Numerical Optimal Transport

http://optimaltransport.github.io

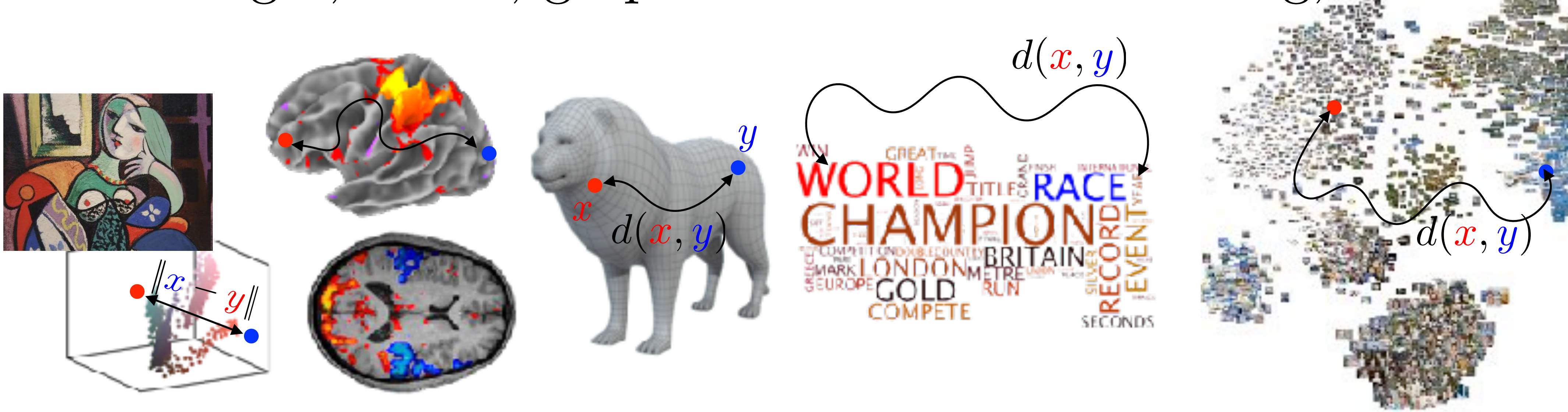
Theoretical Foundations

Gabriel Peyré

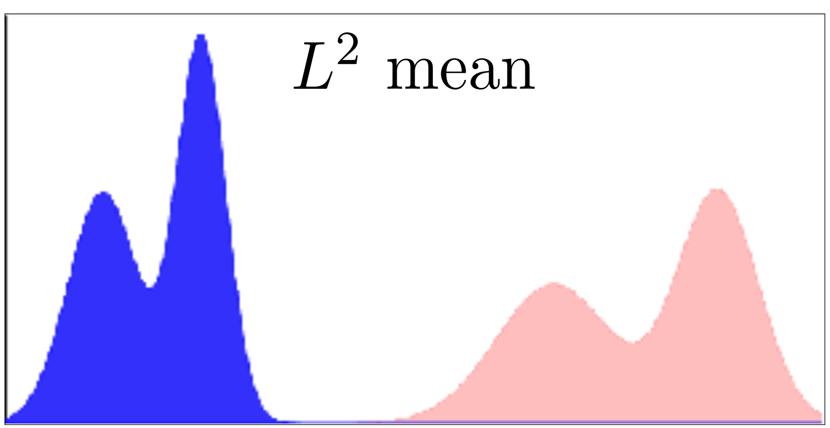
www.numerical-tours.com

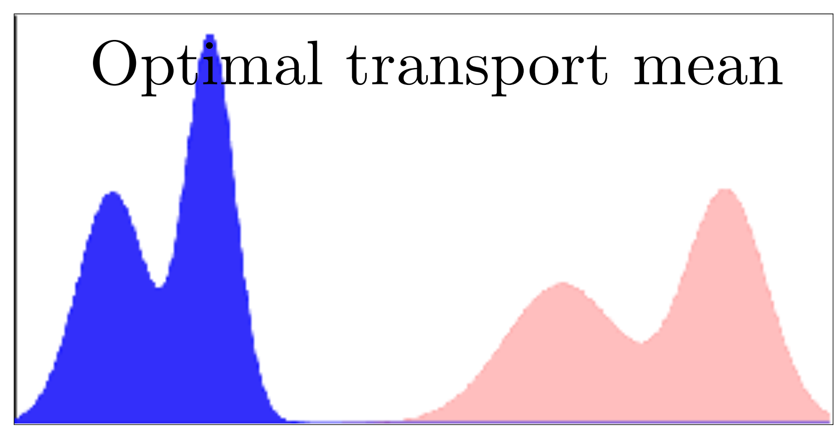
Comparing Measures

→ images, vision, graphics and machine learning, . . .

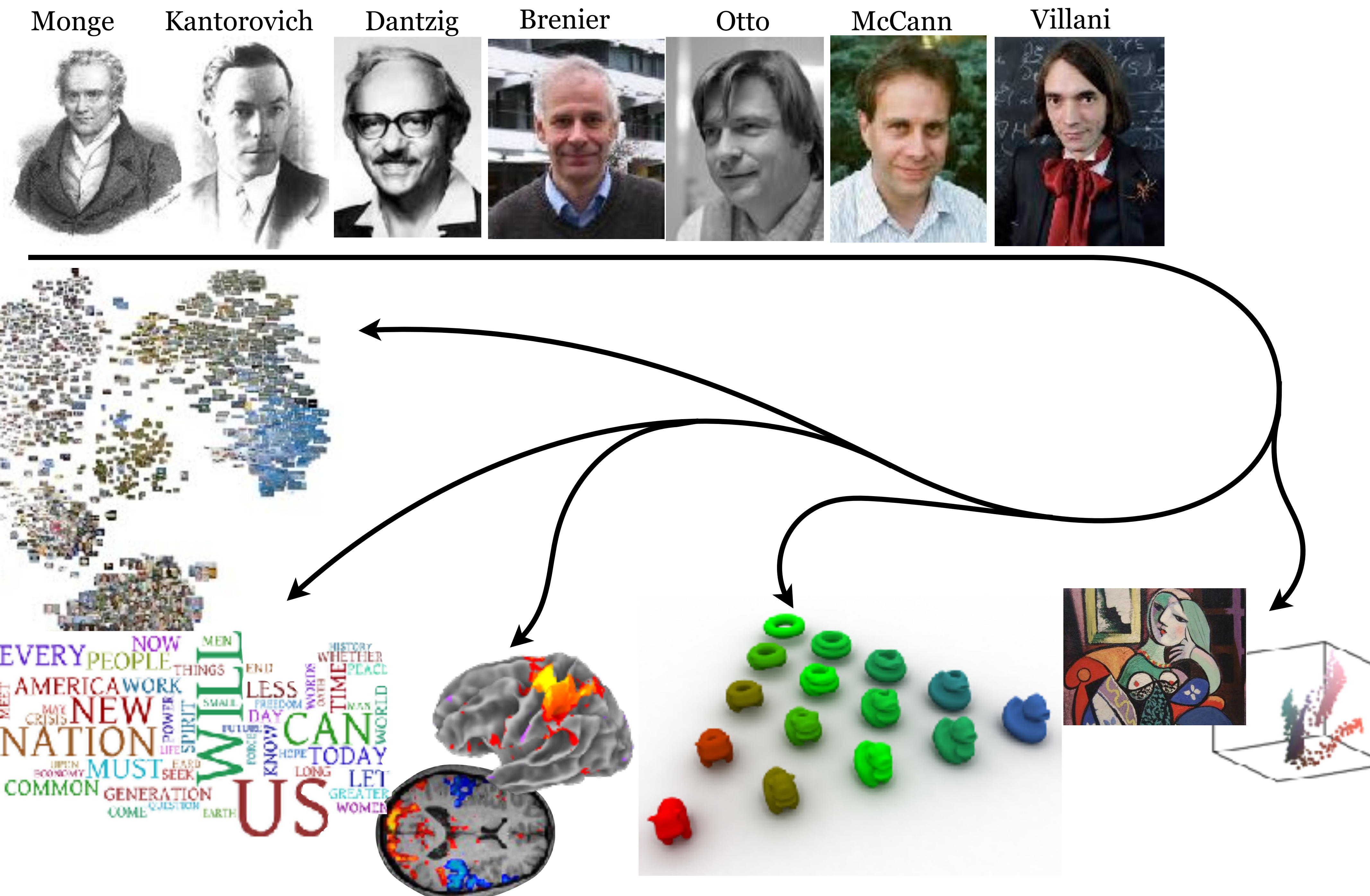


- *Optimal transport*
  → takes into account a metric d.





Toward High-dimensional OT



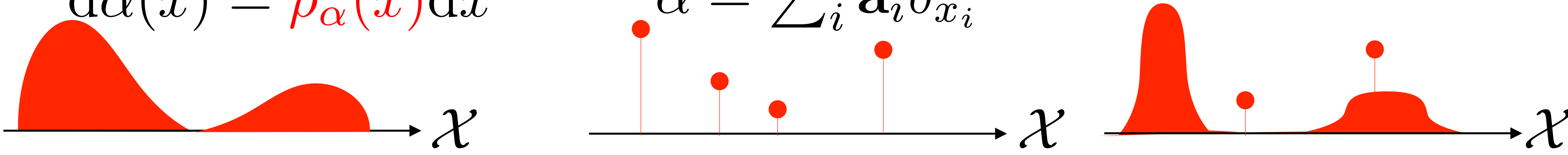
Overview

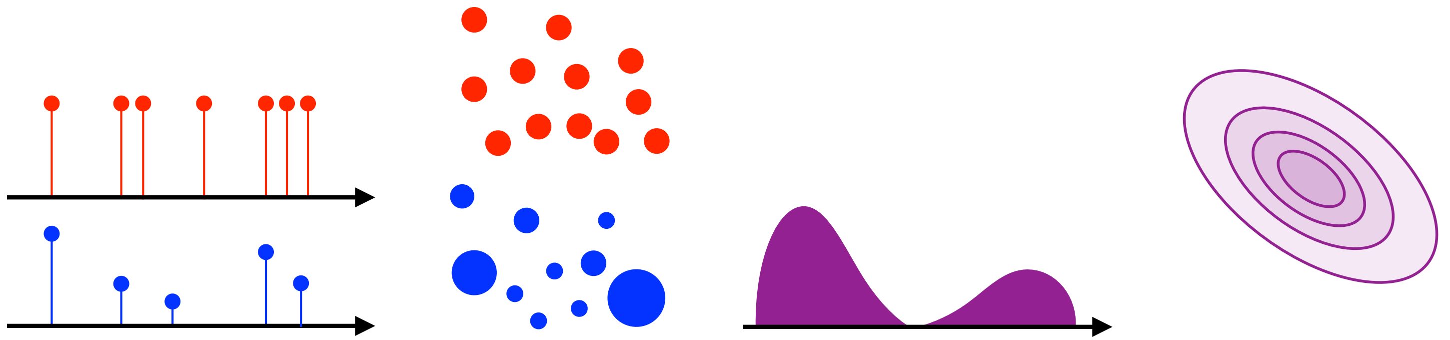
- **Measures and Histograms**
- From Monge to Kantorovitch Formulations
- Special Cases

Probability Measures

Positive Radon measure α on a metric space $\mathcal{X}$.

$d\alpha(x) = {\color{red}\rho_\alpha(x)}dx$ $\qquad$ $\alpha = \sum_i \mathbf{a}_i \delta_{x_i}$





Discrete $d = 1$ $\quad$ Discrete $d = 2$ $\quad$ Density $d = 1$ $\quad$ Density $d = 2$

Measure of sets $A \subset \mathcal{X}$: $\alpha(A) = \int_A d\alpha(x) \geqslant 0$

Integration against continuous functions: $\int_{\mathcal{X}} g(x) d\alpha(x) \geqslant 0$

$$d\alpha(x) = {\color{red}\rho_\alpha(x)}dx \longrightarrow \int_{\mathcal{X}} g d\alpha = \int_{\mathcal{X}} \rho_\alpha(x) dx$$

$$\alpha = \sum_i \mathbf{a}_i \delta_{x_i} \longrightarrow \int_{\mathcal{X}} g d\alpha = \sum_i \mathbf{a}_i g(x_i)$$

Probability (normalized) measure: $\alpha(\mathcal{X}) = \int_{\mathcal{X}} d\alpha(x) = 1$

Measures and Random Variables

Random vectors

$$\mathbb{P}(X \in A)$$

Convergence in law:

$\forall$ set A

$$\mathbb{P}(X_n \in A) \xrightarrow{n \to +\infty} \mathbb{P}(X \in A)$$

=

Radon measures

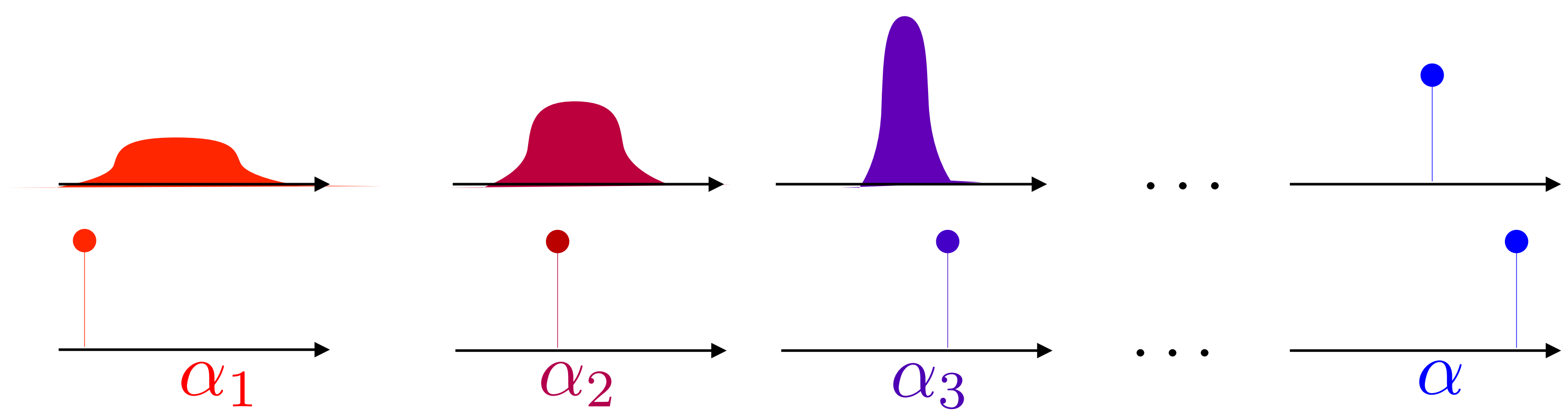
$$\int_A \mathrm{d}\alpha(x)$$

Weak* convergence:

$\forall$ continuous function f

$$\int f \mathrm{d}\alpha_n \xrightarrow{n \to +\infty} \int f \mathrm{d}\alpha$$

Weak convergence:



Convergence of Random Variables

In mean

$$\lim_{n \to +\infty} \mathbb{E}(|X_n - X|^p) = 0$$

Almost sure

$$\mathbb{P}(\lim_{n \to +\infty} X_n = X) = 1$$

$\Downarrow$ (from In mean and Almost sure)

In probability

$$\forall \varepsilon > 0,\ \mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow{n \to +\infty} 0$$

$\Downarrow$

In law

$$\mathbb{P}(X_n \in A) \xrightarrow{n \to +\infty} \mathbb{P}(X \in A)$$

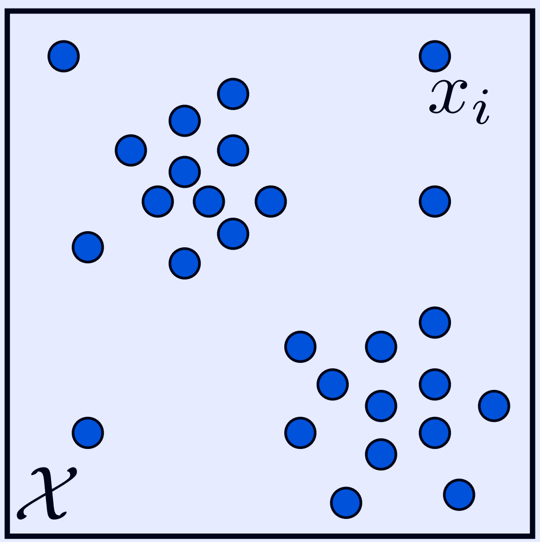
(the X_n can be defined on different spaces)

Lagrangian vs. Eulerian Discretization

Discrete measure: $\alpha = \sum_{i=1}^{n} \mathbf{a}_i \delta_{x_i} \quad x_i \in \mathcal{X}, \quad \sum_i \mathbf{a}_i = 1$

Lagrangian (point clouds)

Constant weights $\mathbf{a}_i = \frac{1}{n}$

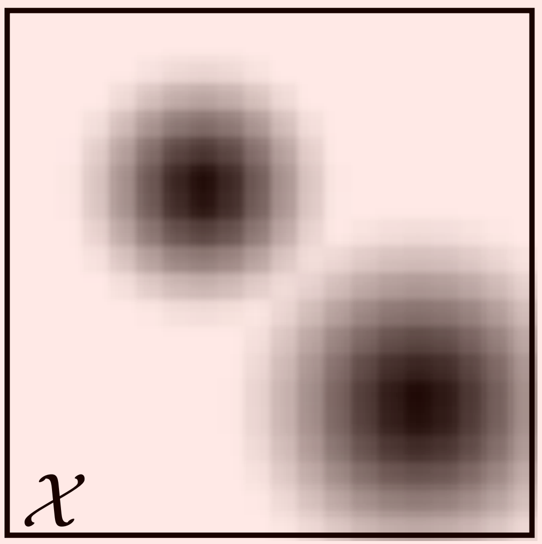


Quotient space:

$\mathcal{X}^n / \mathrm{Perm}(n)$

Eulerian (histograms)

Fixed positions x_i (e.g. grid)



Convex polytope (simplex):

$\{(\mathbf{a}_i)_i \geqslant 0 \; ; \; \sum_i \mathbf{a}_i = 1\}$

Push Forward

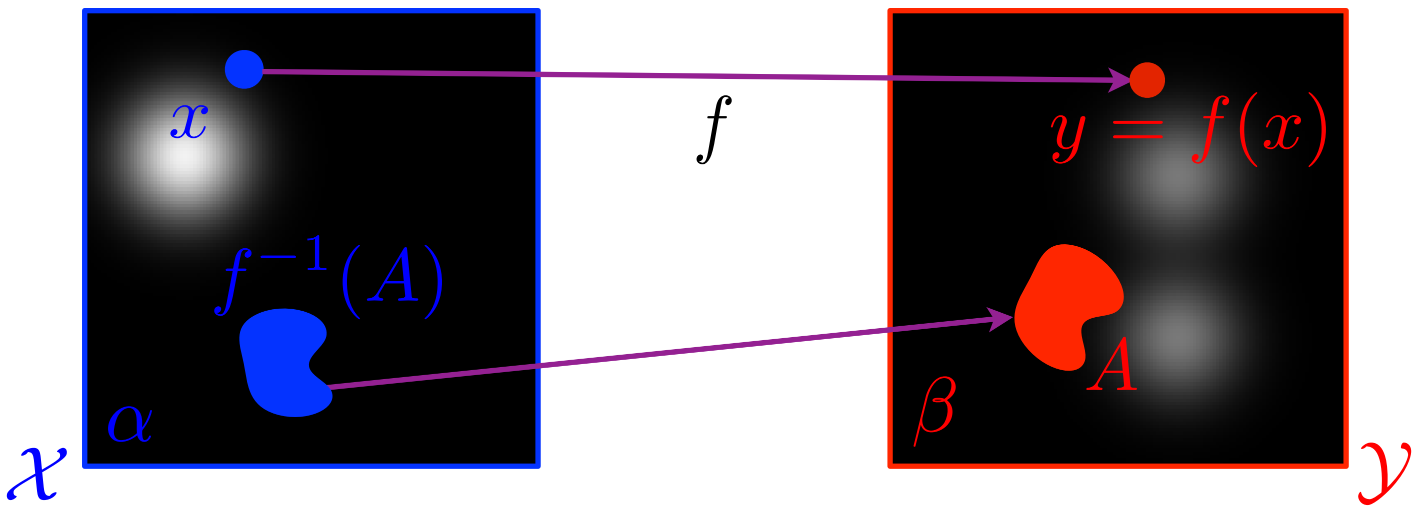
Radon measures (α, β) on $(\mathcal{X}, \mathcal{Y})$.

Transfer of measure by $f : \mathcal{X} \to \mathcal{Y}$: *push forward.*

$\beta = f_\sharp \alpha$ defined by:

$$\beta(A) \overset{\text{def.}}{=} \alpha(f^{-1}(A))$$

$$\Longleftrightarrow \int_{\mathcal{Y}} g(y) \mathrm{d}\beta(y) \overset{\text{def.}}{=} \int_{\mathcal{X}} g(f(x)) \mathrm{d}\alpha(x)$$

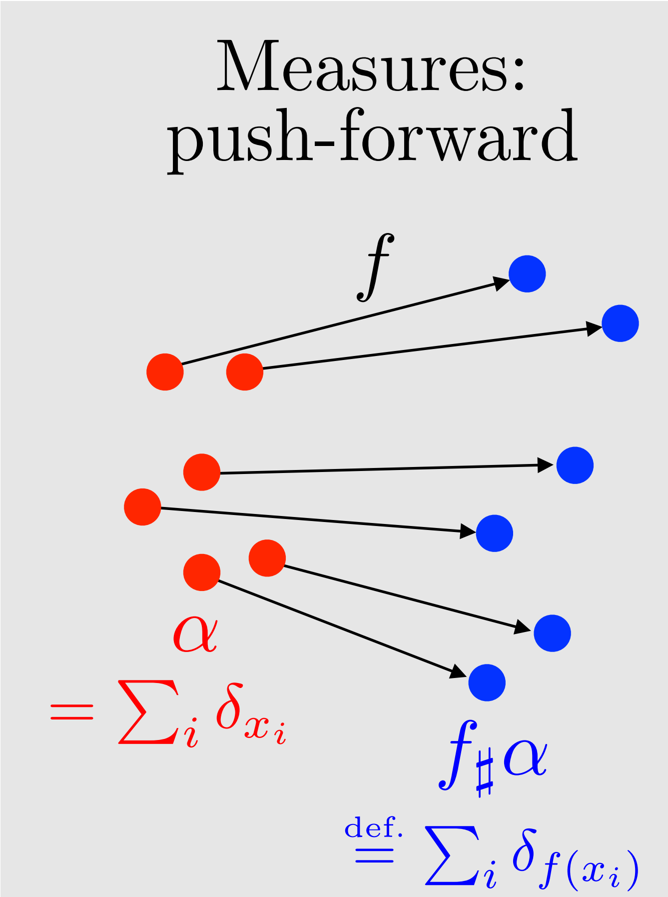


Smooth densities: $\mathrm{d}\alpha = \rho(x)\mathrm{d}x$, $\mathrm{d}\beta = \xi(x)\mathrm{d}x$

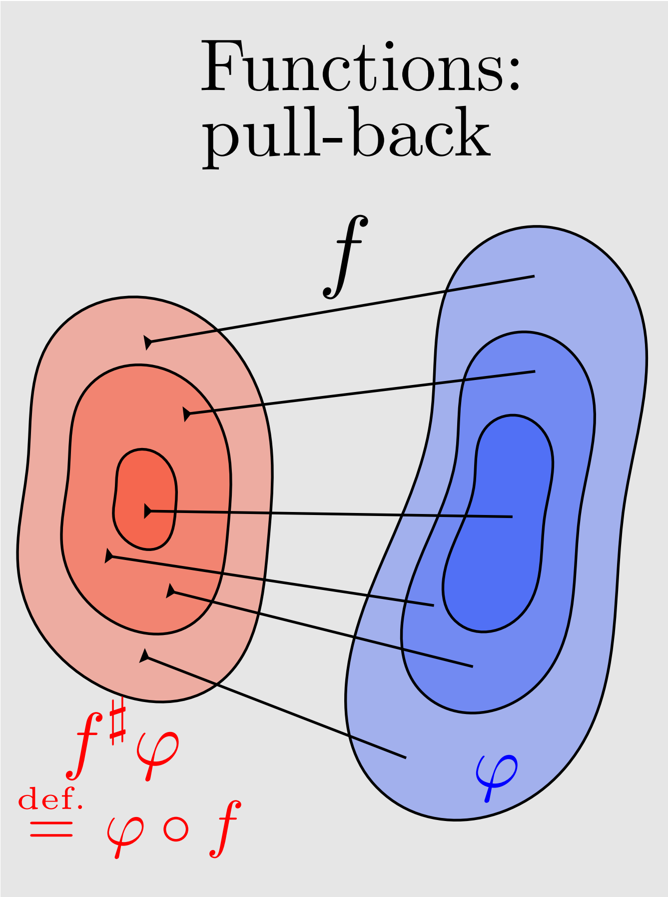
$$f_\sharp \alpha = \beta \Longleftrightarrow \rho(f(x))|\det(\partial f(x))| = \xi(x)$$

Push-forward vs. Pull-back

$f : \mathcal{X} \to \mathcal{Y}$



$f_\sharp : \mathcal{M}(\mathcal{X}) \to \mathcal{M}(\mathcal{Y})$



$f^\sharp : \mathcal{C}(\mathcal{Y}) \to \mathcal{C}(\mathcal{X})$

Remark: $f^\sharp$ and $f_\sharp$ are adjoints

$$\int_{\mathcal{Y}} \varphi \mathrm{d}(f_\sharp \alpha) = \int_{\mathcal{X}} (f^\sharp \varphi) \mathrm{d}\alpha$$

Overview

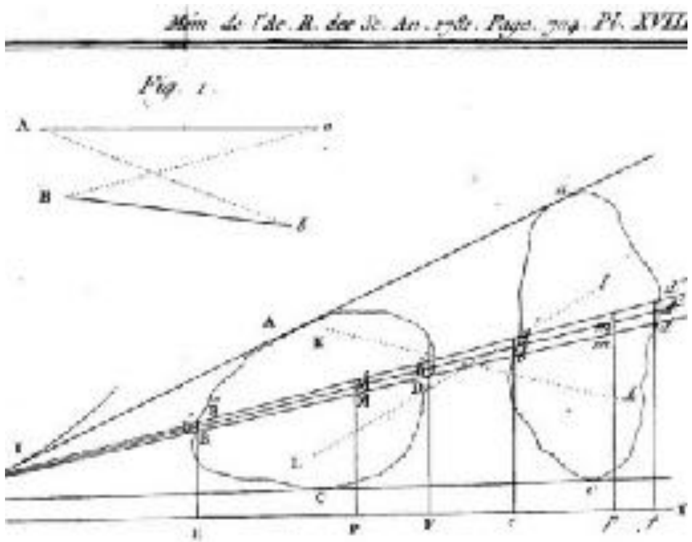
- Measures and Histograms
- **From Monge to Kantorovitch Formulations**
- Special Cases

Gaspard Monge (1746-1818)

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.

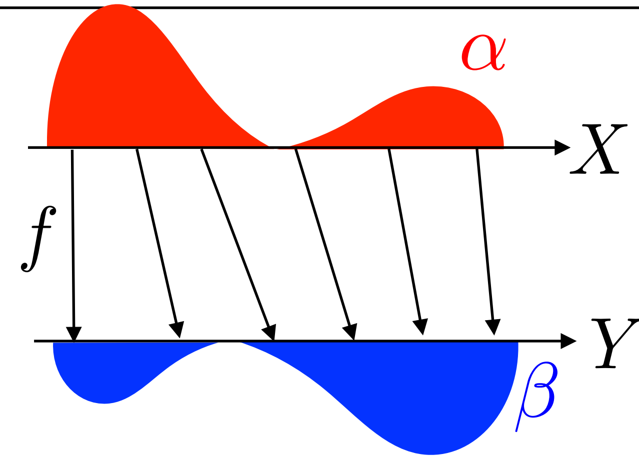
Le prix du tranſport d'une molécule étant, toutes choſes d'ailleurs égales, proportionnel à ſon poids & à l'eſpace qu'on lui fait parcourir, & par conſéquent le prix du tranſport total devant être proportionnel à la ſomme des produits des molécules multipliées chacune par l'eſpace parcouru, il s'enſuit que le déblai & le remblai étant donnés de figure & de poſition, il n'eſt pas indifférent que telle molécule du déblai ſoit tranſportée dans tel ou tel autre endroit du remblai, mais qu'il y a une certaine diſtribution à faire des molécules du premier dans le ſecond, d'après laquelle la ſomme de ces produits ſera la moindre poſſible, & le prix du tranſport total ſera un *minimum*.



(1784)

Monge's Transport

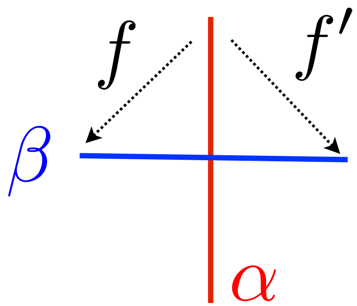
$$\min_{\beta = f_\sharp \alpha} \int_X c(x, f(x)) \mathrm{d}\alpha(x)$$

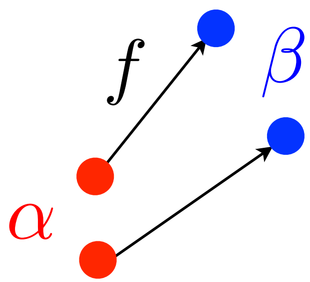


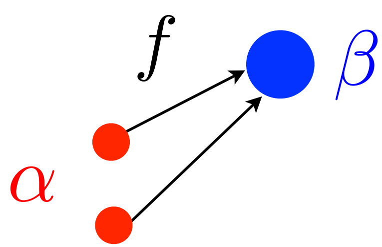
Theorem: [Brenier] for $c(x,y) = \|x-y\|^2$, (α, β) with density, there exists a unique optimal f. One has $f = \nabla\psi$ where ψ is the unique convex function such that $(\nabla\psi)_\sharp \alpha = \beta$

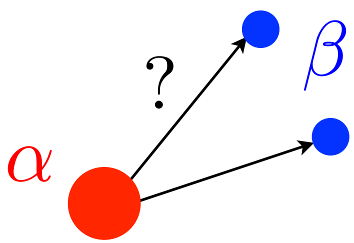
Monge-Ampère equation: $\rho(\nabla\psi) \det(\partial^2\psi) = \xi$

Non-uniqueness / non-existence:



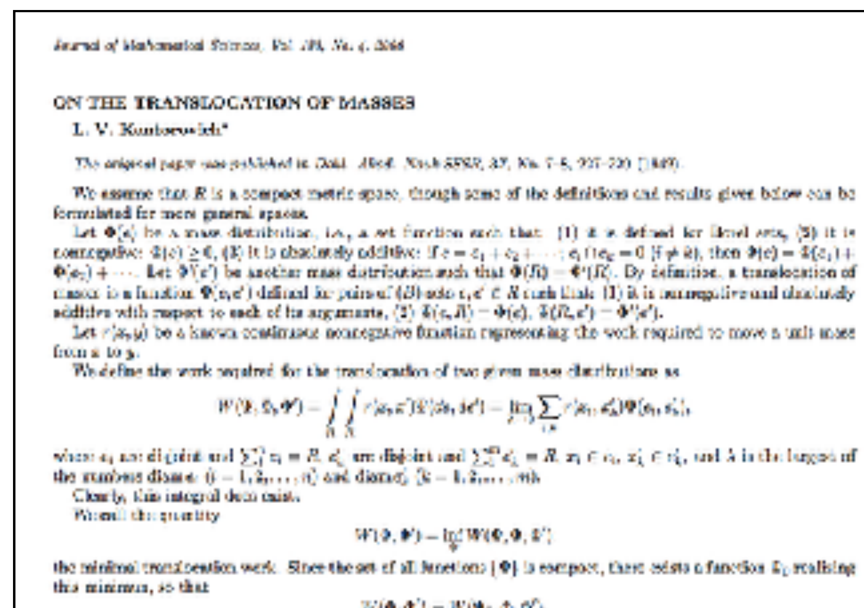






Leonid Kantorovich (1912-1986)

Леонид Витальевич Канторович



ON THE TRANSLOCATION OF MASSES

L. V. Kantorovich*

We assume that R is a compact metric space, though some of the definitions and results given below can be formulated for more general spaces.

Let $\Phi(e)$ be a mass distribution, i.e., a set function such that: (1) it is defined for Borel sets, (2) it is nonnegative: $\Phi(e) \geq 0$, (3) it is absolutely additive: if $e = e_1 + e_2 + \cdots$; $e_i \cap e_k = 0$ $(i \neq k)$, then $\Phi(e) = \Phi(e_1) + \Phi(e_2) + \cdots$. Let $\Phi'(e')$ be another mass distribution such that $\Phi(R) = \Phi'(R)$. By definition, a translocation of masses is a function $\Psi(e, e')$ defined for pairs of (B)-sets $e, e' \in R$ such that: (1) it is nonnegative and absolutely additive with respect to each of its arguments, (2) $\Psi(e, R) = \Phi(e)$, $\Psi(R, e') = \Phi'(e')$.

Let $r(x, y)$ be a known continuous nonnegative function representing the work required to move a unit mass from x to y.

We define the work required for the translocation of two given mass distributions as

Clearly, this integral does exist.

We call the quantity

the minimal translocation work. Since the set of all functions $\{\Psi\}$ is compact, there exists a function Ψ_0 realising this minimum, so that

Before Kantorovitch

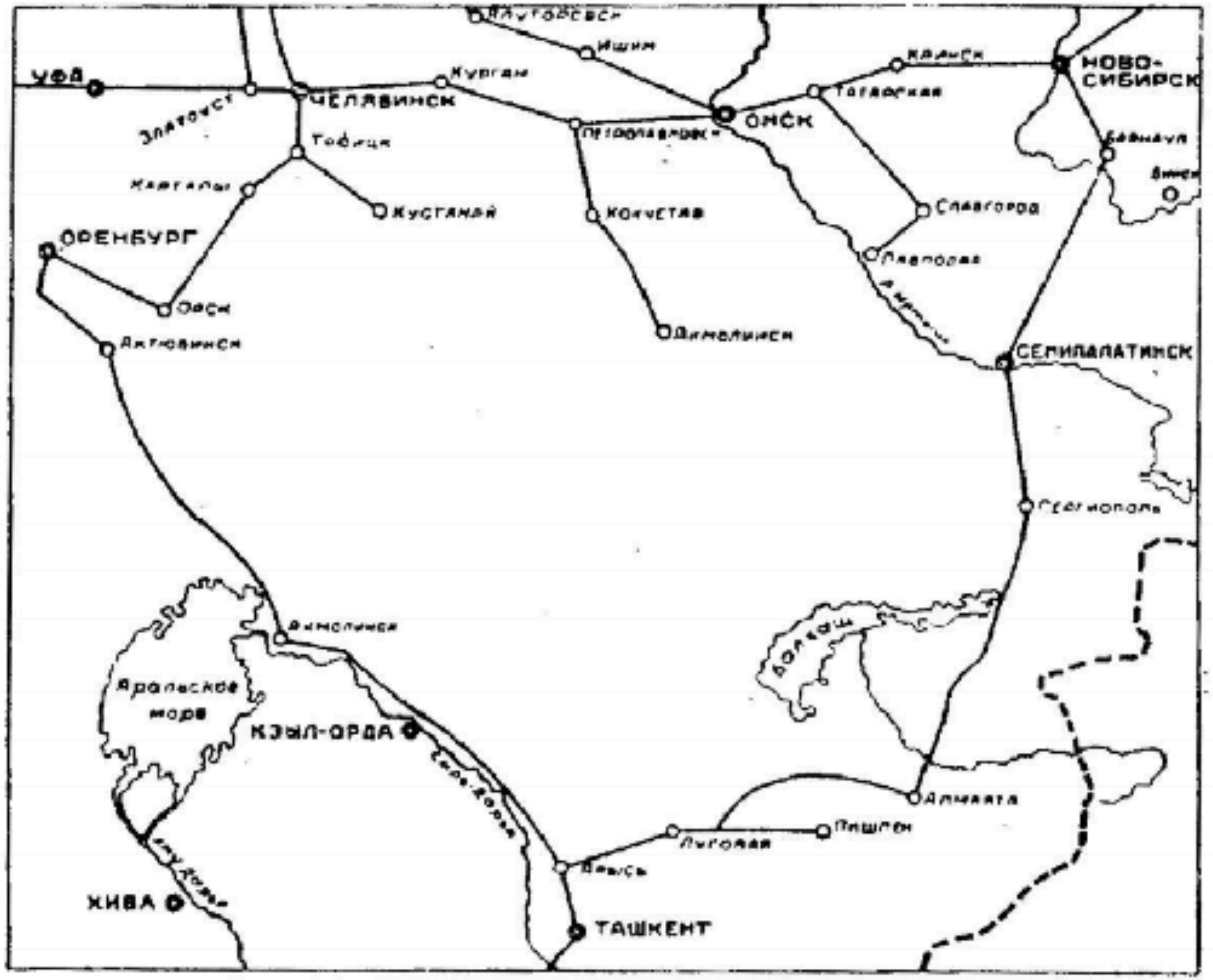


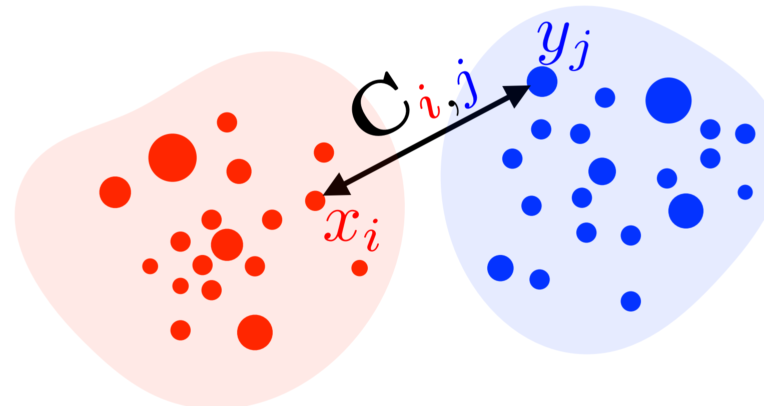
Figure 1: Figure from Tolstoĭ [1930] to illustrate a negative cycle

Optimal Transport was formulated in 1930 by A.N. Tolstoi, 12 years before Kantorovich. He even solved a "large scale" 10×68 instance!

Kantorovitch's Formulation

Input distributions

$\alpha = \sum_{i=1}^{n} \mathbf{a}_i \delta_{x_i} \quad \beta = \sum_{j=1}^{m} \mathbf{b}_j \delta_{y_j}$



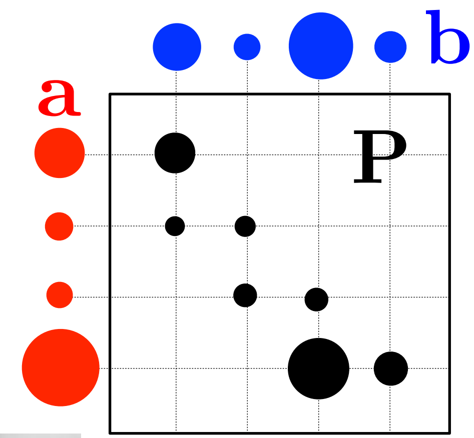
Points $(x_i)_i$, $(y_j)_j$

Weights $\mathbf{a}_i \geqslant 0$, $\mathbf{b}_j \geqslant 0$.

$\sum_{i=1}^{n} \mathbf{a}_i = \sum_{j=1}^{m} \mathbf{b}_j = 1$

Cost:

$\mathbf{C}_{i,j} = c(x_i, y_j)$

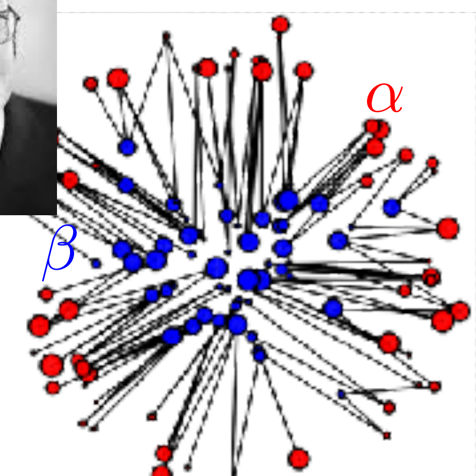


Couplings:

$$U(\mathbf{a},\mathbf{b}) \overset{\text{def.}}{=} \left\{ \mathbf{P} \in \mathbb{R}_+^{n \times m} \;;\; \mathbf{P}\mathbb{1}_n = \mathbf{a}, \mathbf{P}^\top \mathbb{1}_m = \mathbf{b} \right\}$$

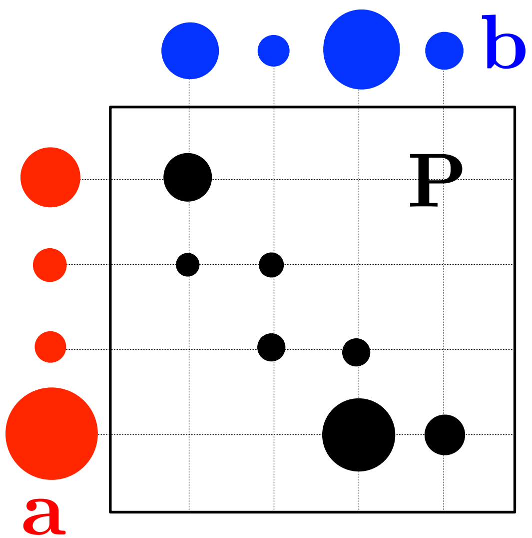
[Kantorovich 1942]

$$L_{\mathbf{C}}(\mathbf{a},\mathbf{b}) \overset{\text{def.}}{=} \min \left\{ \sum_{i,j} \mathbf{P}_{i,j} \mathbf{C}_{i,j} \;;\; \mathbf{P} \in U(\mathbf{a},\mathbf{b}) \right\}$$

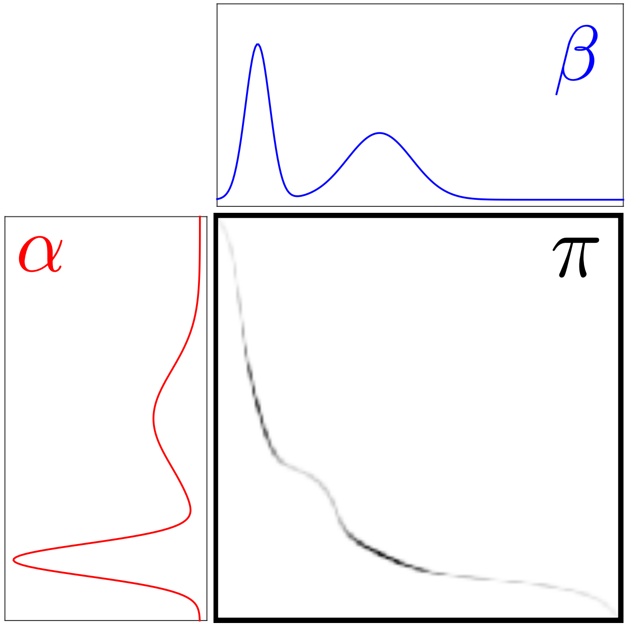


Couplings Between General Measures

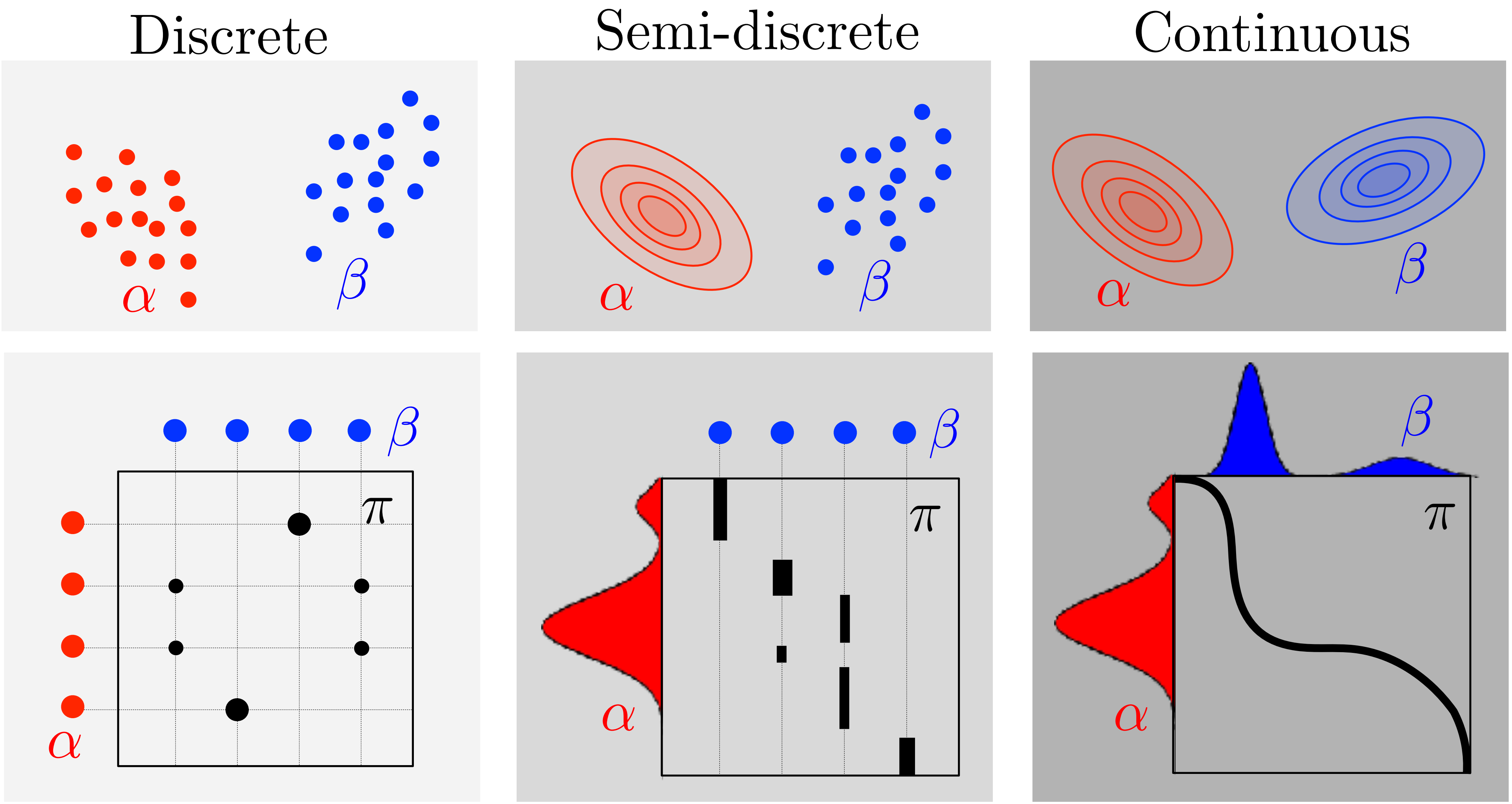
Couplings: $\mathcal{U}(\alpha,\beta) \overset{\text{def.}}{=} \left\{\pi \in \mathcal{M}_+^1(\mathcal{X}\times\mathcal{Y}) \;:\; P_{\mathcal{X}\sharp}\pi = \alpha \quad \text{and} \quad P_{\mathcal{Y}\sharp}\pi = \beta\right\}$



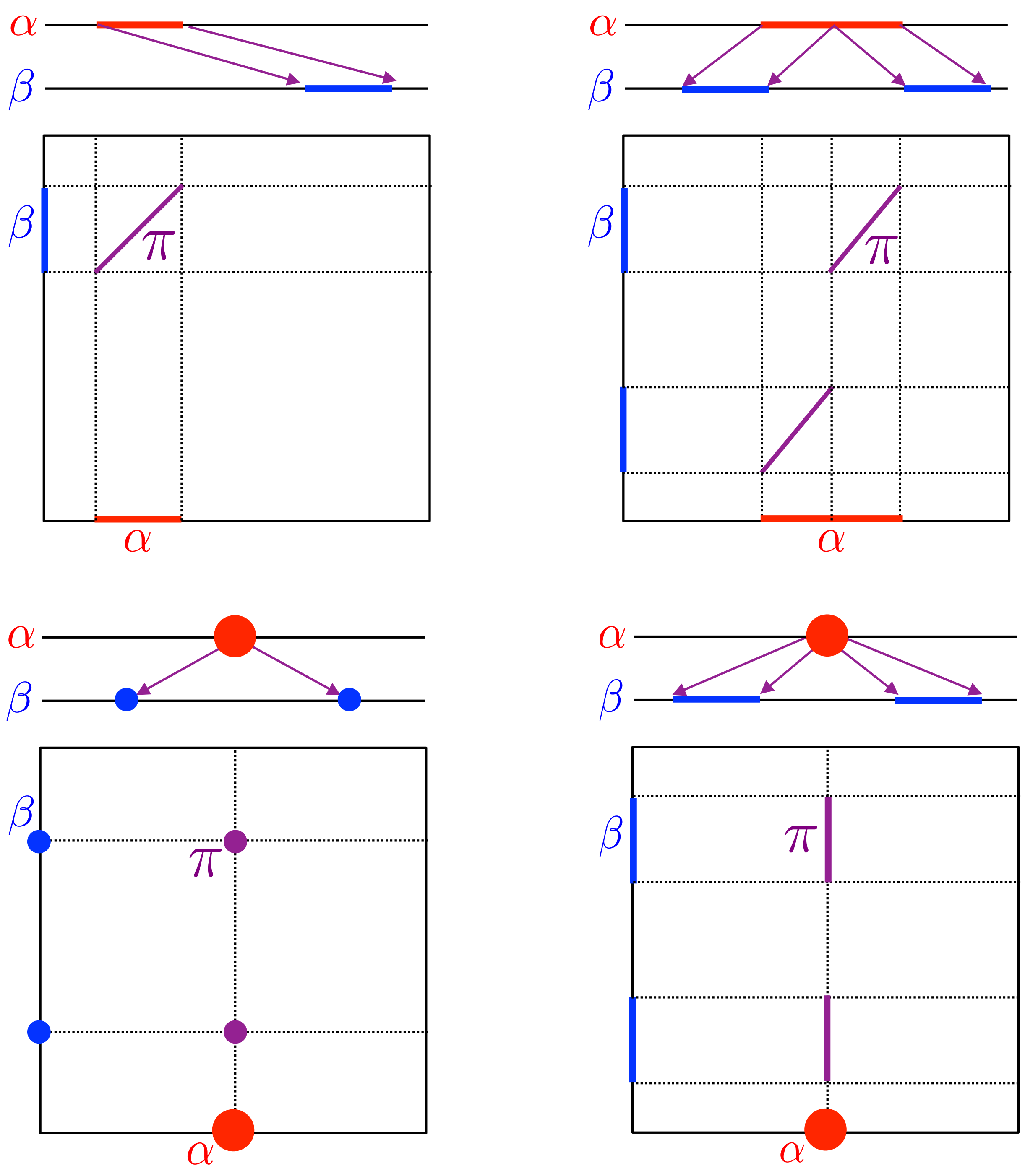
$$\pi = \sum_{i,j} \mathbf{P}_{i,j}\delta_{x_i,y_j}$$



Couplings: the 3 Settings



Examples of Couplings



Kantorovitch Problem for General Measures

Optimal transport: [Kantorovitch 1942]

$$\mathcal{L}_c(\alpha, \beta) \stackrel{\text{def.}}{=} \min_{\pi \in \mathcal{U}(\alpha,\beta)} \int_{\mathcal{X} \times \mathcal{Y}} c(x,y) \mathrm{d}\pi(x,y)$$

Probabilistic interpretation:

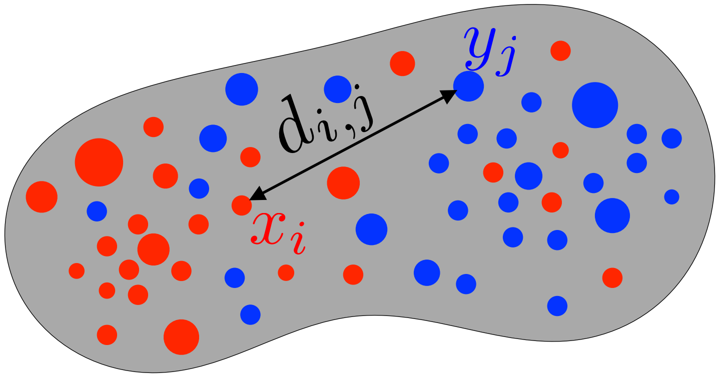
$$\min_{(X,Y)} \left\{ \mathbb{E}_{(X,Y)}(c(X,Y)) \ : \ X \sim \alpha, Y \sim \beta \right\}$$

Wasserstein Distance

Metric spaces $\mathcal{X} = \mathcal{Y}$

Distance $d(x, y)$.

Cost $c(x, y) = d(x, y)^p$, $p \geqslant 1$.



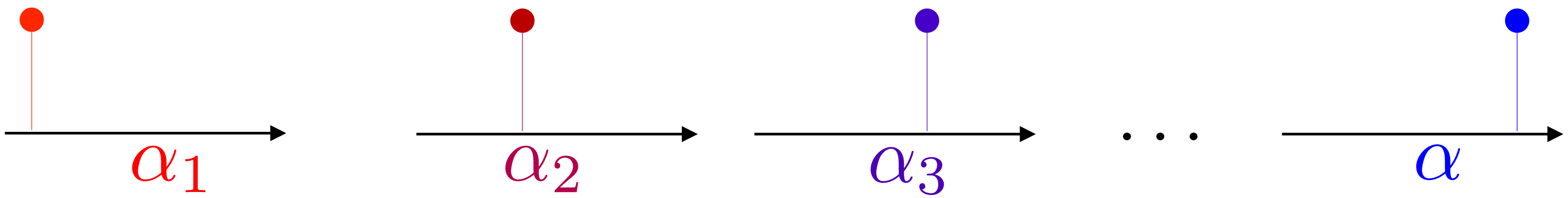
Wasserstein distance:

$$\mathrm{W}_p(\mathbf{a}, \mathbf{b}) \overset{\text{def.}}{=} \mathrm{L}_{\mathbf{D}^p}(\mathbf{a}, \mathbf{b})^{1/p} \qquad \mathcal{W}_p(\alpha, \beta) \overset{\text{def.}}{=} \mathcal{L}_{d^p}(\alpha, \beta)^{1/p}$$

Theorem: W_p and $\mathcal{W}_p$ are distances.

$$\mathcal{W}_p(\alpha_n, \alpha) \to 0 \iff \alpha_n \overset{\text{weak}*}{\to} \alpha$$

Examples: $\mathcal{W}_p(\delta_x, \delta_y) = d(x, y)$.



Overview

- Measures and Histograms
- From Monge to Kantorovitch Formulations
- Special Cases

1-D Optimal Transport

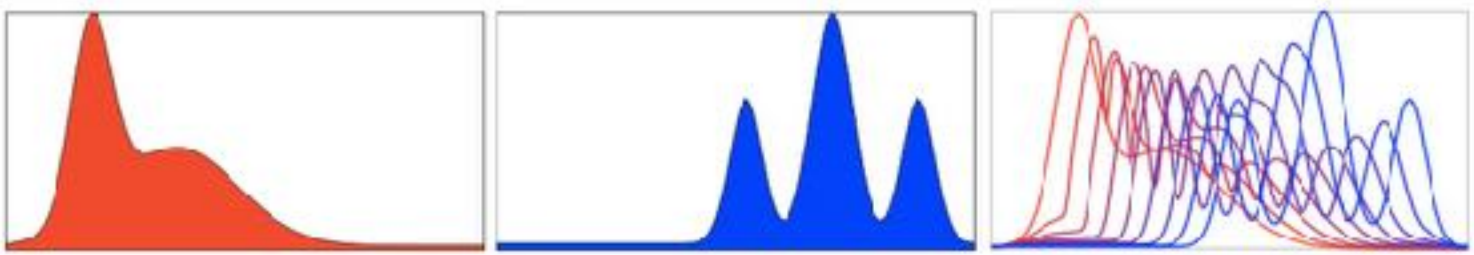
Cumulative function: $\forall\, x \in \mathbb{R}, \quad \mathcal{C}_\alpha(x) \overset{\text{def.}}{=} \int_{-\infty}^{x} \mathrm{d}\alpha,$

Inverse cumulative: $\forall\, r \in [0,1], \quad \mathcal{C}_\alpha^{-1}(r) = \min_x \; \{x \in \mathbb{R} \cup \{-\infty\} \;:\; \mathcal{C}_\alpha(x) \geq r\}$

Theorem:

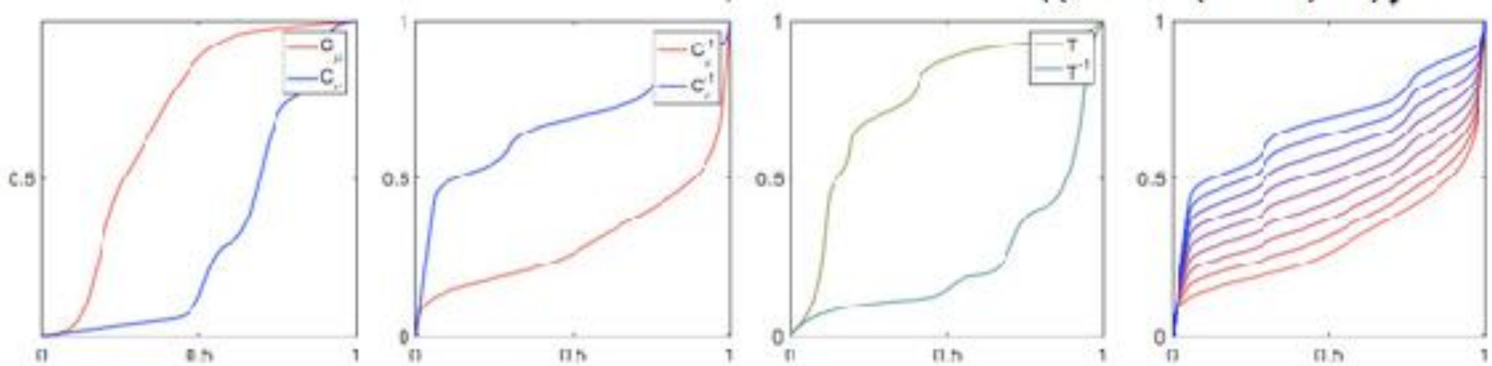
$$\mathcal{W}_p(\alpha,\beta)^p = \left\| \mathcal{C}_\alpha^{-1} - \mathcal{C}_\beta^{-1} \right\|_{L^p([0,1])}^p = \int_0^1 |\mathcal{C}_\alpha^{-1}(r) - \mathcal{C}_\beta^{-1}(r)|^p \mathrm{d}r$$

$$\mathcal{W}_1(\alpha,\beta) = \|\mathcal{C}_\alpha - \mathcal{C}_\beta\|_{L^1(\mathbb{R})} = \int_{\mathbb{R}} |\mathcal{C}_\alpha(x) - \mathcal{C}_\beta(x)| \mathrm{d}x = \int_{\mathbb{R}} \left| \int_{-\infty}^{x} \mathrm{d}(\alpha - \beta) \right| \mathrm{d}x.$$



α $\qquad$ β $\qquad$ $((t\mathrm{Id} + (1-t)\mathrm{Id})_\sharp \alpha$

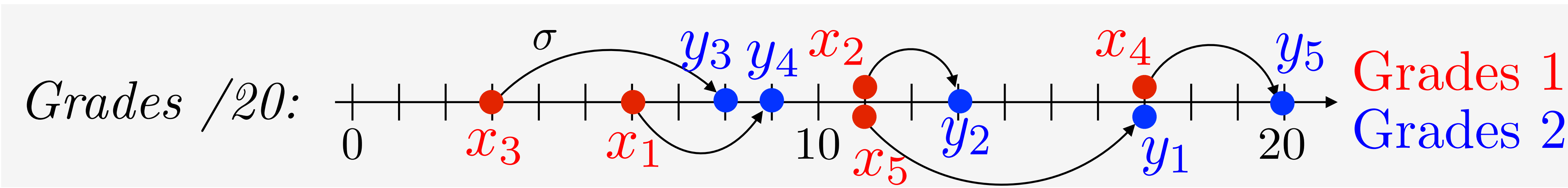
$T = \mathcal{C}_\beta^{-1} \circ \mathcal{C}_\alpha$

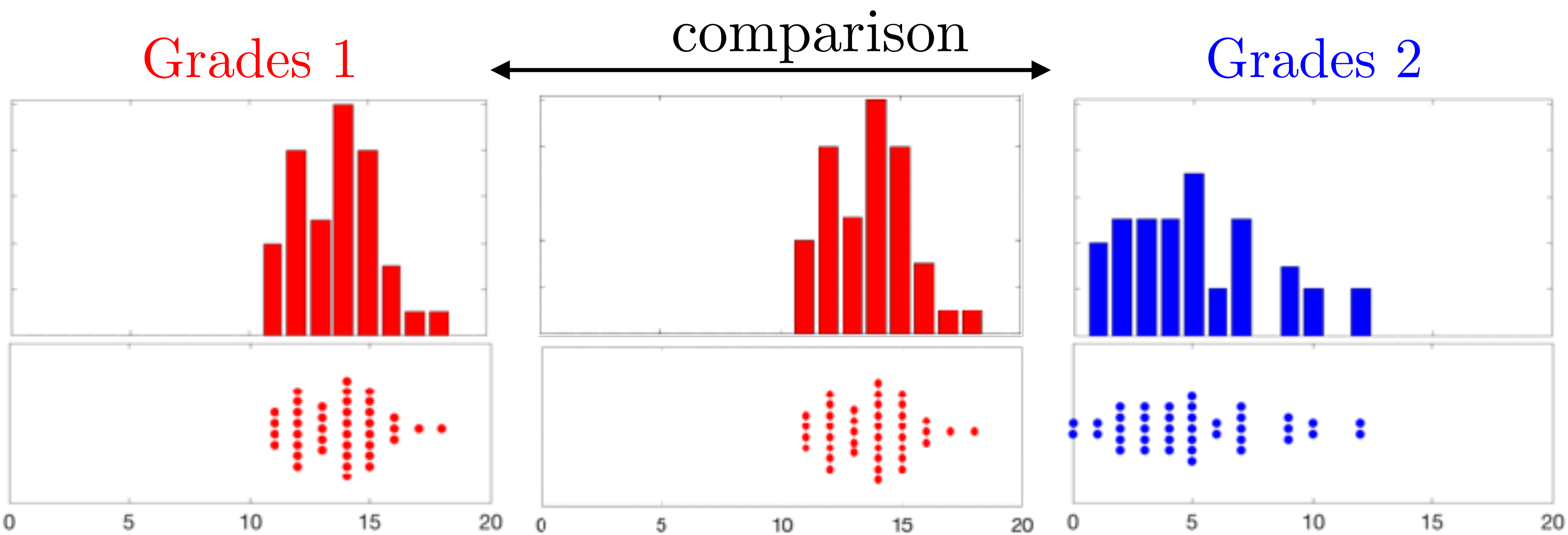


$(\mathcal{C}_\alpha, \mathcal{C}_\beta)$ $\qquad$ $(\mathcal{C}_\alpha^{-1}, \mathcal{C}_\beta^{-1})$ $\qquad$ (T, T^{-1}) $\qquad$ $(1-t)\mathcal{C}_\alpha^{-1} + t\mathcal{C}_\beta^{-1}$

Discrete 1-D Optimal Transport

$$\min_{\sigma \in \Sigma_n} \sum_{i=1}^{n} |x_i - y_{\sigma(i)}|$$





OT Between Gaussians

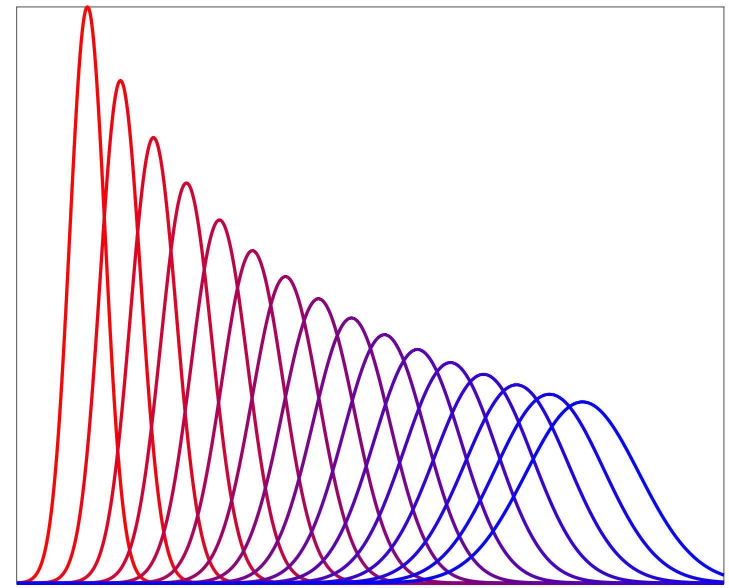
Remark. If $\Omega = \mathbb{R}^d$, $\boldsymbol{c}(x,y) = \|x-y\|^2$, and $\boldsymbol{\mu} = \mathcal{N}(\mathbf{m}_{\boldsymbol{\mu}}, \boldsymbol{\Sigma}_{\boldsymbol{\mu}})$, $\boldsymbol{\nu} = \mathcal{N}(\mathbf{m}_{\boldsymbol{\nu}}, \boldsymbol{\Sigma}_{\boldsymbol{\nu}})$ then

$$W_2^2(\boldsymbol{\mu},\boldsymbol{\nu}) = \|\mathbf{m}_{\boldsymbol{\mu}} - \mathbf{m}_{\boldsymbol{\nu}}\|^2 + B(\boldsymbol{\Sigma}_{\boldsymbol{\mu}}, \boldsymbol{\Sigma}_{\boldsymbol{\nu}})^2$$

where B is the Bures metric

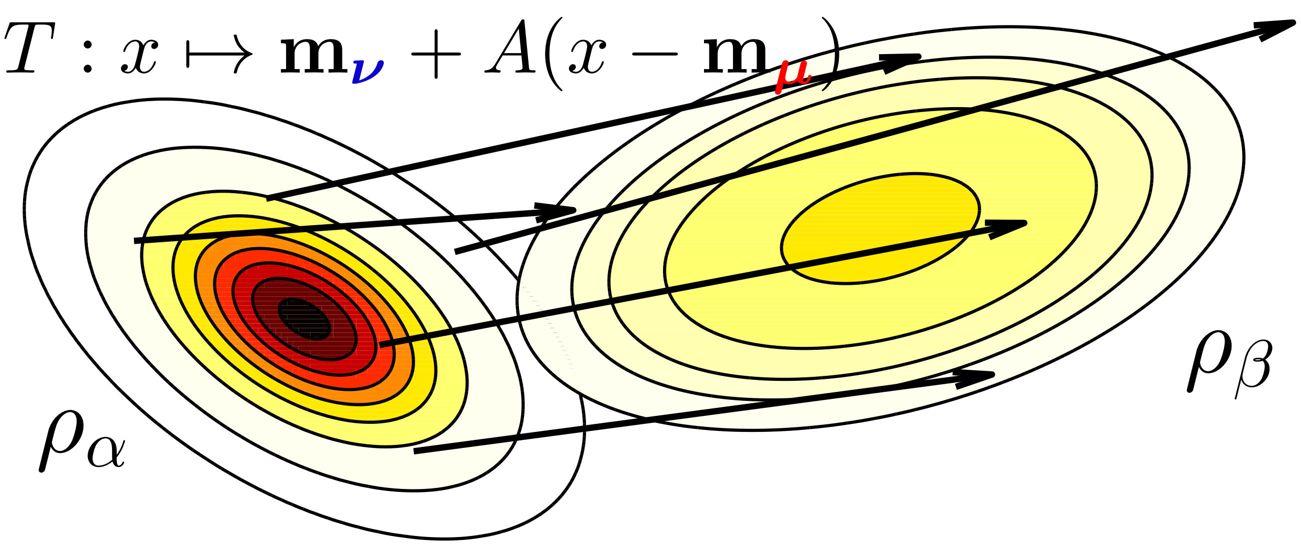
$$B(\boldsymbol{\Sigma}_{\boldsymbol{\mu}}, \boldsymbol{\Sigma}_{\boldsymbol{\nu}})^2 = \operatorname{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\mu}} + \boldsymbol{\Sigma}_{\boldsymbol{\nu}} - 2(\boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{1/2}\boldsymbol{\Sigma}_{\boldsymbol{\nu}}\boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{1/2})^{1/2}).$$

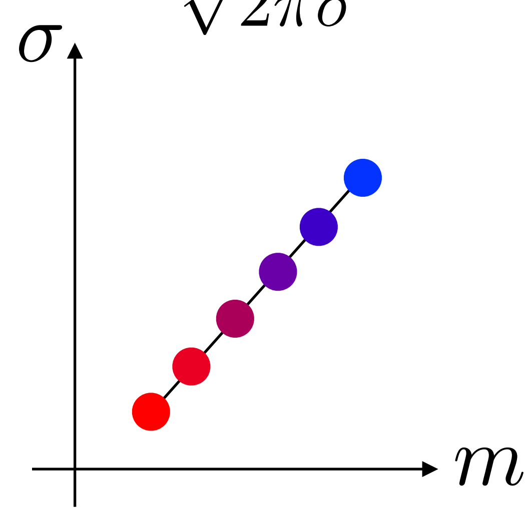
The map $T : x \mapsto \mathbf{m}_{\boldsymbol{\nu}} + A(x - \mathbf{m}_{\boldsymbol{\mu}})$ is **optimal**, where $A = \boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{-\frac{1}{2}} \left(\boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{\frac{1}{2}}\boldsymbol{\Sigma}_{\boldsymbol{\nu}}\boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{\frac{1}{2}}\right)^{\frac{1}{2}} \boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{-\frac{1}{2}}$.



$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

$T : x \mapsto \mathbf{m}_{\boldsymbol{\nu}} + A(x - \mathbf{m}_{\boldsymbol{\mu}})$





OT on Gaussians and Bures' Distance

Remark 2.11 (Distance between Gaussians). If $\alpha = \mathcal{N}(m_\alpha, C_\alpha)$ and $\beta = \mathcal{N}(m_\beta, C_\beta)$, then one can show that

$$\mathcal{W}_2^2(\alpha, \beta) = \|m_\alpha - m_\beta\|^2 + \mathcal{B}(C_\alpha, C_\beta)^2 \tag{2.19}$$

where $\mathcal{B}$ is the so-called Bures metric

$$\mathcal{B}(C_\alpha, C_\beta)^2 \stackrel{\text{def.}}{=} \operatorname{tr}\left(C_\alpha + C_\alpha - 2(C_\alpha^{1/2} C_\beta C_\alpha^{1/2})^{1/2}\right) \tag{2.20}$$