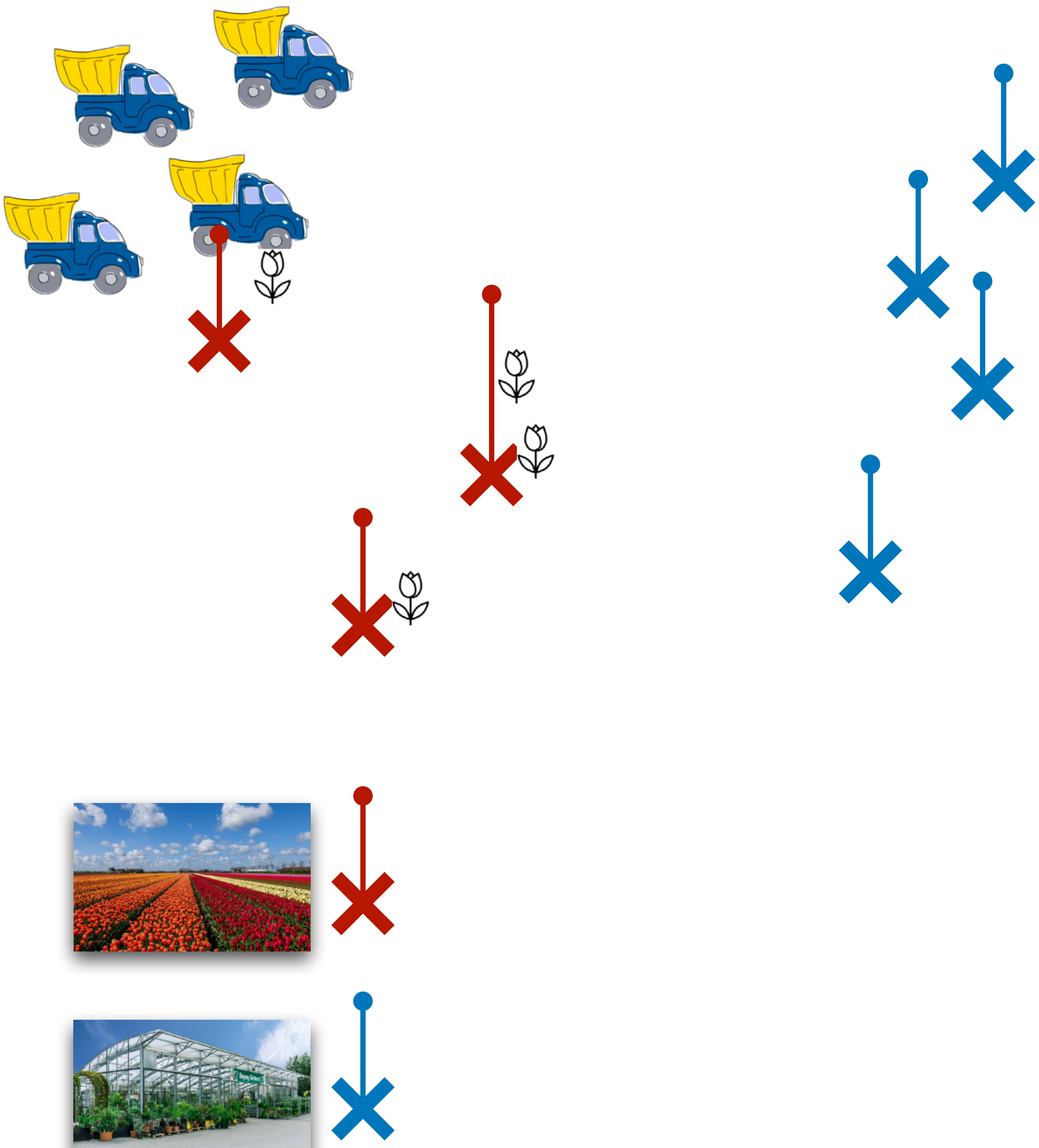
Optimal Transport: Find the most efficient way to transport μ to ν.

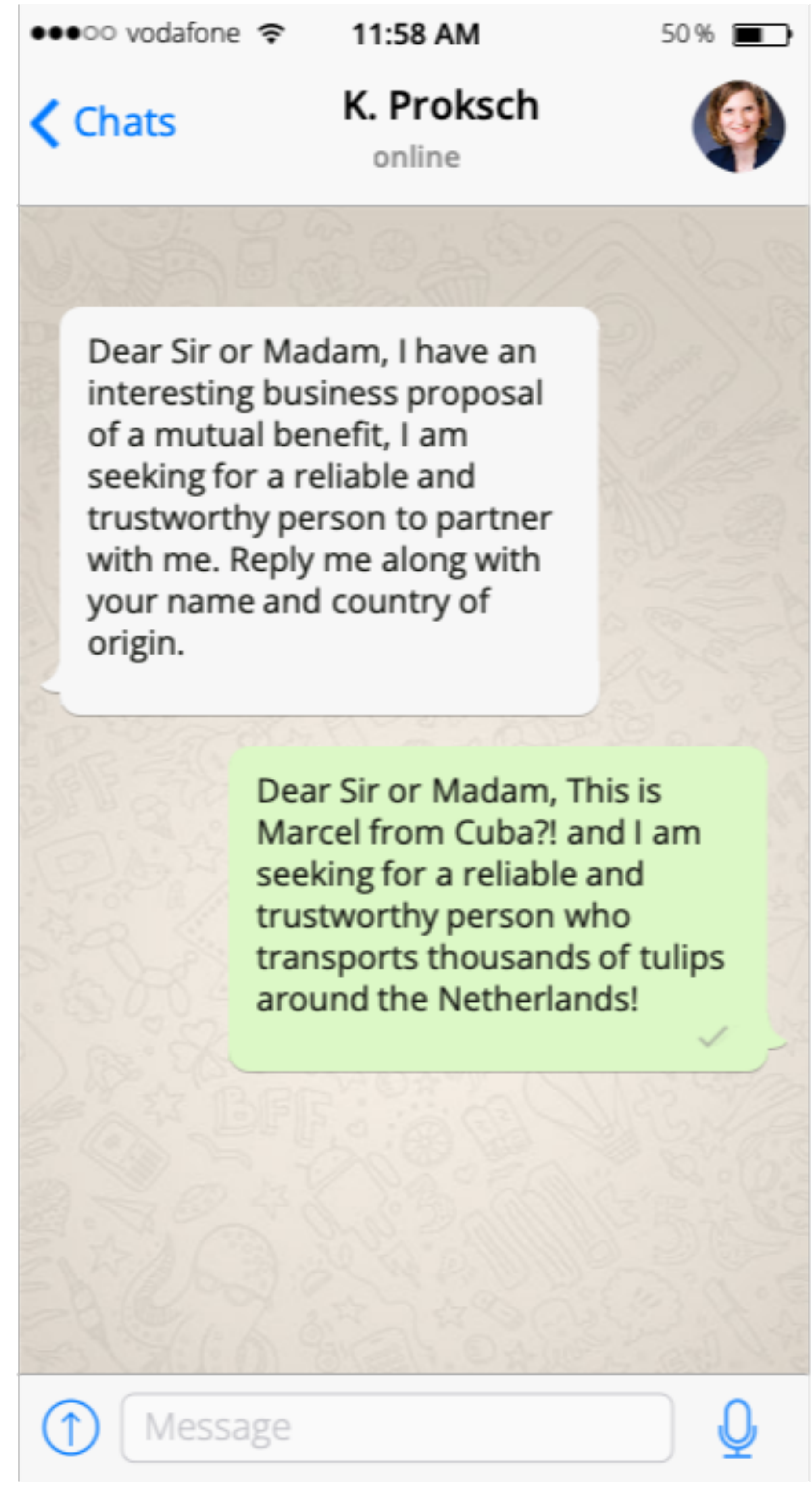
$$\left.\begin{aligned} \sum_{j=1}^{K} \pi(x_i, x_j) &= \mu_i, \forall i \\ \sum_{i=1}^{K} \pi(x_i, x_j) &= \nu_j, \forall j \end{aligned}\right\} \pi \in \Pi(\mu, \nu)$$

$$\mathrm{OT}_d(\mu, \nu) := \min_{\pi \in \Pi(\mu,\nu)} \sum_{i,j=1}^{K} d(x_i, y_j)\pi(x_i, y_j)$$

Optimal Transport: Kantorovich Duality

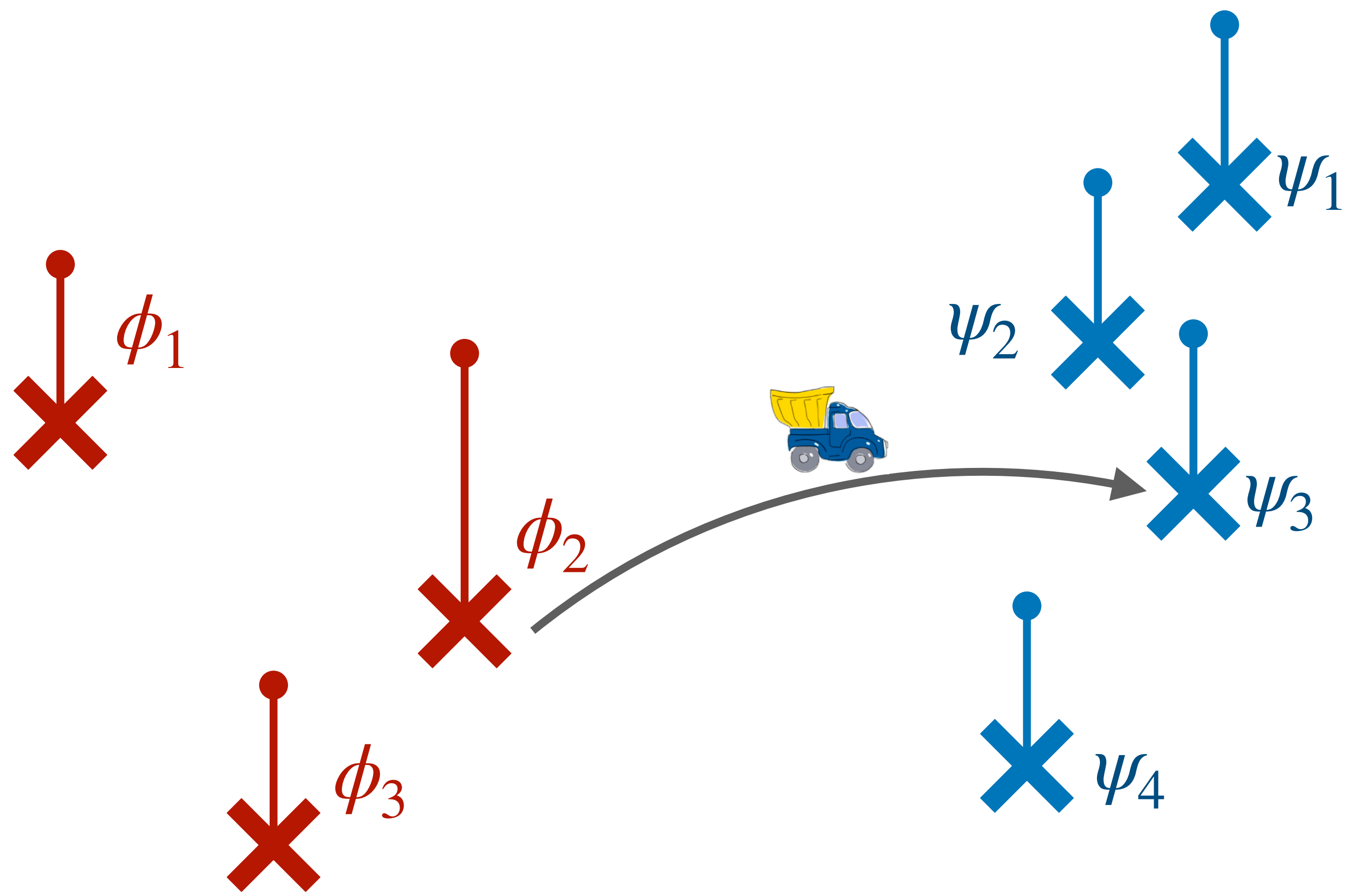
$$\mathrm{OT}_d(\mu, \nu) = \min_{\pi \in \Pi(\mu,\nu)} \sum_{i,j=1}^{K} d(x_i, y_j)\pi(x_i, y_j)$$





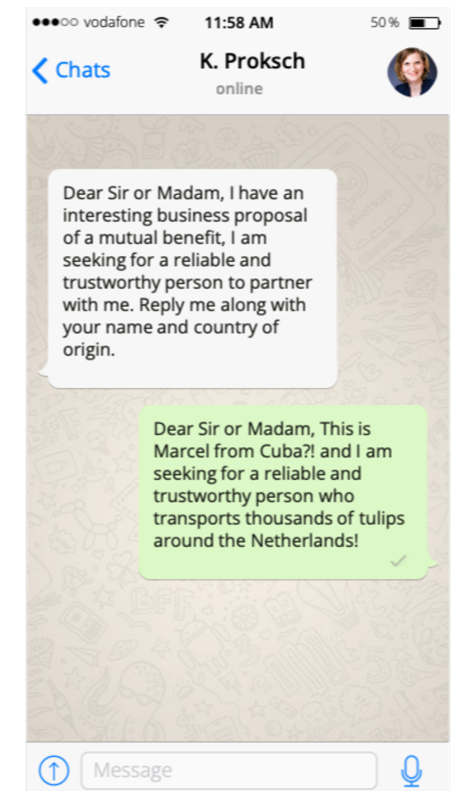
Optimal Transport: Kantorovich Duality

$$\mathrm{OT}_d(\mu, \nu) = \min_{\pi \in \Pi(\mu,\nu)} \sum_{i,j=1}^{K} d(x_i, y_j)\pi(x_i, y_j)$$



ϕ_i Price for loading

ψ_j Price for unloading

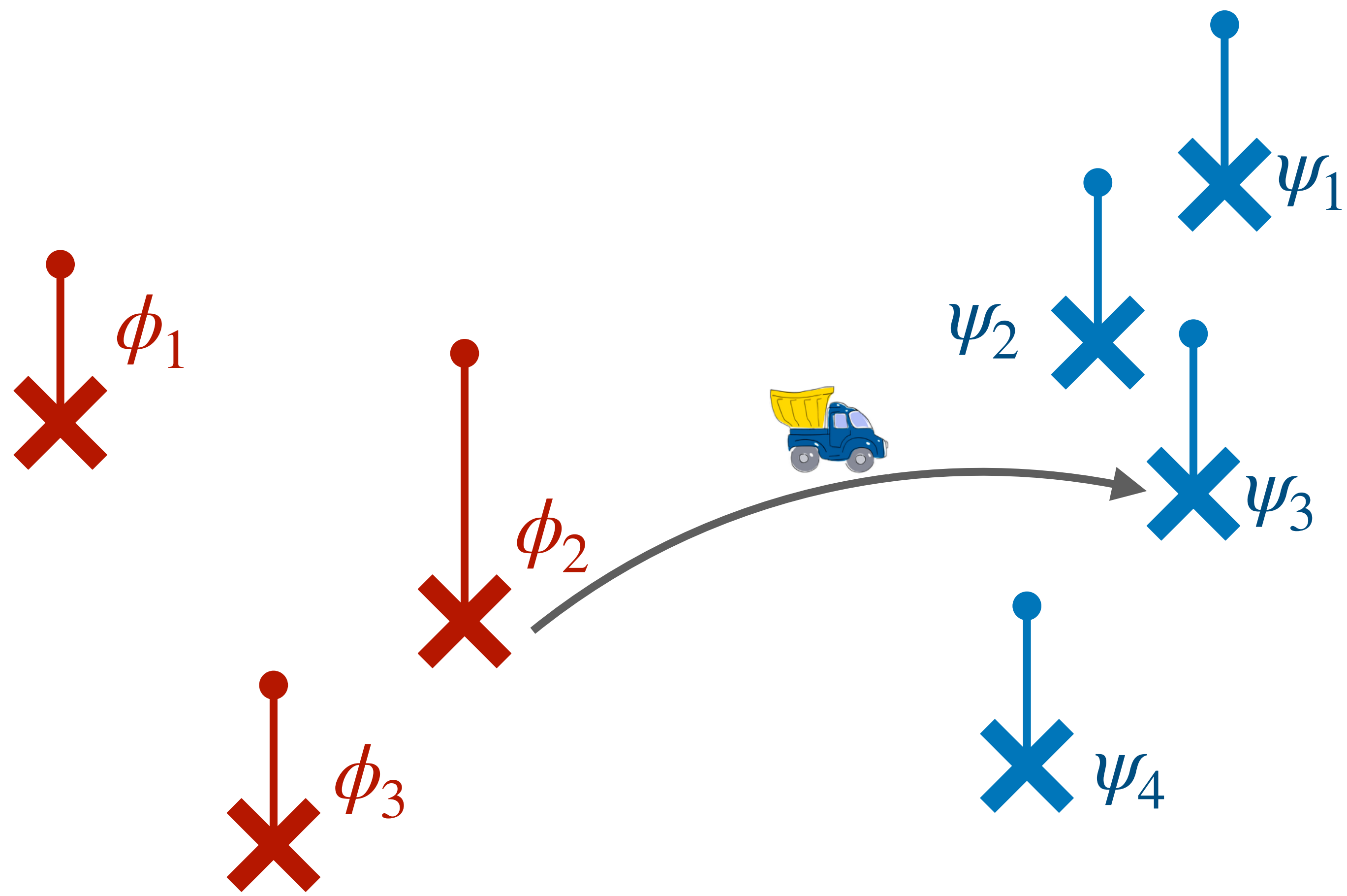


$$\phi_2 + \psi_3 > d(x_2, y_3)$$

$$\phi_i + \psi_j \leq d(x_i, y_j) \; \forall i, j$$

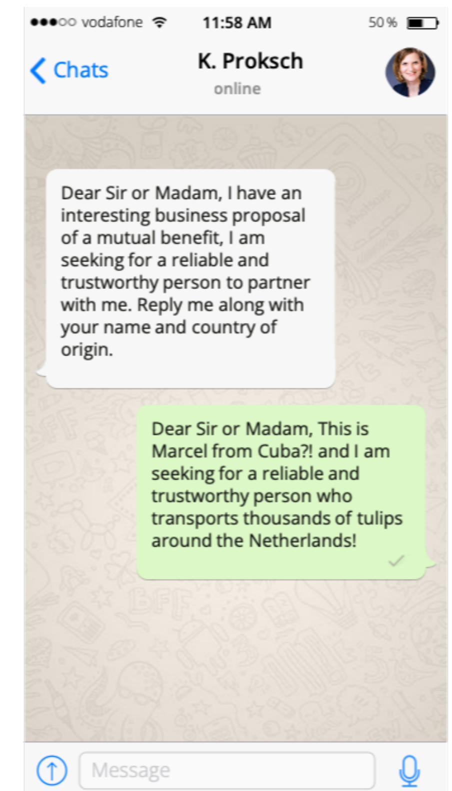
Optimal Transport: Kantorovich Duality

$$\mathrm{OT}_d(\mu, \nu) = \min_{\pi \in \Pi(\mu,\nu)} \sum_{i,j=1}^{K} d(x_i, y_j)\pi(x_i, y_j)$$



ϕ_i Price for loading

ψ_j Price for unloading

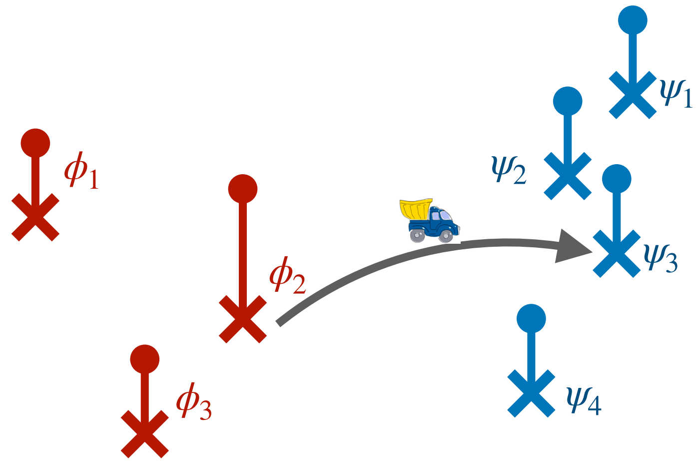


Katharina's profit

$$\max_{\phi_i + \psi_j \le d(x_i, y_j)\, \forall i,j} \sum_{i=1}^{K} \phi_i \mu_i + \sum_{j=1}^{K} \psi_j \nu_j$$

Optimal Transport: Kantorovich Duality

$$\mathrm{OT}_d(\mu, \nu) = \min_{\pi \in \Pi(\mu,\nu)} \sum_{i,j=1}^{K} d(x_i, y_j)\pi(x_i, y_j)$$



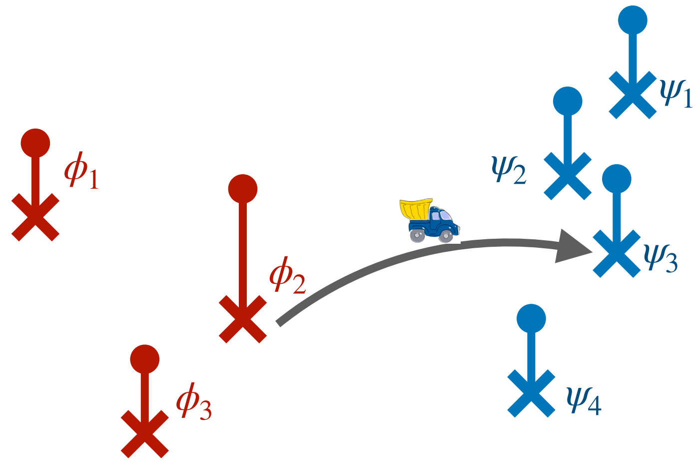
$\geq$

Weak Duality

$$\min_{\pi \in \Pi(\mu,\nu)} \sum_{i,j=1}^{K} d(x_i, y_j)\pi(x_i, y_j) \geq \max_{\phi_i + \psi_j \leq d(x_i, y_j)\, \forall i,j} \sum_{i=1}^{K} \phi_i \mu_i + \sum_{j=1}^{K} \psi_j \nu_j$$

Optimal Transport: Kantorovich Duality

$$\mathrm{OT}_d(\mu, \nu) = \min_{\pi \in \Pi(\mu,\nu)} \sum_{i,j=1}^{K} d(x_i, y_j)\pi(x_i, y_j)$$

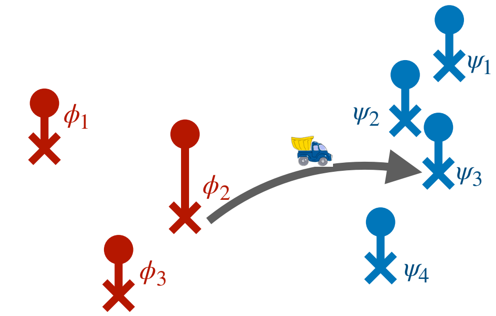


=

Strong Duality

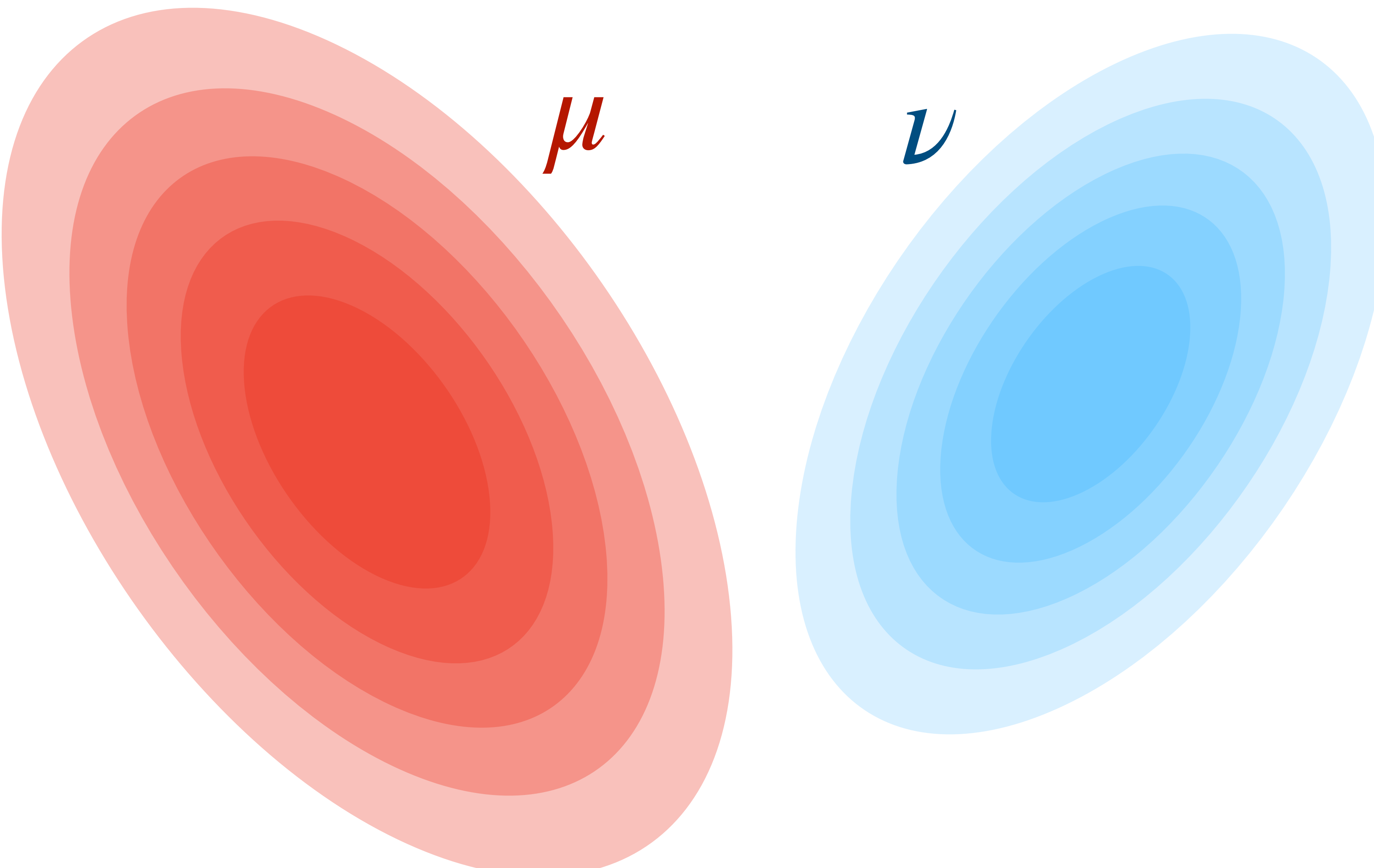
$$\min_{\pi \in \Pi(\mu,\nu)} \sum_{i,j=1}^{K} d(x_i, y_j)\pi(x_i, y_j) = \max_{\phi_i + \psi_j \leq d(x_i, y_j)\, \forall i,j} \sum_{i=1}^{K} \phi_i \mu_i + \sum_{j=1}^{K} \psi_j \nu_j$$

Optimal Transport:
From Discrete to Continuous

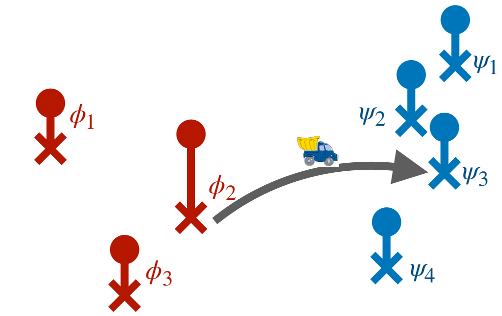


Strong Duality

$$\min_{\pi\in\Pi(\mu,\nu)} \sum_{i,j=1}^{K} d(x_i, y_j)\pi(x_i, y_j) = \max_{\phi_i + \psi_j \le d(x_i, y_j)\, \forall i,\, j} \sum_{i=1}^{K} \phi_i \mu_i + \sum_{j=1}^{K} \psi_j \nu_j$$

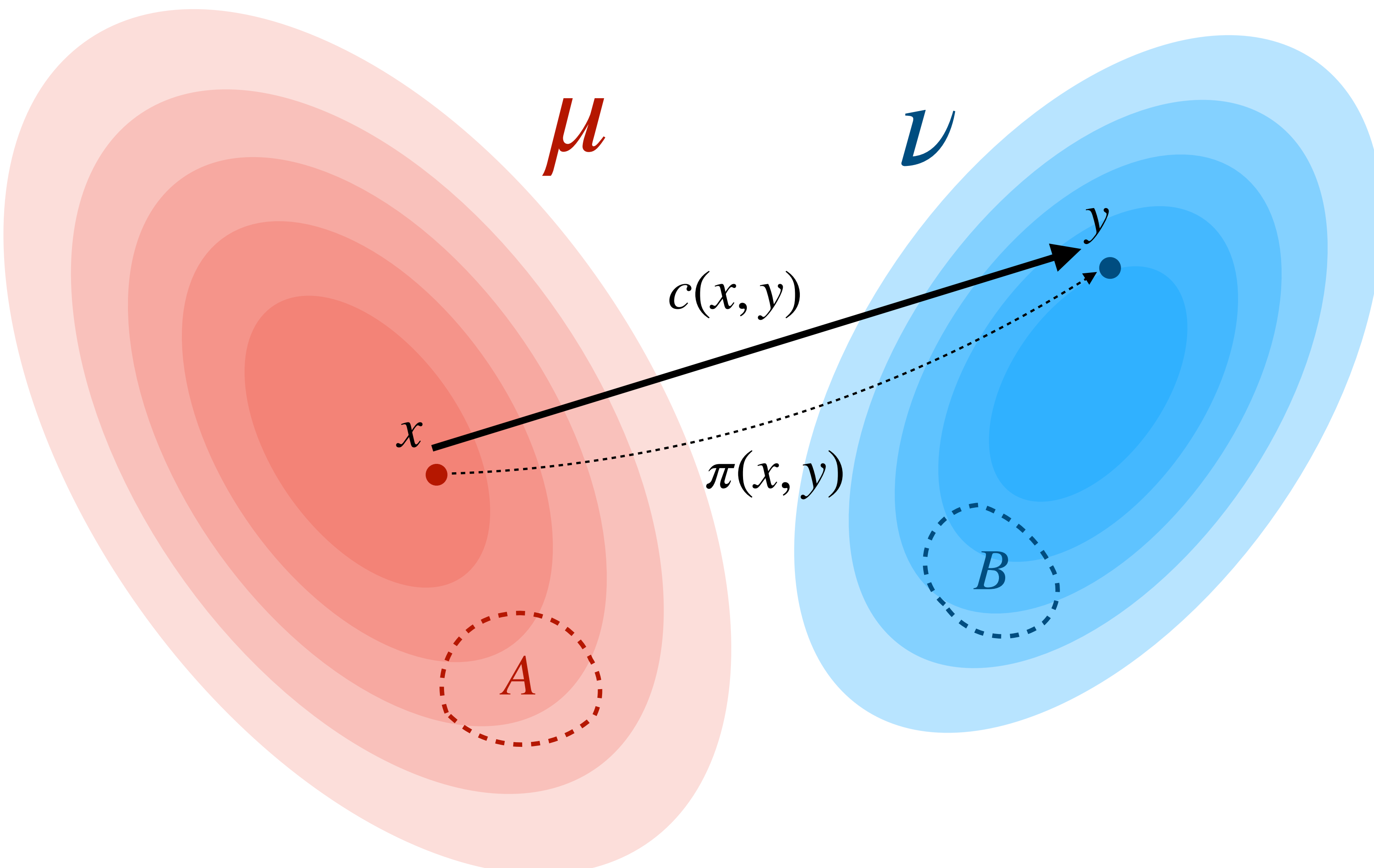


Optimal Transport: From Discrete to Continuous



Strong Duality

$$\min_{\pi \in \Pi(\mu,\nu)} \sum_{i,j=1}^{K} d(x_i, y_j)\pi(x_i, y_j) = \max_{\phi_i + \psi_j \le d(x_i, y_j)\,\forall i,j} \sum_{i=1}^{K} \phi_i \mu_i + \sum_{j=1}^{K} \psi_j \nu_j$$



$$\left.\begin{aligned} \pi(A \times \mathcal{X}) &= \mu(A) \\ \pi(\mathcal{X} \times B) &= \nu(B) \end{aligned}\right\} \pi \in \Pi(\mu, \nu)$$

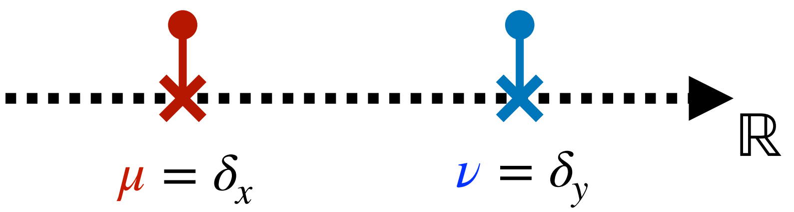
$$\mathrm{OT}_c(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{X}} c(x, y)\, \mathrm{d}\pi(x, y)$$

$$\boxed{!} = \sup_{\substack{\phi, \psi \in C_b(\mathcal{X}) \\ \phi(x) + \psi(y) \le c(x,y)}} \int_{\mathcal{X}} \phi(x)\, \mathrm{d}\mu(x) + \int_{\mathcal{X}} \psi(x)\, \mathrm{d}\nu(x)$$

Optimal Transport: Concrete Costs

$$\mathrm{OT}_c(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{X}} c(x, y)\, \mathrm{d}\pi(x, y)$$

Cost: $c(x, y) = 1_{\{x \neq y\}}$ —> Total Variation



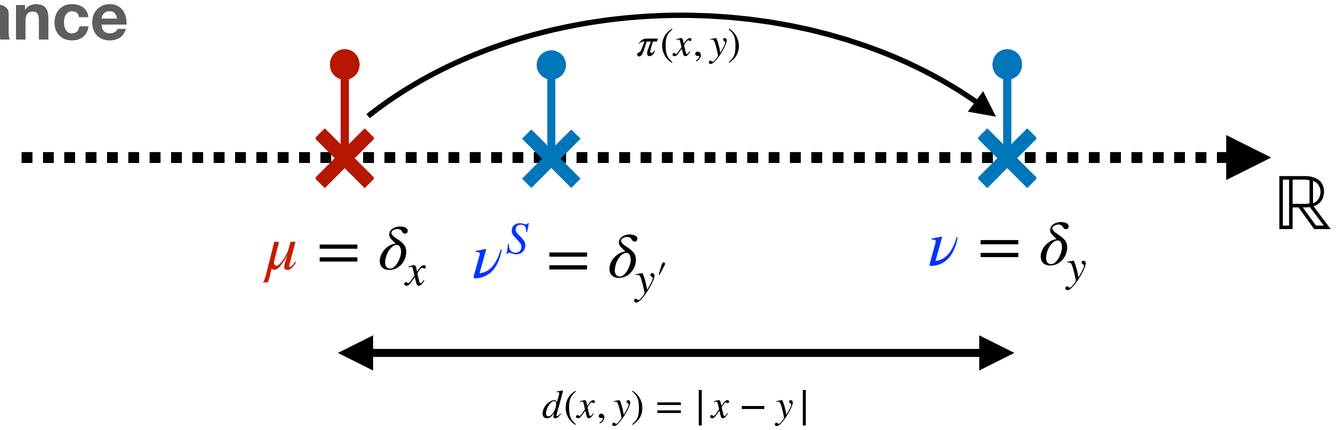
$$\mathrm{OT}_c(\mu, \nu) = \mathrm{TV}(\mu, \nu)$$

Cost: $c(x, y) = d(x, y)$ —> Kantorovich-Rubinstein Formula

$$\mathrm{OT}_d(\mu, \nu) = \sup_{\phi \in \mathrm{Lip}_1(\mathcal{X})} \int_{\mathcal{X}} \phi(x)\, \mathrm{d}(\mu - \nu)(x)$$

… the real line (D=1), squared Euclidean costs, (ultra)metric trees …

Wasserstein (Monge-Kantorovich) Distance



Cost $c(x, y) = d^p(x, y), p \geq 1$ —> Wasserstein Distance

$$W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{X}} d^p(x, y)\, \mathrm{d}\pi(x, y) \right)^{1/p}$$

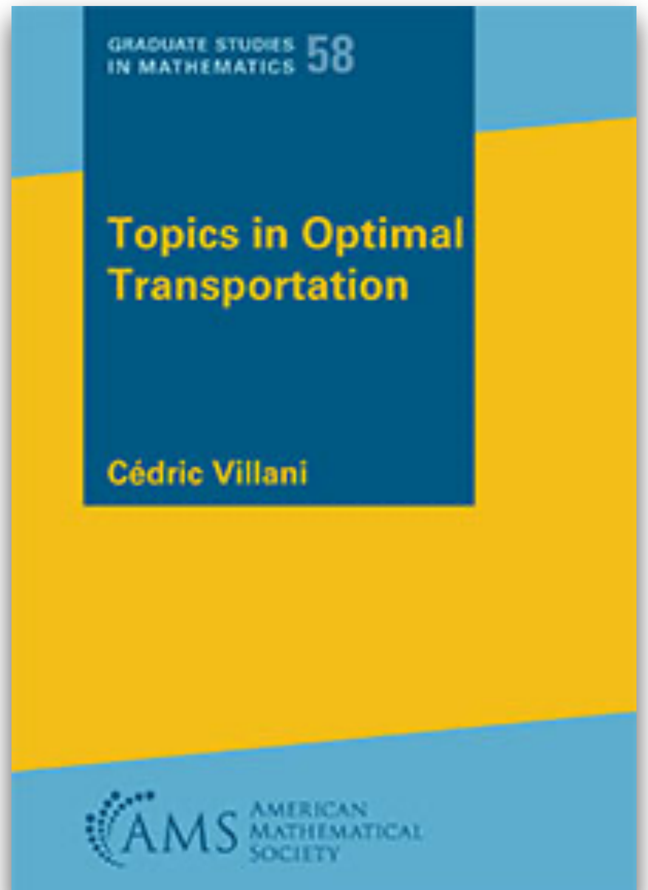
For all $p \in [1, \infty)$ the Wasserstein distance defines a finite metric on the set of probability measures with finite moments of order p, i.e. measures μ such that for some $x_0 \in \mathcal{X}$,

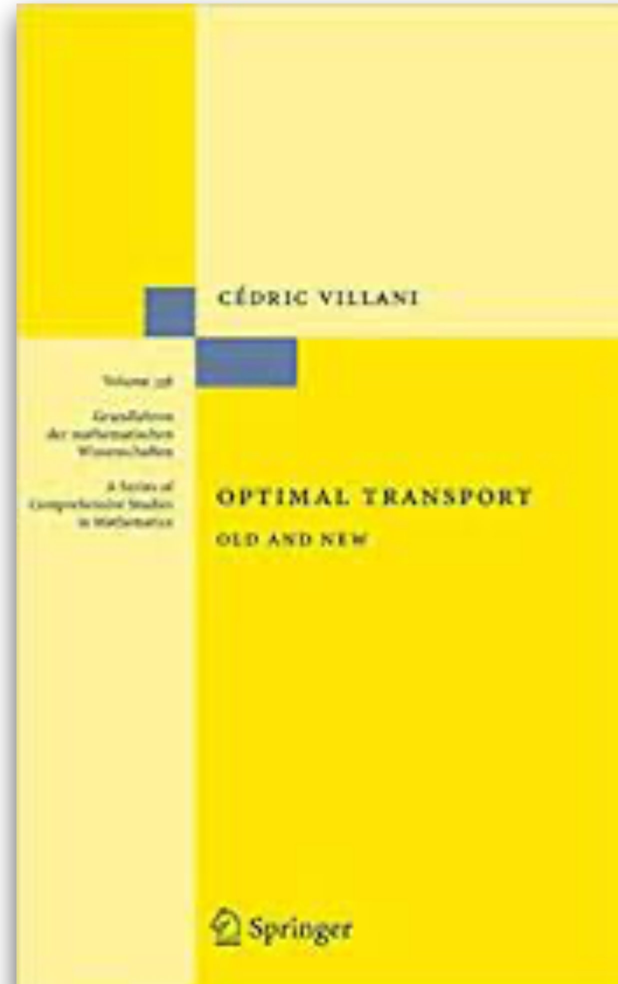
$$\int_{\mathcal{X}} d^p(x_0, x)\, \mathrm{d}\mu(x) < \infty\,.$$

Further Reading…

$$W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{X}} d^p(x, y) \, \mathrm{d}\pi(x, y) \right)^{1/p}$$

1. Theory





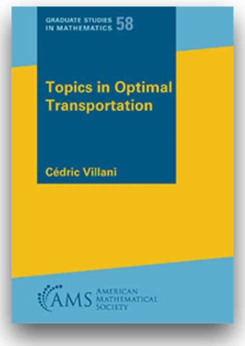
Cédric Villani. Topics in Optimal Transportation. *American Mathematical Society,* 2021

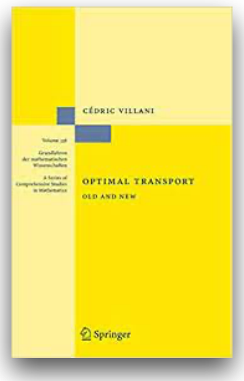
Cédric Villani. Optimal Transport: Old and New. *Springer,* 2009

Further Reading…

$$W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d^p(x, y)\, \mathrm{d}\pi(x, y) \right)^{1/p}$$

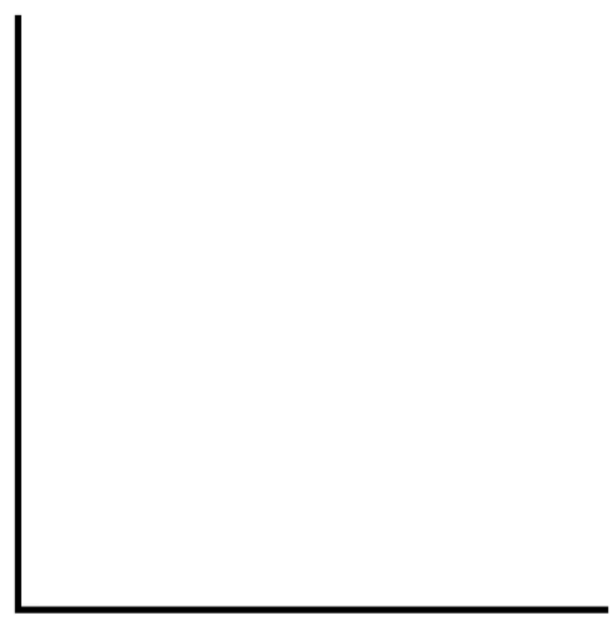
1. Theory





Let μ, ν be two probability distributions with finite second moments over $\mathbb{R}^D$ and $c(x, y) = 1/2\, \|x - y\|_2^2$. Suppose that μ is absolutely continuous w.r.t. Lebesgue measure. Then there exists, unique, an optimal transport map T from μ to ν, and it is of the form

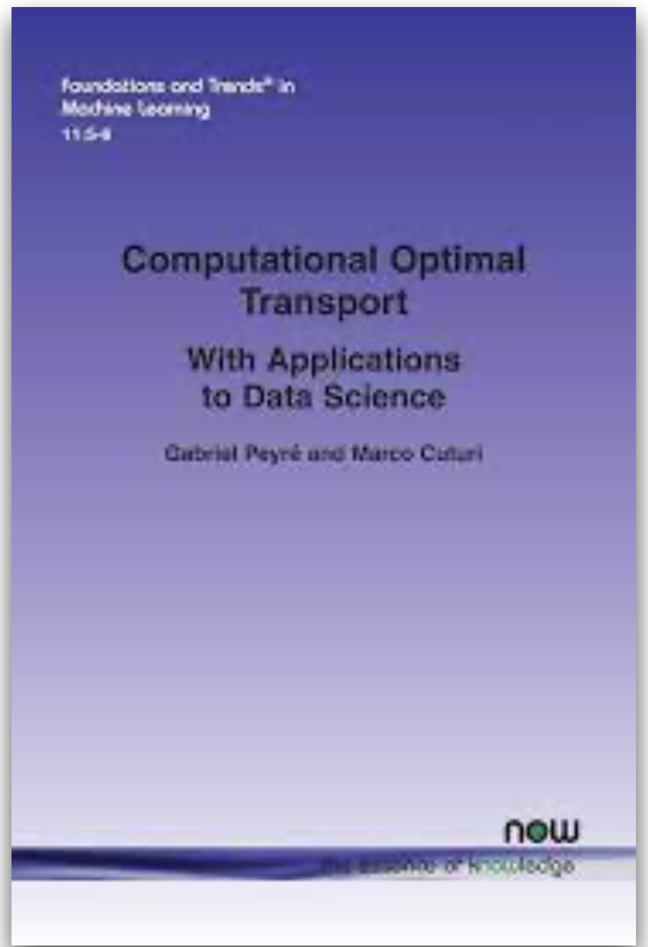
$$T = \nabla u, \text{ for a convex function } u.$$

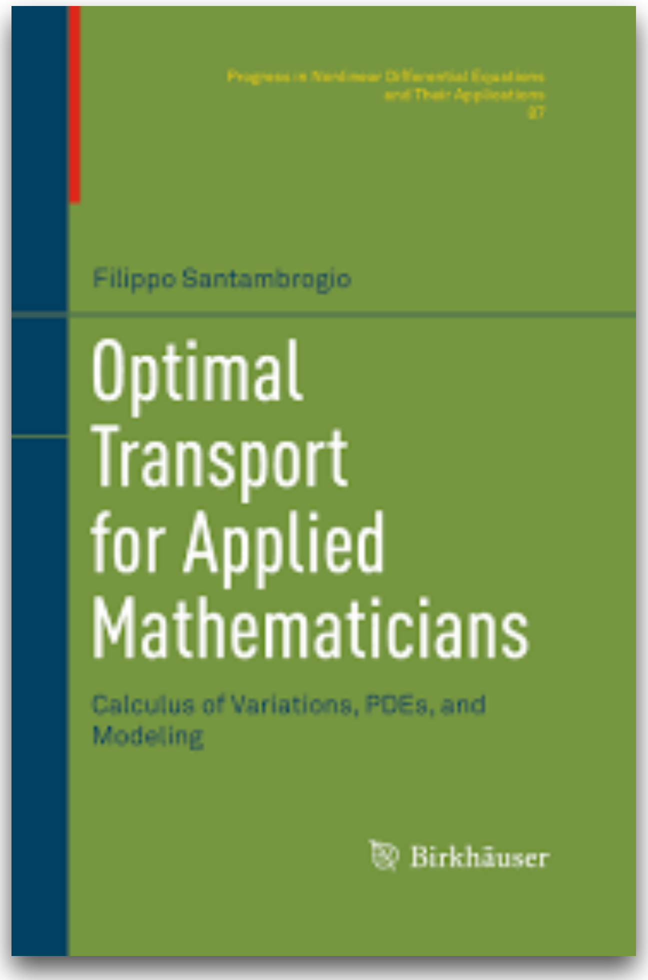


Further Reading…

2. Computations

$$\mathrm{W}_p(\mu,\nu) := \left(\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d^p(x,y)\,\mathrm{d}\pi(x,y) \right)^{1/p}$$





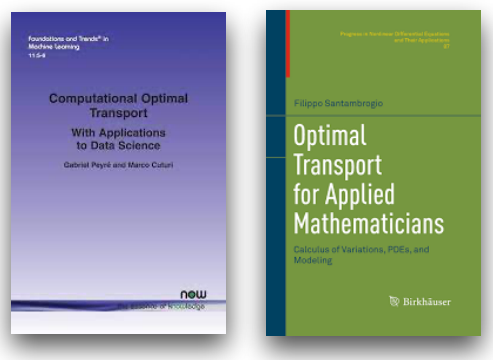
Gabriel Peyré & Marco Cuturi. Computational Optimal Transport: With Applications to Data Science. *Foundations and Trends in Machine Learning,* 2019

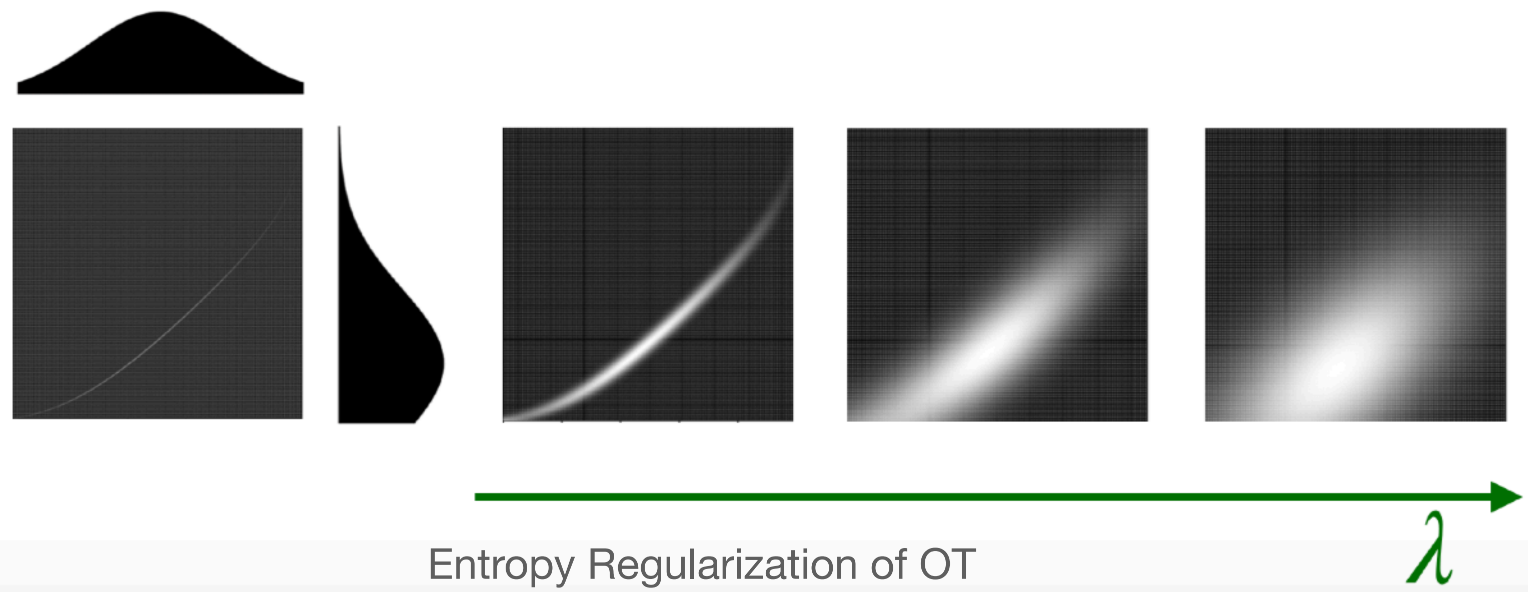
Filippo Santambrogio. Optimal Transport for Applied Mathematicians. *Birkhäuser,* 2015

Further Reading...

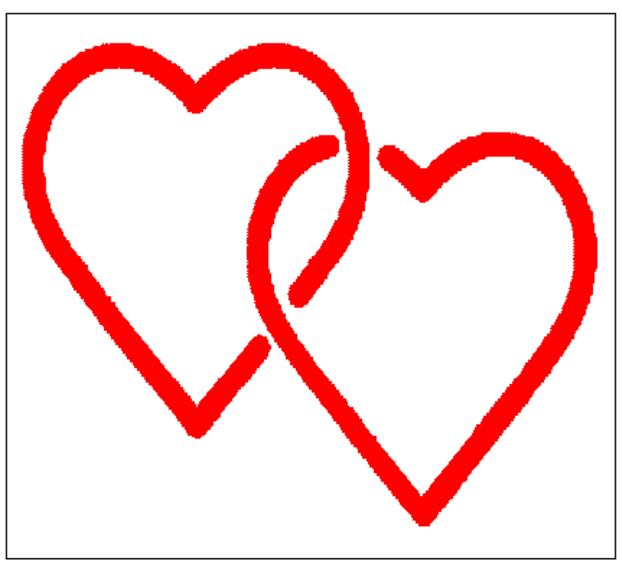
2. Computations

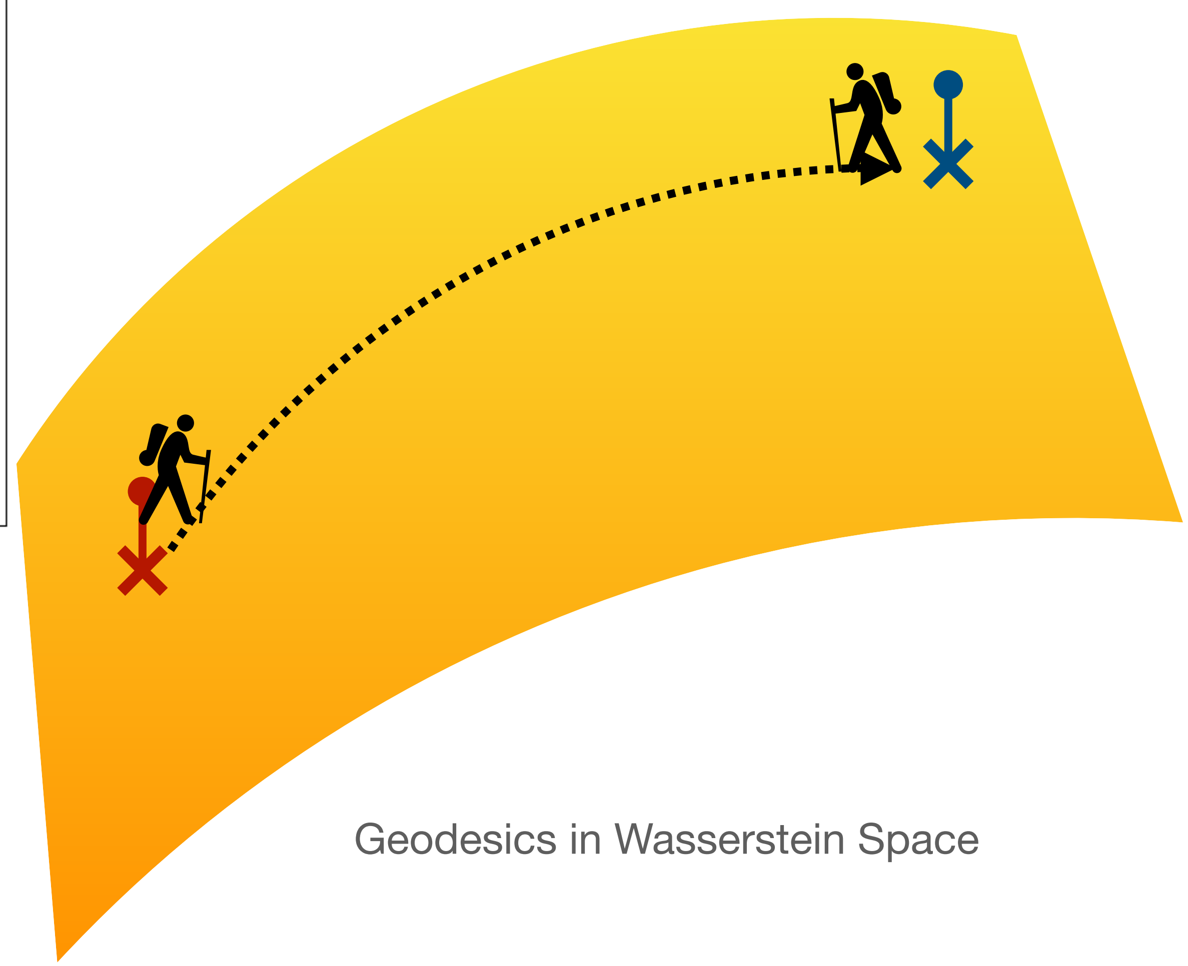
$$\mathrm{W}_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d^p(x,y)\, \mathrm{d}\pi(x,y) \right)^{1/p}$$





Entropy Regularization of OT



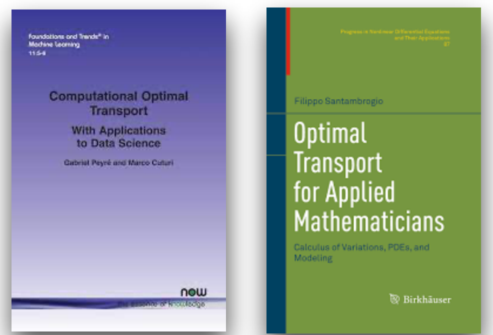


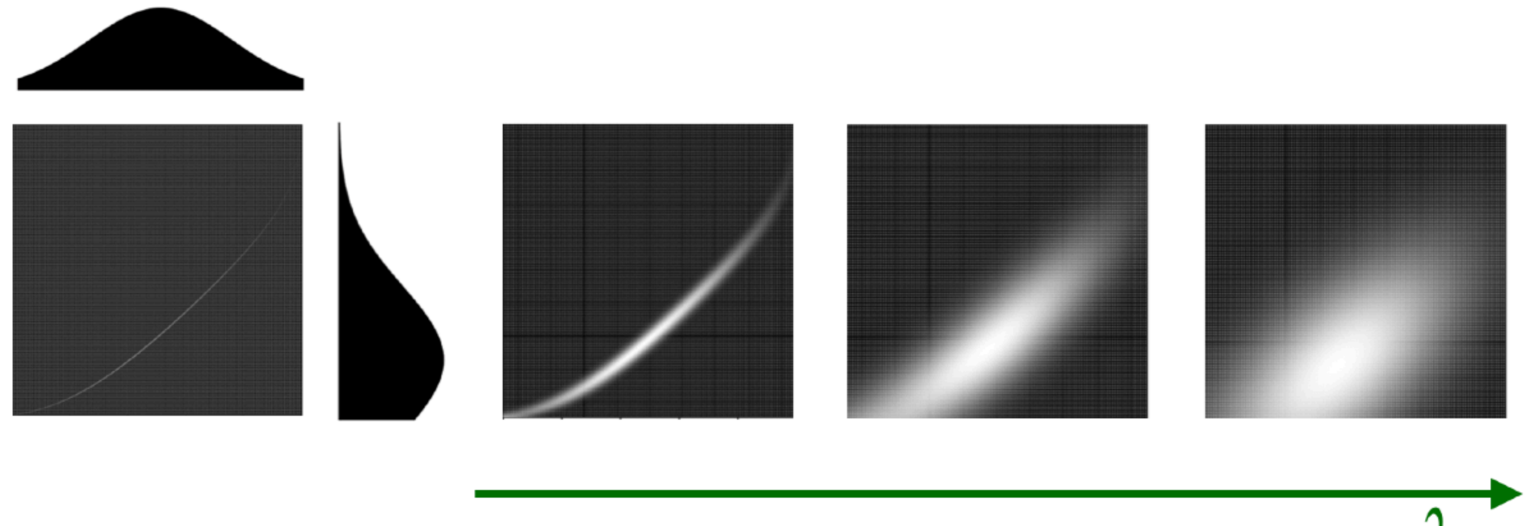
Geodesics in Wasserstein Space

Further Reading…

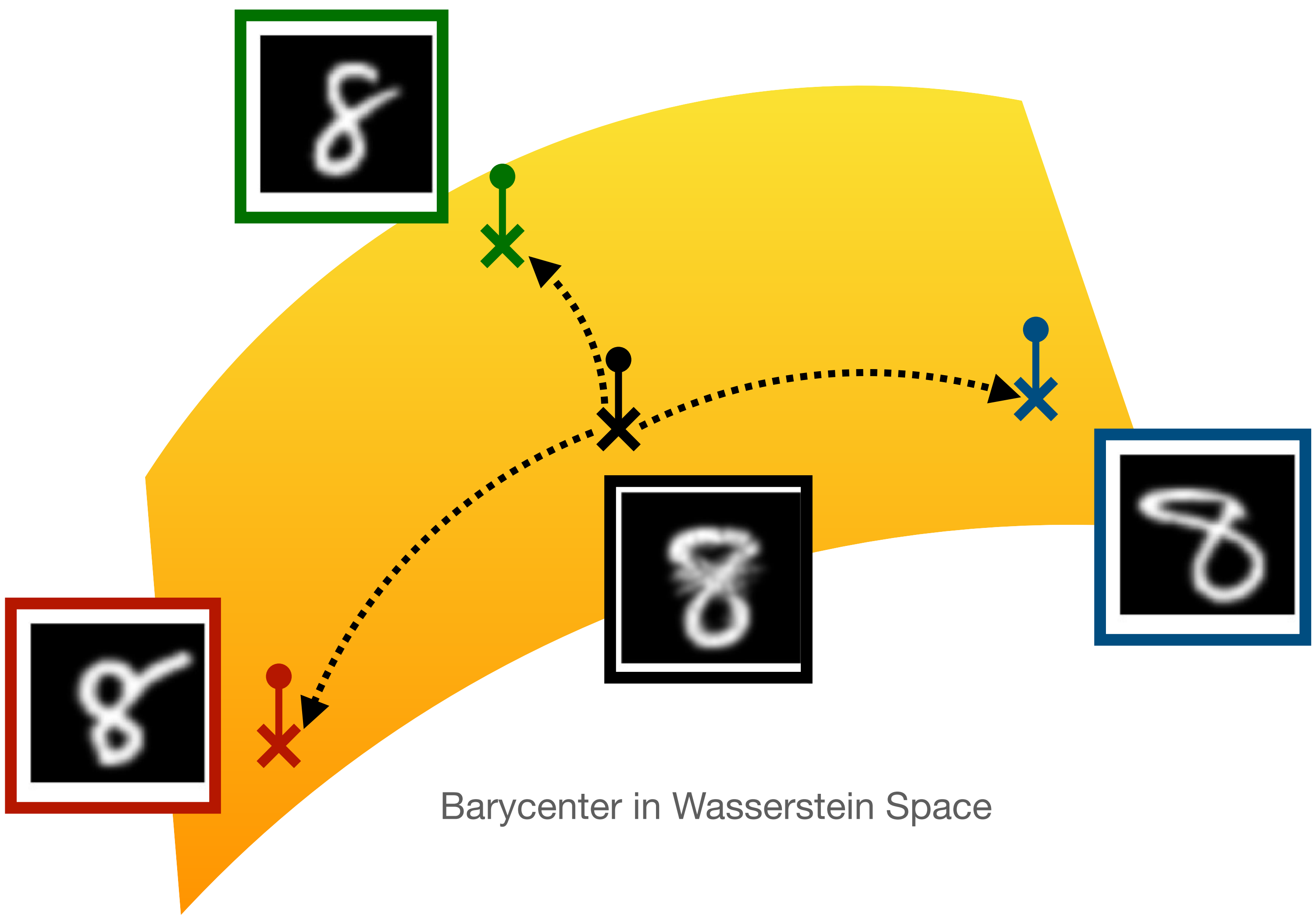
2. Computations

$$W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d^p(x,y)\, \mathrm{d}\pi(x,y) \right)^{1/p}$$





Entropy Regularization of OT

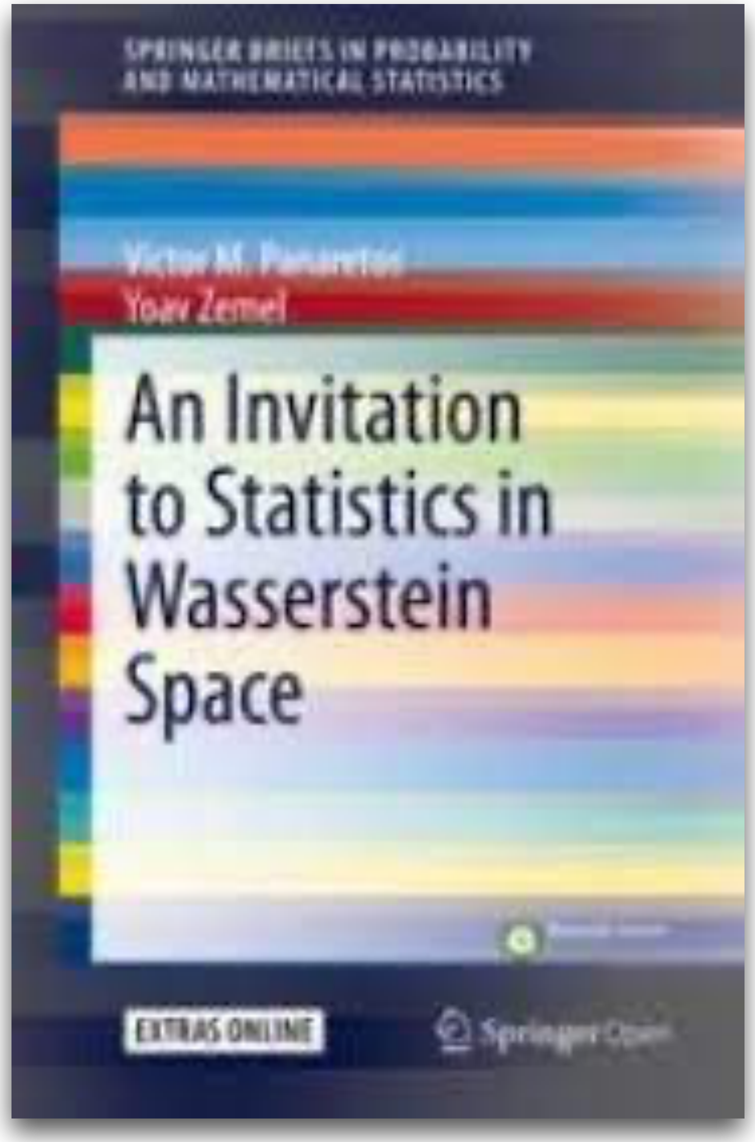


Barycenter in Wasserstein Space

Further Reading…

$$W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d^p(x,y)\, \mathrm{d}\pi(x,y) \right)^{1/p}$$

3. Statistics

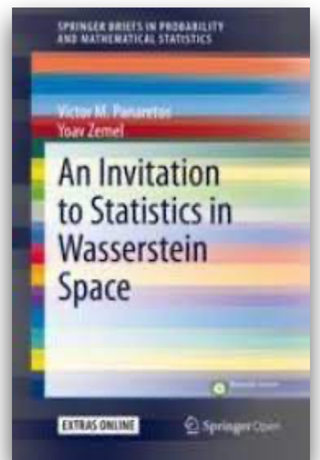


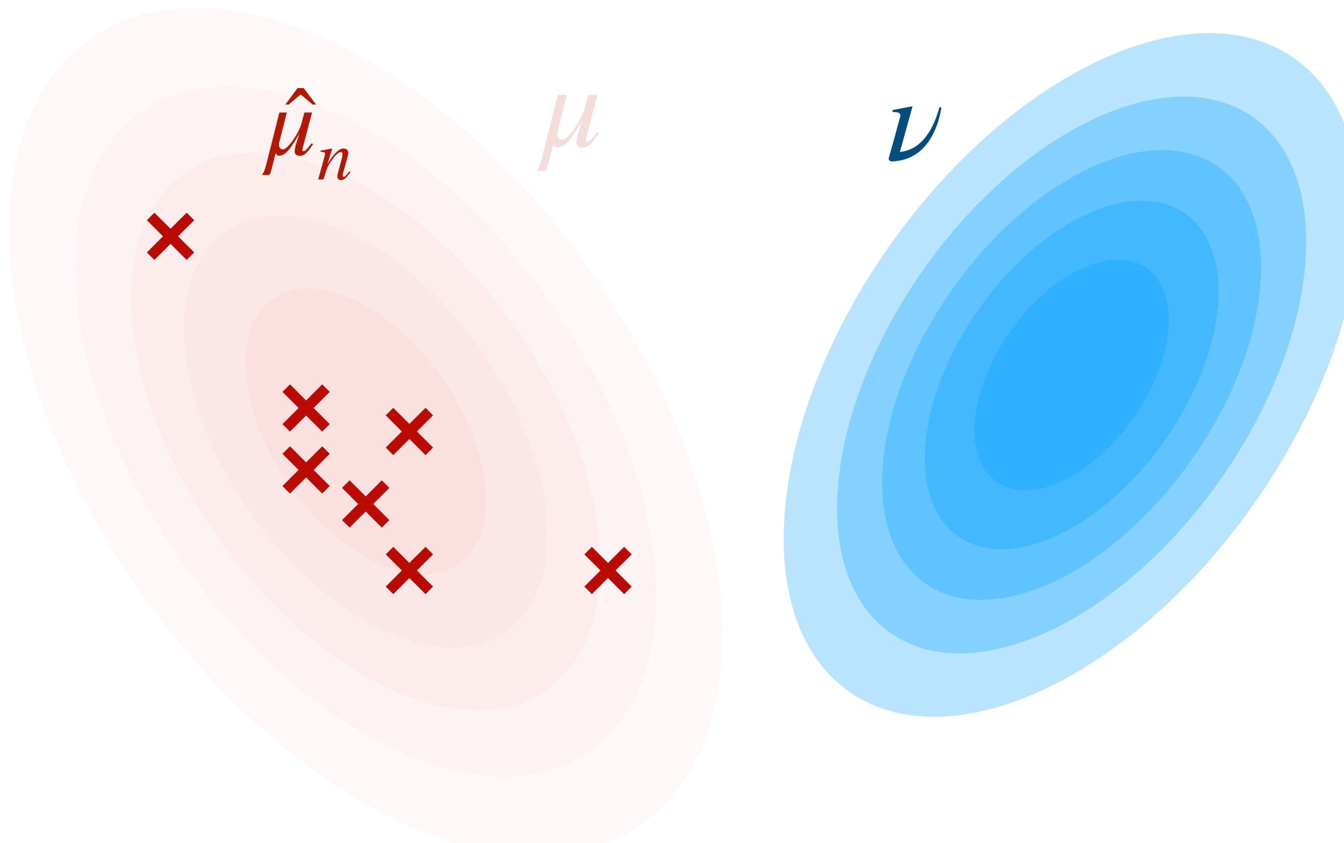
Victor Panaretos & Yoav Zemel. An Invitation to Statistics in Wasserstein Space. *Springer Nature,* 2020

Further Reading...

$$\mathrm{W}_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d^p(x, y)\, \mathrm{d}\pi(x, y) \right)^{1/p}$$

3. Statistics





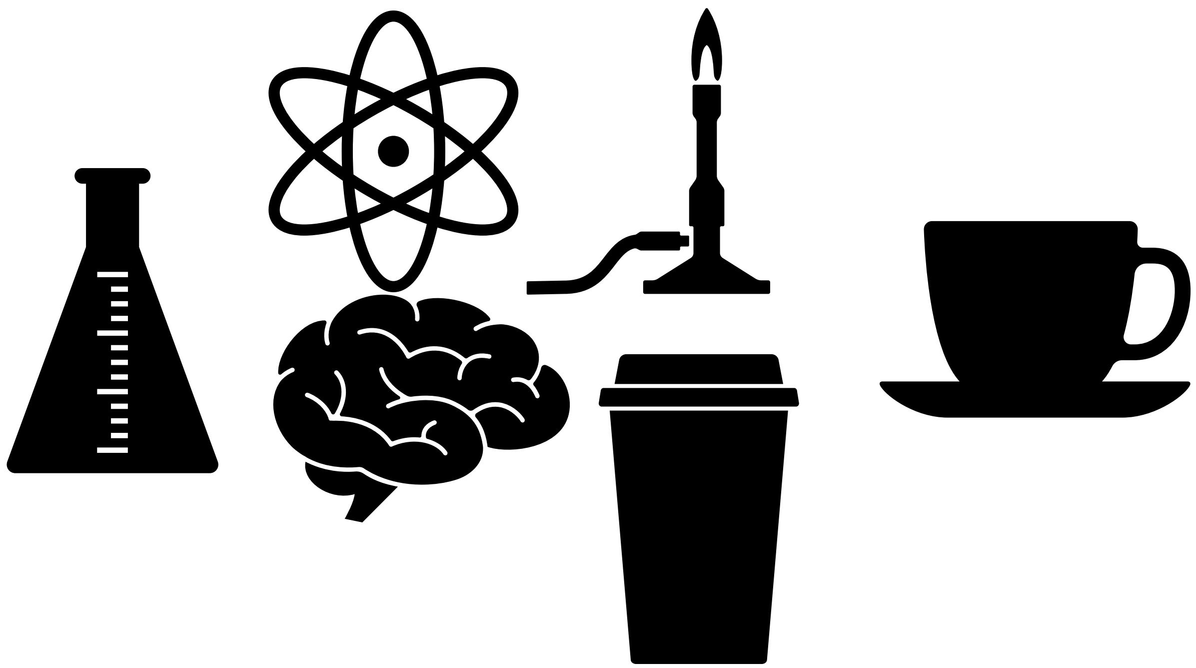
$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mu$$

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

$$\mathrm{W}_p^p(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d^p(x, y)\, \mathrm{d}\pi(x, y)$$

↑ ??

$$\mathrm{W}_p^p(\hat{\mu}_n, \nu) = \inf_{\pi \in \Pi(\hat{\mu}_n,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d^p(x, y)\, \mathrm{d}\pi(x, y)$$



… Sunday
(Zi-4033)

Do you have no idea what to do now that corona measurements have eased? Have you always been interested in dancing? Are you planning a trip to an exotic land or just want to impress your partner? Join us for a **FREE** introduction to Cuban Salsa with Marcel from Cuba. Don't let this opportunity miss you and sign up for this fun event that P-NUT has brought you. Notice that you don't need to have a partner beforehand.