

Gromov-Wasserstein Distances: Statistical & Computational Advancements via Duality Theory

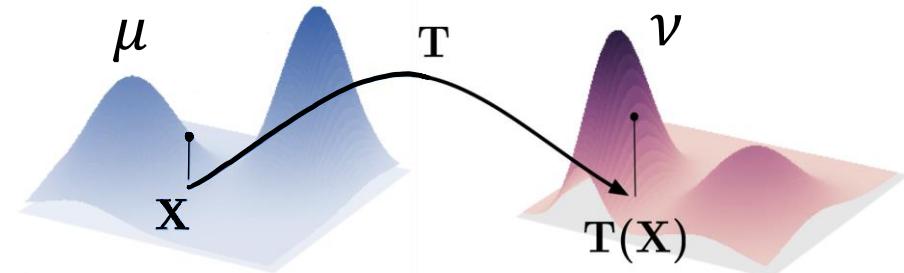
Ziv Goldfeld
Cornell University

2023 North American School of Information Theory

Primer: Optimal Transport Theory

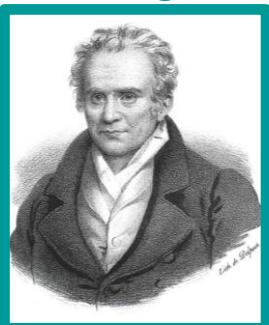
Why Optimal Transport?

Broad interest: Pure math, applied math, economics, comp. biology, machine learning...

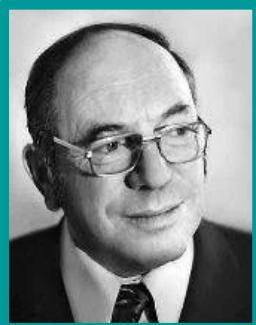


Rich history:

Monge



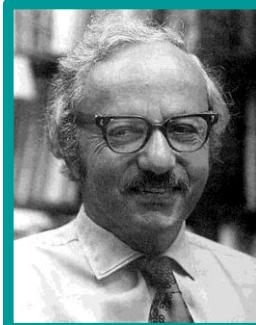
Kantorovich



Koopmans



Dantzig



Caffarelli



Otto



Villani



Figalli



Nobel '75

NMoS '75

Abel '23

Liebniz '06

Fields '10

Fields '18

Has a bit of everything: Theory, statistics, algorithms, applications

Optimal Transport

Distributions: $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$

Cost: $c: \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, \infty)$

Transport map: $T: \mathbb{R}^d \rightarrow \mathbb{R}^d$ s.t. $T_\# \mu = \nu$

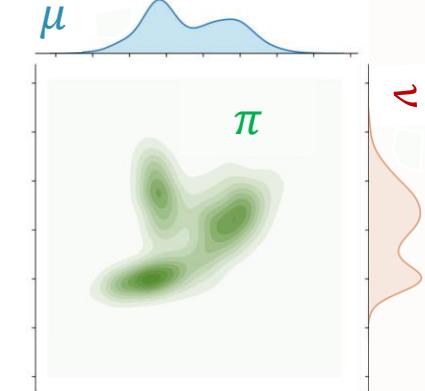
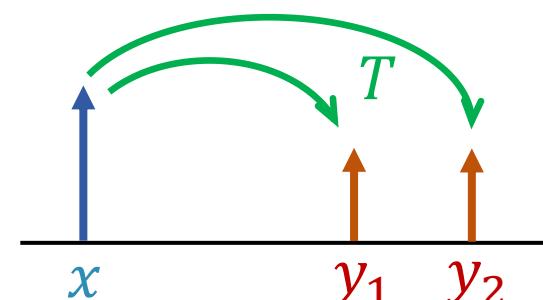
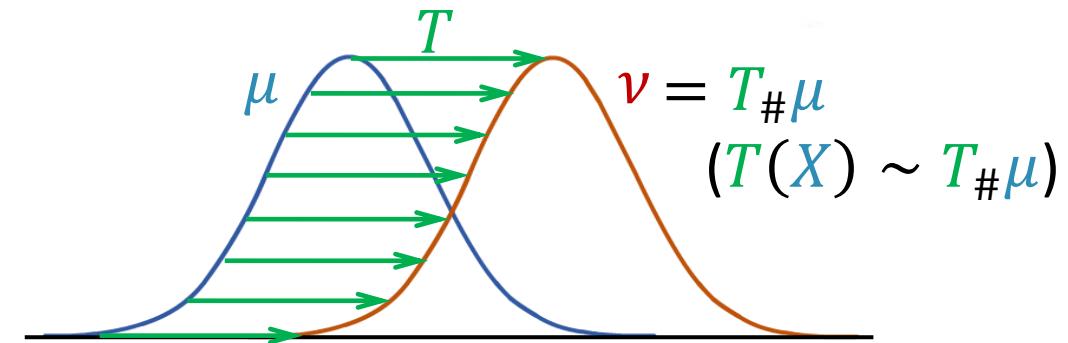
OT problem (Monge 1781): $M_c(\mu, \nu) := \inf_{T: T_\# \mu = \nu} \int_{\mathbb{R}^d} c(x, T(x)) d\mu(x)$

🚫 $\{T: T_\# \mu = \nu\}$ may be empty, not closed, non-linear problem, ...

Coupling: $\Pi(\mu, \nu) = \{\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d): \pi(\cdot \times \mathbb{R}^d) = \mu, \pi(\mathbb{R}^d \times \cdot) = \nu\}$

Optimal Transport (Kantorovich '42)

$$\text{OT}_c(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\pi(x, y)$$



The Wasserstein Distance

Construction: Kantorovich OT with distance cost (or power thereof) $c(x, y) = \|x - y\|^p$

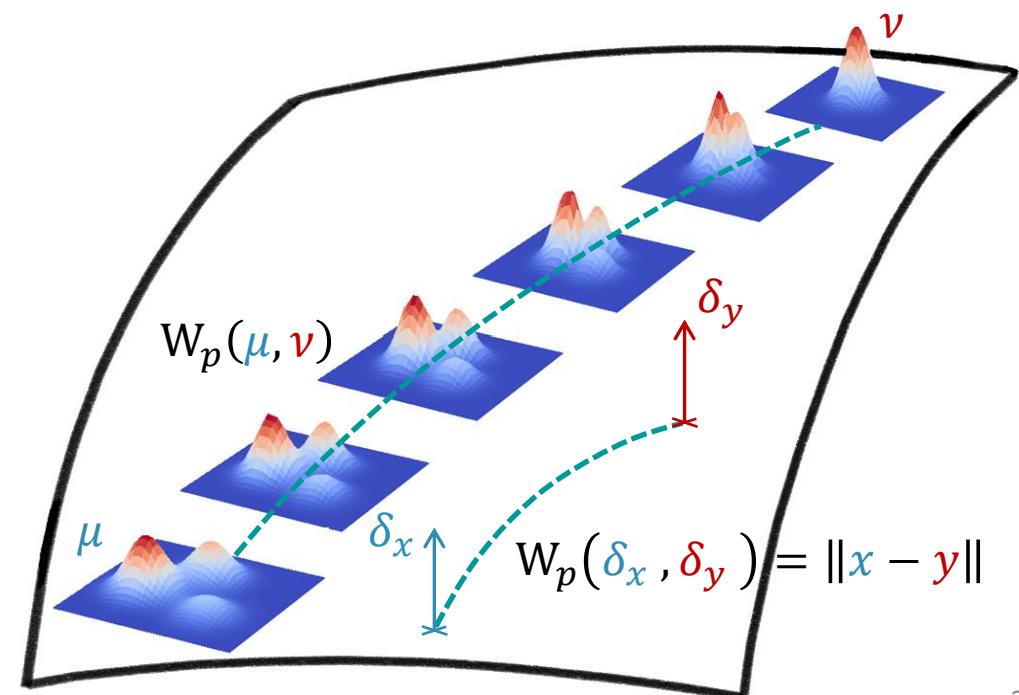
p -Wasserstein Distance

For $p \in [1, \infty)$ and $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$: $W_p(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p d\pi(x, y) \right)^{1/p}$

Wasserstein space: $\mathfrak{W}_p = (\mathcal{P}_p(\mathbb{R}^d), W_p)$ metric space

Wasserstein geometry:

- Euclidean geometry
- Geodesic curves (shortest paths)
- Barycenters (averages)
- Gradient flows



The Wasserstein Metric: Difficulties

$$W_p(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p d\pi(x, y) \right)^{1/p}$$

Statistical: Data $\implies \hat{\mu}_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ & $\hat{\nu}_n := \frac{1}{n} \sum_{i=1}^n \delta_{Y_i} \implies W_p(\mu, \nu) \approx W_p(\hat{\mu}_n, \hat{\nu}_n)?$

Theorem (Dudley '69, Boissard-Le Gouic '14, Fournier-Guillin '14,...)

For $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ and $d > 2p$: $\mathbb{E}[|W_p(\mu, \nu) - W_p(\hat{\mu}_n, \hat{\nu}_n)|] \asymp n^{-\frac{1}{d}}$

🚫 **Too slow** for $d \gg 1$

Computational: Kantorovich OT is LP \implies Network flow solvers $O(n^3 \log n)$

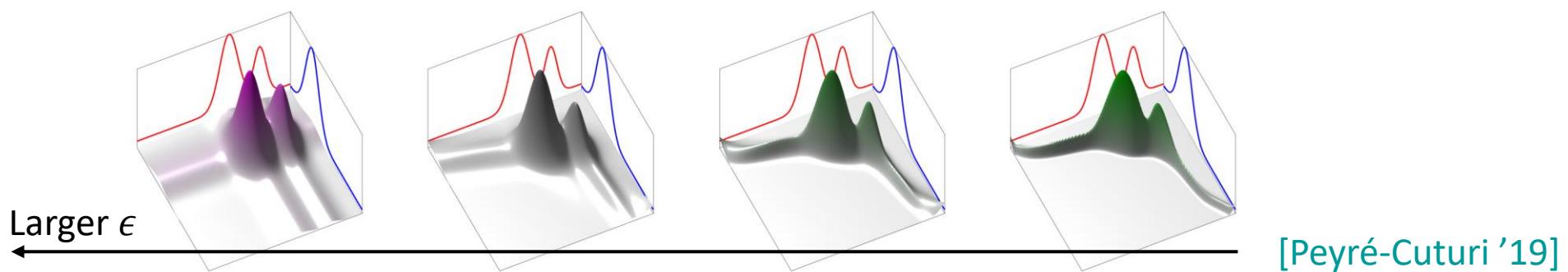
🚫 **Infeasible** for large scale problems

Entropic Optimal Transport

Entropic Optimal Transport

$$\text{For } \epsilon > 0: \text{EOT}_{\epsilon,c}(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \mathbb{E}_\pi[c(X, Y)] - \epsilon H(\pi)$$

- **Entropic penalty:** Encourage randomness of π



- **Approximation error:** $|\text{EOT}_{\epsilon,c}(\mu, \nu) - \text{OT}_c(\mu, \nu)| \lesssim \epsilon \log(1/\epsilon)$

→ Strongly convex optimization problem with a unique and smooth solution

Entropic Optimal Transport: Estimation

Setting: $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ & $\hat{\nu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$ $\implies \text{EOT}_{\epsilon,c}(\mu, \nu) \approx \text{EOT}_{\epsilon,c}(\hat{\mu}_n, \hat{\nu}_n)$?

Duality: $\text{EOT}_{\epsilon,c}(\mu, \nu) := \sup_{(\varphi, \psi) \in L^1(\mu) \times L^1(\nu)} \int \varphi d\mu + \int \psi d\nu - \underbrace{\epsilon \left(\int e^{\frac{\varphi(x) + \psi(y) - c(x,y)}{\epsilon}} d\mu \otimes \nu - 1 \right)}_{= 0 \text{ for optimal } (\varphi, \psi)}$

Empirical convergence analysis: Standard technique

- Regularity of EOT potentials:** $(\varphi, \psi) \in \mathcal{F}_s \times \mathcal{G}_s$ for Hölder classes of arbitrary smoothness
- Suprema of emp. process:** Decompose

$$\mathbb{E}[|\text{EOT}_{\epsilon,c}(\mu, \nu) - \text{EOT}_{\epsilon,c}(\hat{\mu}_n, \hat{\nu}_n)|] \leq \mathbb{E} \left[\sup_{\varphi \in \mathcal{F}_s} \left| \mathbb{E}_\mu[\varphi] - \frac{1}{n} \sum_{i=1}^n \varphi(X_i) \right| \right] + \mathbb{E} \left[\sup_{\psi \in \mathcal{G}_s} \left| \mathbb{E}_\mu[\psi] - \frac{1}{n} \sum_{i=1}^n \psi(Y_i) \right| \right]$$

Bound Dudley entropy integral of \mathcal{F}_s and \mathcal{G}_s (Hölder) with $s = \left\lceil \frac{d_x}{2} \right\rceil + 1$

$\lesssim 1/\sqrt{n}$

Entropic Optimal Transport: Computation

Setting: Compute EOT between discrete measures $\mu = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ & $\nu = \frac{1}{n} \sum_{i=1}^n \delta_{y_i}$

$$\text{EOT}_{\epsilon,c}(\mu, \nu) = \min_{\pi \in \Pi(\mu, \nu)} \langle \pi, C \rangle - \epsilon H(\pi)$$

Coupling matrix $[\pi]_{i,j} = \pi(x_i, y_j)$
(vectorized)

Cost matrix $[C]_{i,j} = c(x_i, y_j)$
(vectorized)

Proposition

Optimal $\pi_\epsilon^* \in \Pi(\mu, \nu)$ is unique & $\exists \mathbf{a}, \mathbf{b} \in \mathbb{R}_{\geq 0}^n$ s.t. $\pi_\epsilon^* = \text{diag}(\mathbf{a}) K \text{diag}(\mathbf{b})$, $[K]_{i,j} = e^{-\frac{[C]_{i,j}}{\epsilon}}$

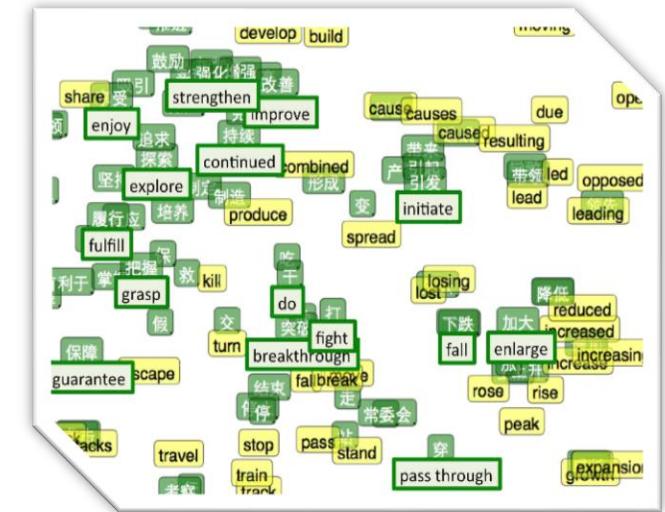
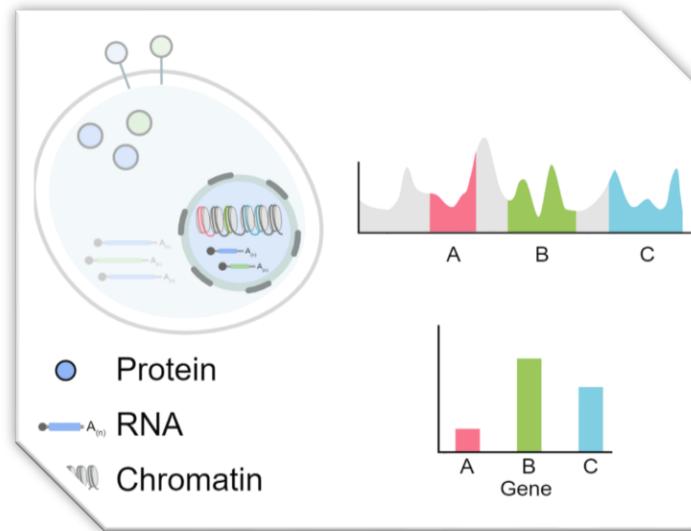
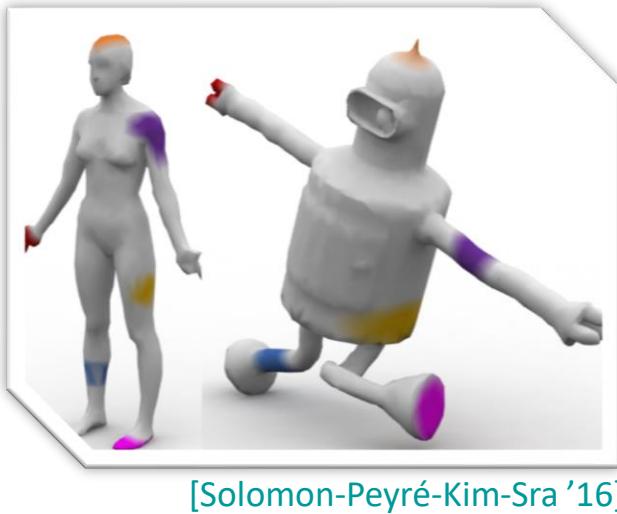
$$\pi_\epsilon^* \in \Pi(\mu, \nu) \iff \begin{cases} \mathbf{a} = \mu / K \mathbf{b} \\ \mathbf{b} = \nu / K \mathbf{a} \end{cases}$$

\implies **Fixed point (Sinkhorn) algorithm:** $O(n^2)$ time & highly parallelizable [Cuturi '13]

Gromov-Wasserstein Distance

Heterogeneous & Structured Data

Dataset Matching: Various applications require matching heterogeneous & structured datasets



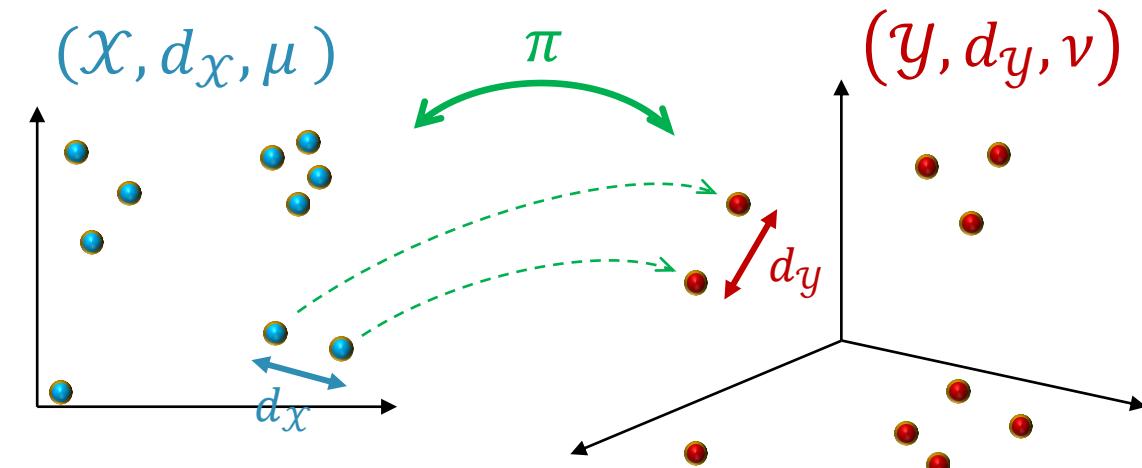
- Goals:**
1. Compare how similar/different two datasets are
 2. Obtain matching/alignment

Gromov-Wasserstein Distance

- Datasets as metric measure spaces
⇒ $(\mathcal{X}, d_{\mathcal{X}}, \mu)$ & $(\mathcal{Y}, d_{\mathcal{Y}}, \nu)$
- Find matching (transport map) $T: \mathcal{X} \rightarrow \mathcal{Y}$
⇒ $\nu = T_{\#}\mu$ (if $X \sim \mu$ then $T(X) \sim T_{\#}\mu$)

- Preserve distances (minimize distance distortion)

$$\Rightarrow \text{cost} = \left| d_{\mathcal{X}}(x_i, x_j)^q - d_{\mathcal{Y}}(T(x_i), T(x_j))^q \right|$$



Gromov-Wasserstein Distance (Memoli '11)

The (p, q) -GW distance between mm spaces $(\mathcal{X}, d_{\mathcal{X}}, \mu)$ and $(\mathcal{Y}, d_{\mathcal{Y}}, \nu)$ is

$$D_{p,q}(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \left(\mathbb{E}_{\substack{(X, Y) \sim \pi \\ (X', Y') \sim \pi}} \left[|d_{\mathcal{X}}(X, X')^q - d_{\mathcal{Y}}(Y, Y')^q|^p \right] \right)^{1/p}$$

Gromov-Wasserstein Distance

$$D_{p,q}(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \left(\mathbb{E}_{\substack{(X,Y) \sim \pi \\ (X',Y') \sim \pi}} \left[|d_X(X, X')^q - d_Y(Y, Y')^q|^p \right] \right)^{1/p}$$

Comments: Relaxation of Gromov-Hausdorff distance between metric spaces ($p = \infty, q = 1$)

- **Finiteness:** $D_{p,q}(\mu, \nu) < \infty \forall \mu, \nu$ with $\int_{\mathcal{X} \times \mathcal{X}} d_X(x, x')^{pq} d\mu \otimes \nu(x, x') < \infty$ & resp. for ν
 - **Identification:** $D_{p,q}(\mu, \nu) = 0 \iff \exists$ isometry $T: \mathcal{X} \rightarrow \mathcal{Y}$ with $T_\# \mu = \nu$ (invariances)
 - **Metric:** Metrizes space of equivalence classes of mm spaces with finite size
 - **Computation:** $D_{p,q} \left(\frac{1}{n} \sum_{i=1}^n \delta_{x_i}, \frac{1}{n} \sum_{i=1}^n \delta_{y_i} \right)^p = \frac{1}{n^2} \min_{\sigma \in S_n} \sum_{i,j=1}^n |d_X(x_i, x_j)^q - d_Y(y_{\sigma(i)}, y_{\sigma(j)})^q|^p$
- 🚫 Quadratic assignment problem (non-convex) [Commander '05] \implies NP complete

Entropic Gromov-Wasserstein Distance

Approach: Variants/reformulations of GW problem for computational tractability

- **Sliced GW:** Avg/max of GW btw low-dimensional projections [Vayer-Flamary-Tavenard '20]
- **Unbalanced GW:** Relax marginal constraints via f -div. penalty [Séjourné-Vialard-Peyré '23]
- **Entropic GW:** Add entropic penalty to GW cost [Peyré-Cuturi-Solomon '16]

Entropic Gromov-Wasserstein Distance

$$S_{p,q}^\epsilon(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \mathbb{E}_{\pi \otimes \pi} \left[|d_\chi(X, X')^q - d_y(Y, Y')^q|^p \right] - \epsilon H(\pi)$$

✳ Computed via mirror-descent with Sinkhorn iterations

↳ Convergence to stationary point (asymptotic)

Entropic Gromov-Wasserstein Theory

Open Questions:

1. Convexity regimes in ϵ ?
2. Algorithms with (global) convergence rates?
3. Sample complexity for statistical estimation?

Approach: New duality theory to relate EGW to EOT

Duality for Entropic GW Distance

Setting: Quadratic cost over Euclidean spaces

- **mm-spaces:** $(\mathbb{R}^{d_x}, \|\cdot\|, \mu)$ and $(\mathbb{R}^{d_y}, \|\cdot\|, \nu)$ with $M_4(\mu) := \int \|x\|^4 d\mu(x), M_4(\nu) < \infty$
- **Quadratic EGW:** $S_\epsilon(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \iint \left| \|x - x'\|^2 - \|y - y'\|^2 \right|^2 d\pi \otimes \pi - \epsilon H(\pi)$

Decomposition: Assume w.l.o.g. that μ, ν are centered (invariance to translation); then

$$S_\epsilon(\mu, \nu) = S_1(\mu, \nu) + S_{2,\epsilon}(\mu, \nu)$$

where $S_1(\mu, \nu) = \int \|x - x'\|^4 d\mu \otimes \mu + \int \|y - y'\|^4 d\nu \otimes \nu - 4 \int \|x\|^2 \|y\|^2 d\mu \otimes \nu$

$$S_{2,\epsilon}(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int -4\|x\|^2 \|y\|^2 d\pi - 8 \sum_{\substack{1 \leq i \leq d_x \\ 1 \leq j \leq d_y}} \left(\int x_i y_j d\pi \right)^2 - \epsilon H(\pi)$$

→ Derive a dual form for $S_{2,\epsilon}(\mu, \nu)$!



Duality Theory for Entropic GW Distance

Approach: Linearize quadratic term using auxiliary variables

$$S_{2,\epsilon}(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int -4\|x\|^2\|y\|^2 d\pi - 8 \sum_{\substack{1 \leq i \leq d_x \\ 1 \leq j \leq d_y}} \left(\int x_i y_j d\pi \right)^2 - \epsilon H(\pi)$$

Optimality at
 $a_{ij}^*(\pi) = 0.5 \int x_i y_j d\pi$

and define

$$M_{\mu, \nu} = \sqrt{M_2(\mu)M_2(\nu)}$$

$\mathcal{D}_{M_{\mu, \nu}}$ = entry-wise bdd
 $d_x \times d_y$ -sized matrices

$$\begin{aligned} &= \inf_{\pi \in \Pi(\mu, \nu)} \int -4\|x\|^2\|y\|^2 d\pi + 32 \sum_{\substack{1 \leq i \leq d_x \\ 1 \leq j \leq d_y}} \inf_{-\frac{M_{\mu, \nu}}{2} \leq a_{ij} \leq \frac{M_{\mu, \nu}}{2}} \left(a_{ij}^2 - \int a_{ij} x_i y_j d\pi \right) - \epsilon H(\pi) \\ &= \inf_{\mathbf{A} \in \mathcal{D}_{M_{\mu, \nu}}} 32\|\mathbf{A}\|_F^2 + \underbrace{\inf_{\pi \in \Pi(\mu, \nu)} \int (-4\|x\|^2\|y\|^2 - 32\mathbf{x}^T \mathbf{A} \mathbf{y}) d\pi - \epsilon H(\pi)}_{=: c_{\mathbf{A}}(x, y)} = \text{EOT}_{\epsilon, c_{\mathbf{A}}}(\mu, \nu) \end{aligned}$$

Theorem (Zhang-G.-Mroueh-Sriperumbudur '23)

Fix $\epsilon > 0$, $(\mu, \nu) \in \mathcal{P}_4(\mathbb{R}^{d_x}) \times \mathcal{P}_4(\mathbb{R}^{d_y})$, and any $M \geq \sqrt{M_2(\mu)M_2(\nu)}$, we have

$$S_{2,\epsilon}(\mu, \nu) = \inf_{\mathbf{A} \in \mathcal{D}_M} 32\|\mathbf{A}\|_F^2 + \text{EOT}_{\epsilon, c_{\mathbf{A}}}(\mu, \nu)$$

Sample Complexity of Entropic GW

Theorem (Zhang-G.-Mroueh-Sriperumbudur '23)

Fix $\epsilon > 0$ and let $(\mu, \nu) \in \mathcal{P}(\mathbb{R}^{d_x}) \times \mathcal{P}(\mathbb{R}^{d_y})$ be 4-sub-Weibull with parameter $\sigma^2 > 0$. Then

$$\mathbb{E}[|S_\epsilon(\mu, \nu) - S_\epsilon(\hat{\mu}_n, \hat{\nu}_n)|] \lesssim_{d_x, d_y} \underbrace{\frac{1 + \sigma^4}{\sqrt{n}}}_{S_1 \text{ rate} + \text{centering bias}} + \epsilon \left(1 + \left(\frac{\sigma}{\sqrt{\epsilon}} \right)^{9[(d_x \vee d_y)/2] + 11} \right) \frac{1}{\sqrt{n}}$$

Comments:

- **Optimality:** Rate is parametric and hence minimax optimal
- **Entropic OT:** Rate matches that for EOT (assuming compact support or sub-Gaussianity)
- **One-sample:** When only μ is estimated, rate is similar but with d_x instead of $d_x \vee d_y$

Sample Complexity of Entropic GW: Proof Outline

Decomposition: Split S_ϵ into $S_1 + S_{2,\epsilon}$ and center empirical measures

$$\mathbb{E}[|S_\epsilon(\mu, \nu) - S_\epsilon(\hat{\mu}_n, \hat{\nu}_n)|] \leq \mathbb{E}[|S_1(\mu, \nu) - S_1(\hat{\mu}_n, \hat{\nu}_n)|] + \mathbb{E}[|S_{2,\epsilon}(\mu, \nu) - S_{2,\epsilon}(\hat{\mu}_n, \hat{\nu}_n)|] + \frac{\sigma^2}{\sqrt{n}}$$

S_1 Analysis: Involves only estimation of moments \implies Rate is parametric $\asymp 1/\sqrt{n}$

$S_{2,\epsilon}$ Analysis: Hinges on dual form + regularity analysis of optimal potentials

- EOT reduction:** $|S_{2,\epsilon}(\mu, \nu) - S_{2,\epsilon}(\hat{\mu}_n, \hat{\nu}_n)| \leq \sup_{A \in \mathcal{D}_M} |EOT_{\epsilon,c_A}(\mu, \nu) - EOT_{\epsilon,c_A}(\hat{\mu}_n, \hat{\nu}_n)|$ *
- Dual potentials:** $\forall A \in \mathcal{D}_M, (\varphi_A, \psi_A) \in \mathcal{F}_s \times \mathcal{G}_s$ for Hölder classes of arbitrary smoothness
- Empirical processes:** $\mathbb{E}[\textcircled{*}] \leq \mathbb{E} \left[\sup_{\varphi \in \mathcal{F}_s} |(\mu - \hat{\mu}_n)\varphi| \right] + \mathbb{E} \left[\sup_{\psi \in \mathcal{G}_s} |(\nu - \hat{\nu}_n)\psi| \right] \lesssim 1/\sqrt{n}$

From Stability Analysis to Convexity

$$S_\epsilon(\mu, \nu) = S_1(\mu, \nu) + \min_{\mathbf{A} \in \mathcal{D}_M} \left\{ \underbrace{32\|\mathbf{A}\|_F^2}_{=: \Phi(\mathbf{A})} + \text{EOT}_{\epsilon, c_A}(\mu, \nu) \right\}$$

- Analysis:**
- Fréchet derivatives $D\Phi_{[\mathbf{A}]}$ and $D^2\Phi_{[\mathbf{A}]}$
 - Bound $\lambda_{max}(D^2\Phi_{[\mathbf{A}]}) \leq 64$ & $\lambda_{min}(D^2\Phi_{[\mathbf{A}]}) \geq 32^2\epsilon^{-1}\sqrt{M_4(\mu)M_4(\nu)} - 64$

Theorem (Rioux-G.-Kato '23)

1. For $M \geq \sqrt{M_2(\mu)M_2(\nu)}$, all minimizers of Φ are in \mathcal{D}_M
2. Φ is strictly convex whenever $\epsilon > 16\sqrt{M_4(\mu)M_4(\nu)}$
3. Φ is L -smooth on \mathcal{D}_M with $L \leq 64 \vee (32^2\epsilon^{-1}\sqrt{M_4(\mu)M_4(\nu)} - 64)$

First-Order Inexact Oracle Methods

$$\min_{\mathbf{A} \in \mathcal{D}_M} 32\|\mathbf{A}\|_F^2 + \text{EOT}_{\epsilon, c_{\mathbf{A}}}(\mu, \nu)$$

First-order methods: Gradient of objective at $\mathbf{A} \in \mathcal{D}_M$ depends on optimal EOT coupling $\pi^{\mathbf{A}}$

$$D\Phi_{[\mathbf{A}]} = 64\mathbf{A} - 32\sum_{i,j=1}^n x_i y_j^T \pi_{i,j}^{\mathbf{A}}$$

Inexact oracle (Sinkhorn): $\tilde{\pi}^{\mathbf{A}}$ s.t. $\|\pi^{\mathbf{A}} - \tilde{\pi}^{\mathbf{A}}\|_{\infty} \leq \delta$

- Gradient approximation $\tilde{D}\Phi_{[\mathbf{A}]}$ ($\tilde{\pi}^{\mathbf{A}}$ instead of $\pi^{\mathbf{A}}$)
- First-order method under convexity [d'Aspremont '08]

→ Computes EGW cost and (approx.) coupling

Algorithm 1 Fast gradient method with inexact oracle

```
Fix  $L = 64$  and let  $\alpha_k = \frac{k+1}{2}$ , and  $\tau_k = \frac{2}{k+3}$ 
1:  $k \leftarrow 0$ 
2:  $\mathbf{A}_0 \leftarrow \mathbf{0}$ 
3:  $\mathbf{G}_0 \leftarrow \tilde{D}\Phi_{[\mathbf{A}_0]}$ 
4:  $\mathbf{W}_0 \leftarrow \alpha_0 \mathbf{G}_0$ 
5: while stopping condition is not met do
6:    $\mathbf{B}_k \leftarrow \frac{M}{2} \text{sign}(\mathbf{A}_k - L^{-1} \mathbf{G}_k) \min\left(\frac{2}{M} |\mathbf{A}_k - L^{-1} \mathbf{G}_k|, 1\right)$ 
7:    $\mathbf{C}_k \leftarrow \frac{M}{2} \text{sign}(-L^{-1} \mathbf{W}_k) \min\left(\frac{2}{M} |L^{-1} \mathbf{W}_k|, 1\right)$ 
8:    $\mathbf{A}_{k+1} \leftarrow \tau_k \mathbf{C}_k + (1 - \tau_k) \mathbf{B}_k$ 
9:    $\mathbf{G}_{k+1} \leftarrow \tilde{D}\Phi_{[\mathbf{A}_{k+1}]}$ 
10:   $\mathbf{W}_{k+1} \leftarrow \mathbf{W}_k + \alpha_{k+1} \mathbf{G}_{k+1}$ 
11:   $k \leftarrow k + 1$ 
12: return  $\mathbf{B}_k$ 
```

Global Convergence Guarantees (Convex)

$$\min_{\mathbf{A} \in \mathcal{D}_M} 32\|\mathbf{A}\|_F^2 + \text{EOT}_{\epsilon, c_{\mathbf{A}}}(\mu, \nu)$$

Theorem (Rioux-G.-Kato '23)

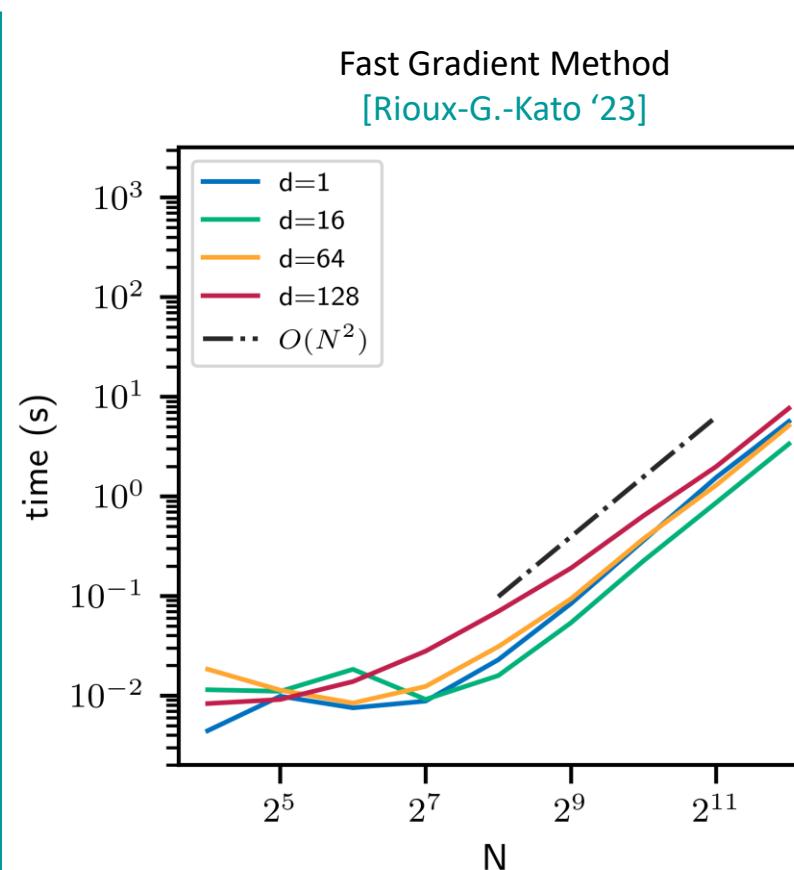
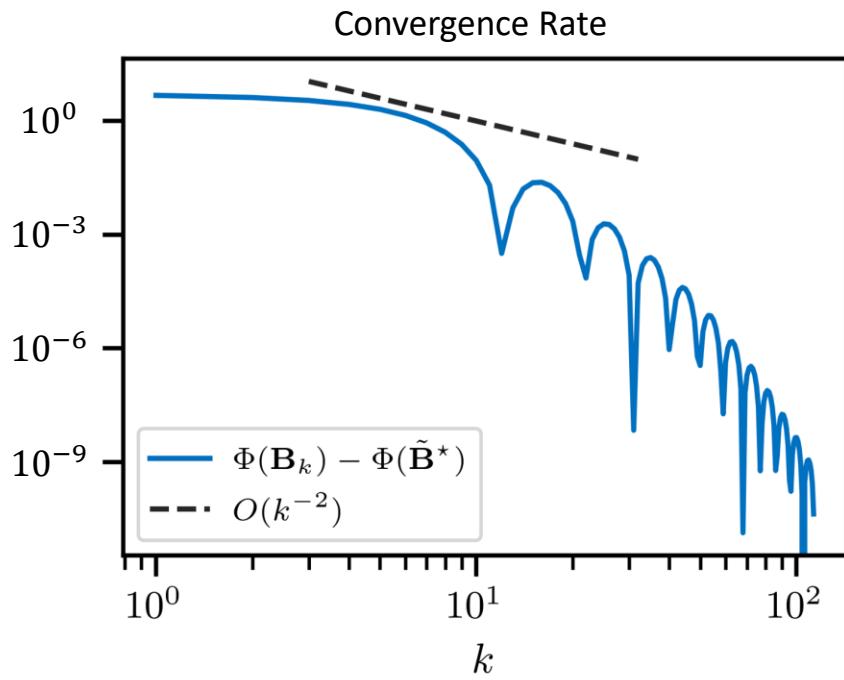
If Φ is convex and L -smooth on \mathcal{D}_M with global min \mathbf{B}_* , then \mathbf{B}_k from Algorithm 1 satisfies

$$\Phi(\mathbf{B}_k) - \Phi(\mathbf{B}_*) \leq \frac{2L\|\mathbf{B}_*\|_F^2}{(k+1)(k+2)} + O(M\delta)$$

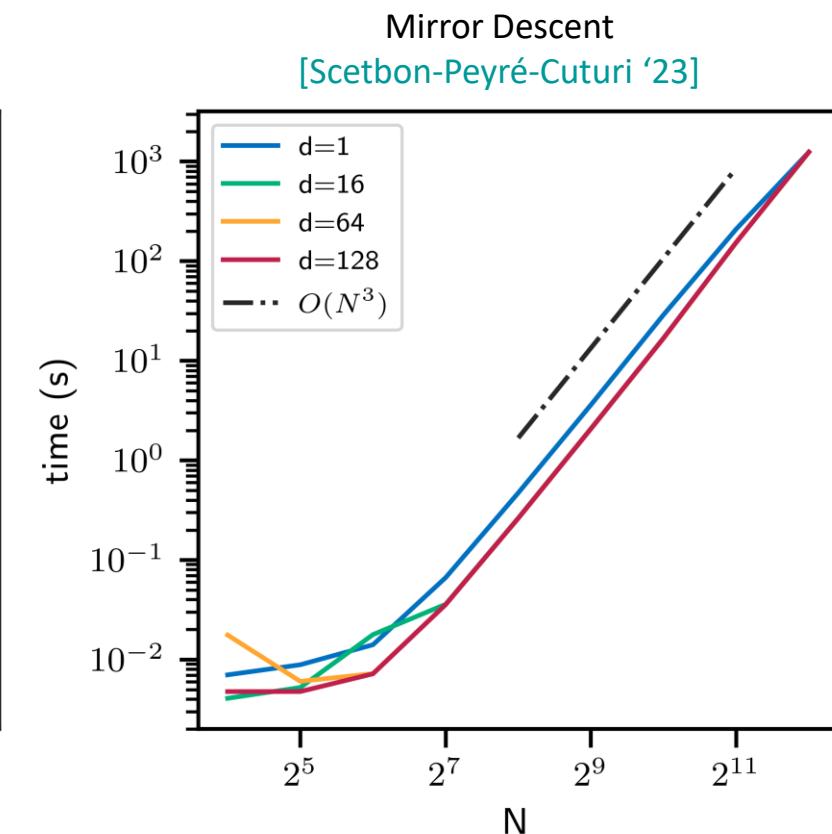
Comments:

- **Optimality:** Optimal complexity of $O(1/k^2)$ for smooth constrained opt. [Nesterov '03]
- **Non-convex regime:** Via smooth non-convex opt. with inexact oracle [Ghadimi-Lan '16]
 - ↳ Adapts to convexity of Φ (yields improved rates if convex)

Numerical Results

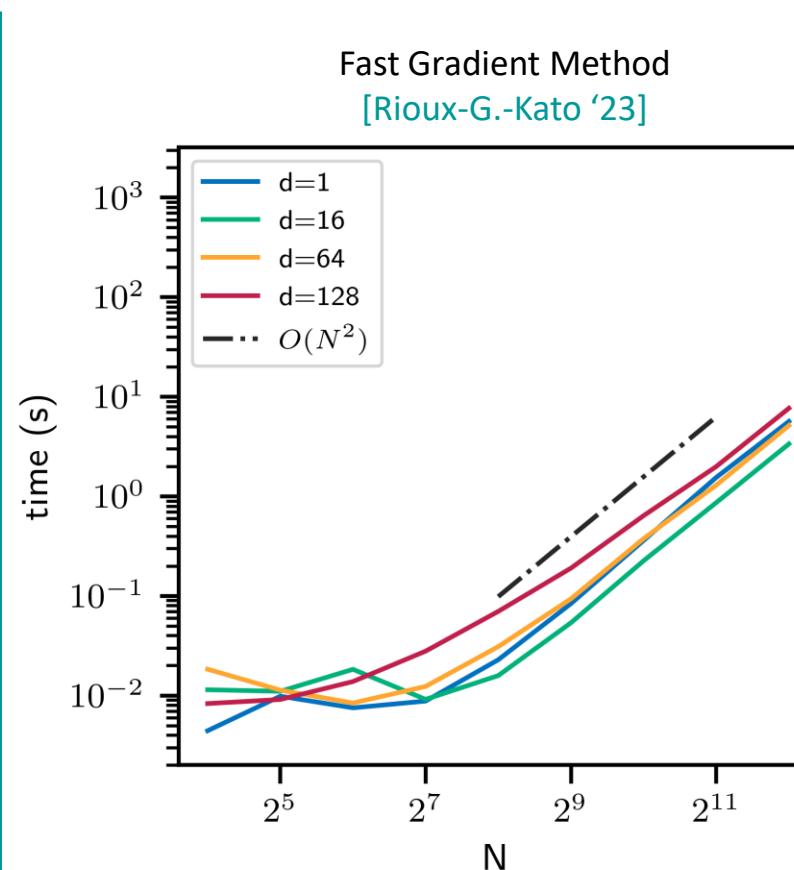
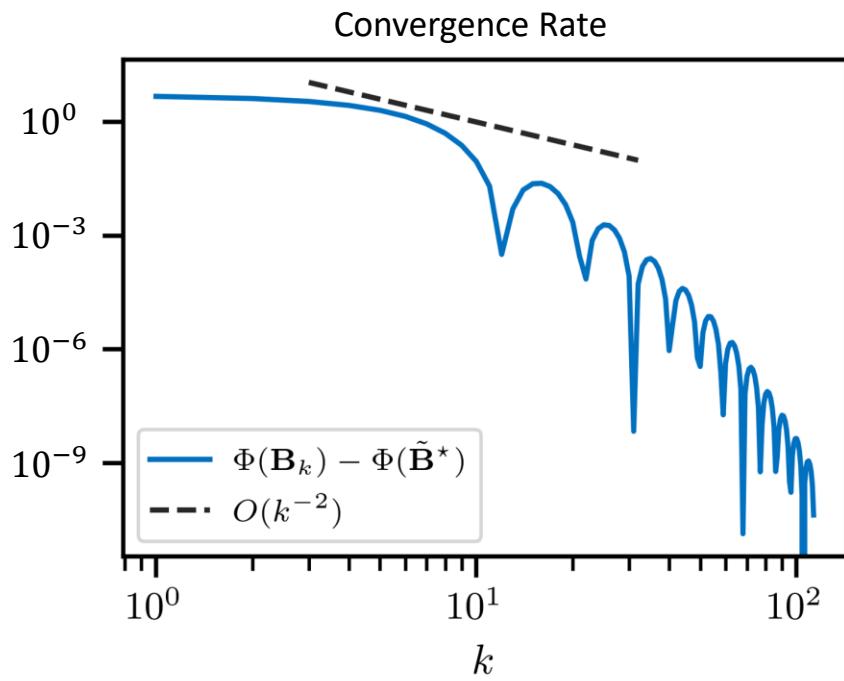


$$\begin{aligned} \text{Time} &= \text{iteration} \times \text{Sinkhorn} \\ &= k \times O(N^2) \end{aligned}$$

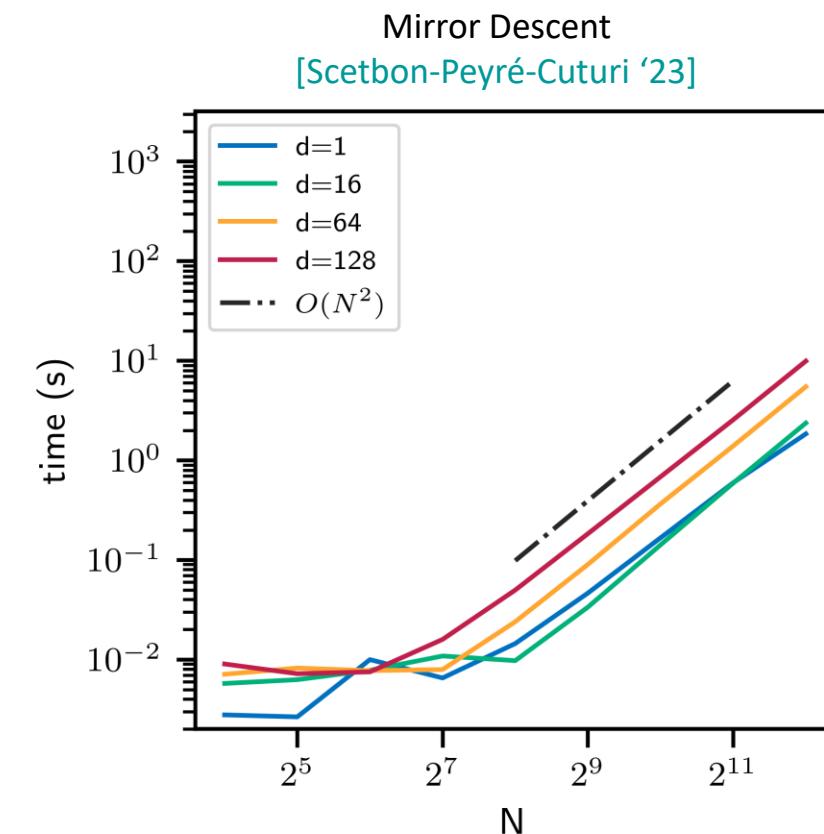


$$\begin{aligned} \text{Time} &= \text{iteration} \times \text{cost update} \\ &= k \times O(N^3) \end{aligned}$$

Numerical Results



$$\text{Time} = \text{iteration} \times \text{Sinkhorn} \\ = k \times O(N^2)$$



$$\text{Time} = \text{iteration} \times \text{cost update} \\ = k \times d \times O(N^2)$$

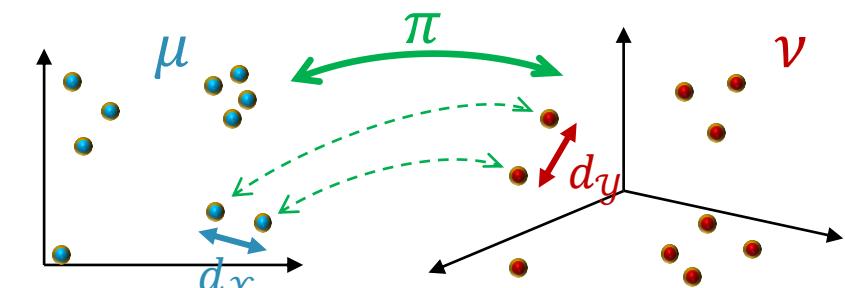
Summary

Gromov-Wasserstein Distance: Quantifies discrepancy between mm spaces

- Applications in ML and beyond for heterogeneous data
- Foundational statistical & computational questions open

Contributions: Duality, empirical rates, and algorithms

- Dual form that connects to EOT
- First sample complexity result for EGW (quadratic cost over Euclidean spaces)
- First algorithms with convergence rates (global optimality under convexity)
- Duality and empirical rates also derived for non-entropic GW



[A] Zhang, Goldfeld, Mroueh, Sriperumbudur, "Gromov-Wasserstein distances: entropic regularization, duality, and sample complexity", ArXiv: 2212.12848

[B] Rioux, Goldfeld, Kato, "Entropic Gromov-Wasserstein distances: stability, algorithms, and distributional limits", ArXiv: 2306.00182

Thank you!