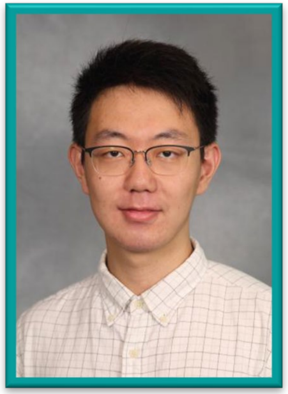
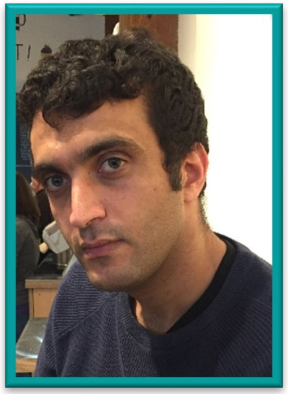
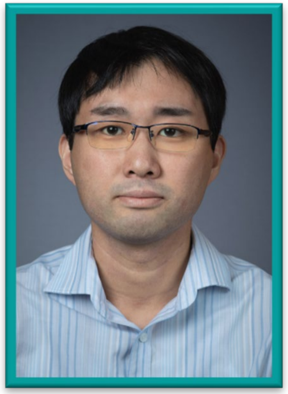
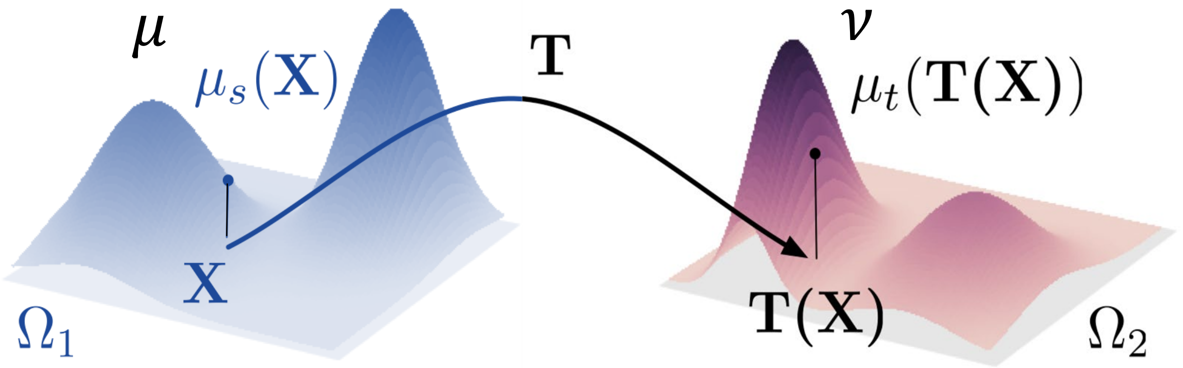
Gromov-Wasserstein Alignment:
Statistical & Computational Advancements via Duality

Ziv Goldfeld
Cornell University

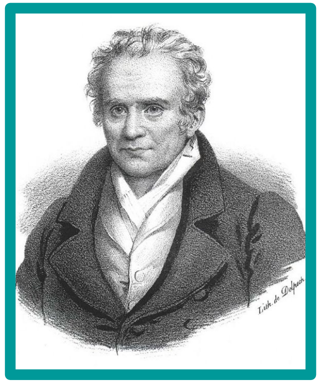



Optimal Transport



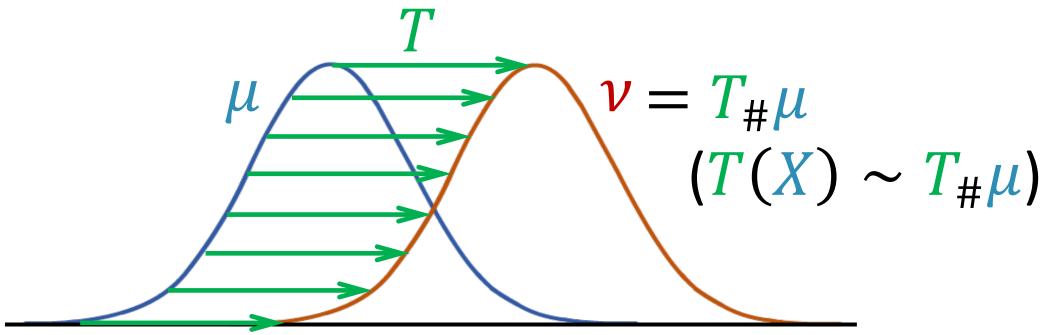
Optimal Transport



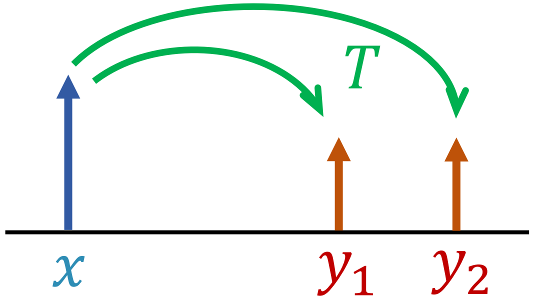
Monge (1781)

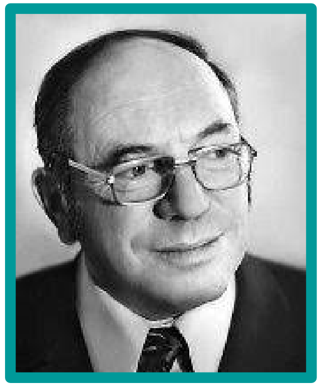
$$\mathrm{M}_c(\mu,\nu) := \inf_{\substack{T:\mathcal{X}\to\mathcal{Y} \\ T_\#\mu=\nu}} \int c(x,T(x))\,d\mu(x)$$

Transport map



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





Kantorovich (1942)

Kantorovich Optimal Transport

$$\mathrm{OT}_c(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \iint c\,d\pi = \sup_{\substack{(\varphi,\psi)\in L^1(\mu)\times L^1(\nu): \\ \varphi(x)+\psi(y)\le c(x,y)}} \int \varphi\,d\mu + \int \psi\,d\nu$$

Coupling (transport plan)

The Wasserstein Distance

Construction: Kantorovich OT with distance cost (or power) $c(x,y) = \|x-y\|^p,\ p \in [1,\infty)$

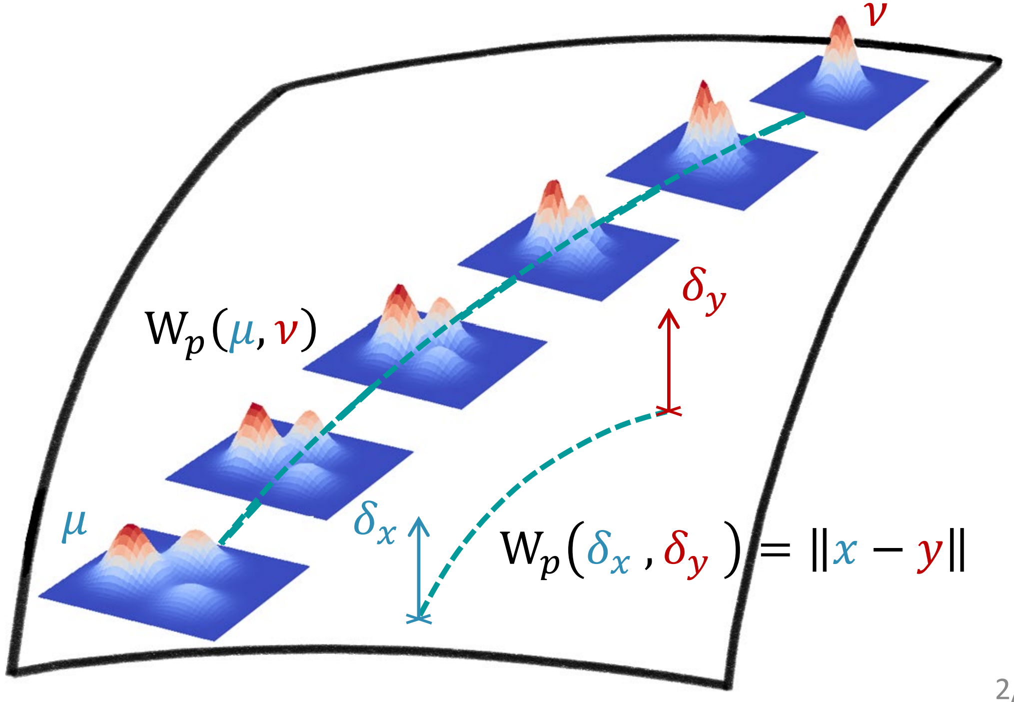
p-Wasserstein Distance

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

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

Wasserstein geometry:

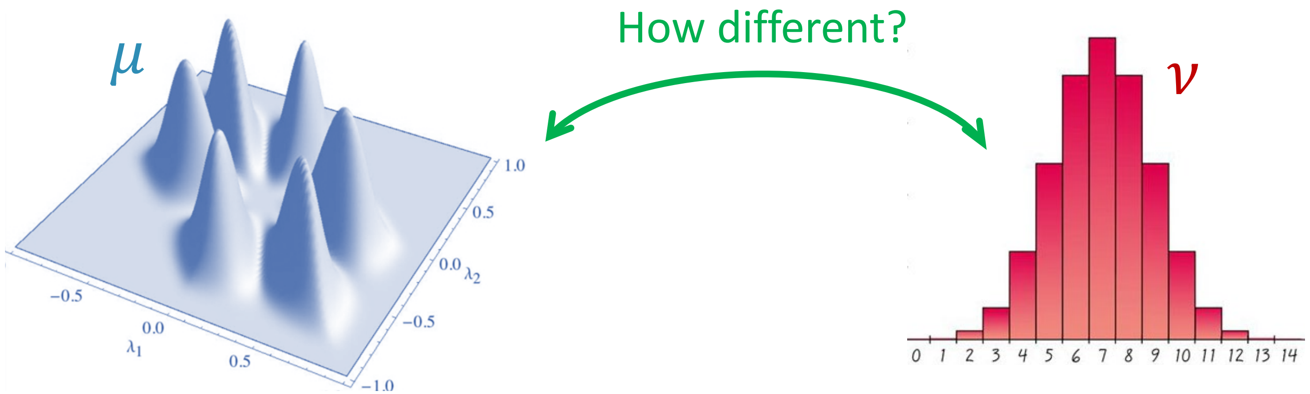
- Euclidean geometry
- Geodesic curves
- Barycenters
- Gradient flows



Beyond OT and Wasserstein Distances

Structure Preserving Interpolation:

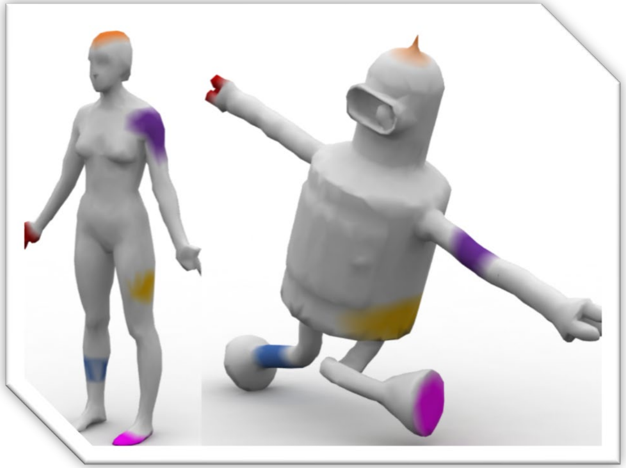
Discrepancy quantification btw incompatible spaces:



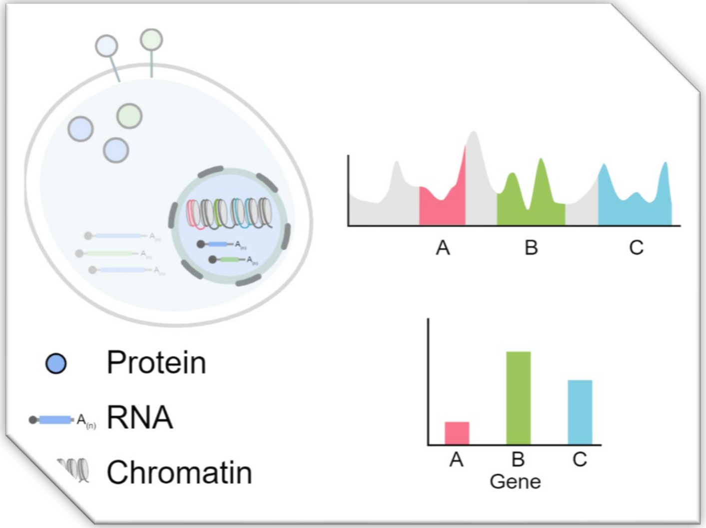
Gromov-Wasserstein Alignment

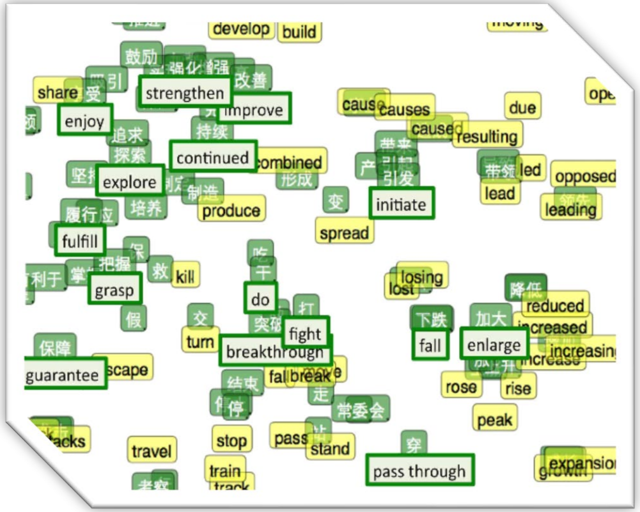
Heterogeneous & Structured Data

Dataset Matching: Various applications require matching heterogeneous **&** structured datasets



[Solomon-Peyré-Kim-Sra ’16]





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

Gromov-Wasserstein Distance

- Datasets as metric measure spaces

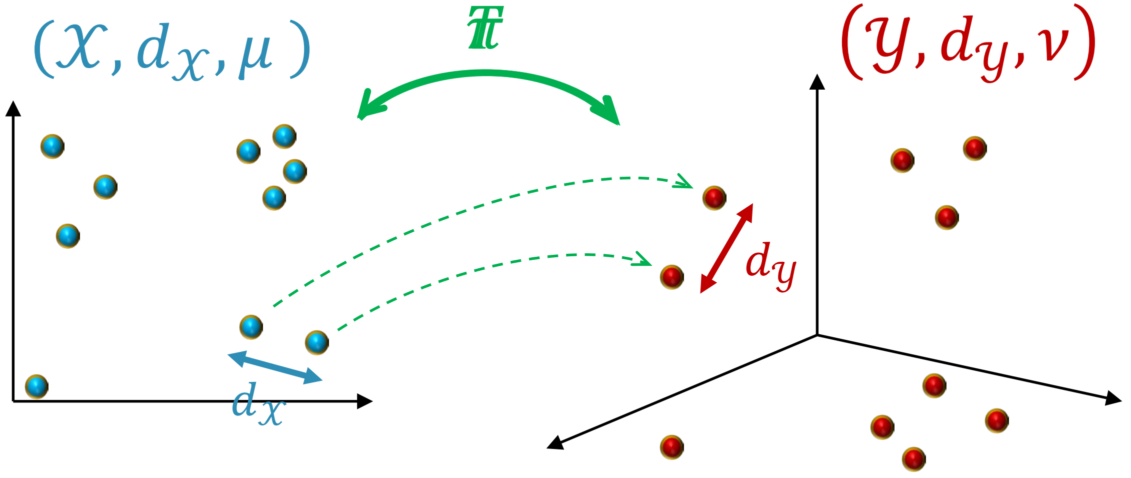
$\Longrightarrow$ $(\mathcal{X}, d_{\mathcal{X}}, \mu)$ & $(\mathcal{Y}, d_{\mathcal{Y}}, \nu)$

- Find matching (transport map) $T: \mathcal{X} \to \mathcal{Y}$

$\Longrightarrow$ $\nu = T_{\#}\mu$ (if $X \sim \mu$ then $T(X) \sim T_{\#}\mu$)

- Preserve distances (minimize distance distortion)

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



(p, q)-Gromov-Wasserstein Distance (Memoli '11)

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

Gromov-Wasserstein Distance

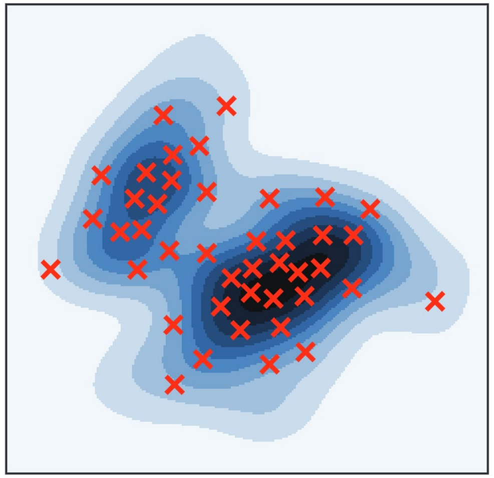
$$\mathrm{D}_{p,q}(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \left(\mathbb{E}_{\substack{(X,Y)\sim\pi \\ (X',Y')\sim\pi}} \left[\left| d_{\mathcal{X}}(X,X')^q - d_{\mathcal{Y}}(Y,Y')^q \right|^p \right] \right)^{1/p}$$

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

- **Finiteness:** $\mathrm{D}_{p,q}(\mu,\nu) < \infty \ \forall \mu,\nu$ with $\mathbb{E}_{(X,X')\sim\mu\otimes\mu}[d_{\mathcal{X}}(X,X')^{pq}] < \infty$ **&** resp. for ν
- **Identification:** $\mathrm{D}_{p,q}(\mu,\nu) = 0 \iff \exists$ isometry $T: \mathcal{X} \to \mathcal{Y}$ with $T_{\#}\mu = \nu$ (invariances)
- **Metric:** Metrizes space of equivalence classes of mm spaces with finite size

Estimation from Data

Question: μ, ν are unknown; we sample $X_1, \dots, X_n \sim \mu$ **&** $Y_1, \dots, Y_n \sim \nu$



- **Empirical measures:** $\hat{\mu}_n \coloneqq \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ and $\hat{\nu}_n \coloneqq \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$

$\Longrightarrow$ Can we approximate $\mathrm{D}_{p,q}(\mu,\nu) \approx \mathrm{D}_{p,q}(\hat{\mu}_n,\hat{\nu}_n)$?

Asymptotic Ans: **Yes!** For μ, ν w/ finite pq–size, $\mathrm{D}_{p,q}(\hat{\mu}_n,\hat{\nu}_n) \to \mathrm{D}_{p,q}(\mu,\nu)$ a.s. [Mémoli '11]

Non-Asymptotic Regime: What is the **rate** at which $\mathbb{E}\left[\left|\mathrm{D}_{p,q}(\mu,\nu) - \mathrm{D}_{p,q}(\hat{\mu}_n,\hat{\nu}_n)\right|\right]$ decays?

🚫 **Open question:** Statistical (sample complex.) **&** computational (time complex.) implications

Duality for Quadratic GW Distance

Setting: (2,2)-GW btw $(\mathbb{R}^{d_x}, \|\cdot\|, \mu)$ and $(\mathbb{R}^{d_y}, \|\cdot\|, \nu)$ with $M_4(\mu) \coloneqq \int \|x\|^4 d\mu(x), M_4(\nu) < \infty$

$$\mathrm{D}(\mu,\nu)^2 = \inf_{\pi\in\Pi(\mu,\nu)} \iint \left| \|x-x'\|^2 - \|y-y'\|^2 \right|^2 d\pi\otimes\pi$$

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

$$\mathrm{D}(\mu,\nu)^2 = \mathrm{S}_1(\mu,\nu) + \mathrm{S}_2(\mu,\nu)$$

where $$\mathrm{S}_1(\mu,\nu) = \int \|x-x'\|^4 d\mu\otimes\mu + \int \|y-y'\|^4 d\nu\otimes\nu - 4\int \|x\|^2\|y\|^2 d\mu\otimes\nu$$

$$\mathrm{S}_2(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int -4\|x\|^2\|y\|^2 d\pi - 8\sum_{\substack{1\le i\le d_x\\ 1\le j\le d_y}} \left(\int x_i y_j d\pi\right)^2$$

⟹ Derive a dual form for $\mathrm{S}_2(\mu,\nu)$!

Duality for the GW Distance

Approach: Linearize quadratic term using auxiliary variables

$$\mathsf{S}_2(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int -4\|x\|^2\|y\|^2 d\pi - 8\sum_{\substack{1\le i\le d_x\\1\le j\le d_y}} \left(\int x_i y_j d\pi\right)^2$$

$$= \inf_{\pi\in\Pi(\mu,\nu)} \int -4\|x\|^2\|y\|^2 d\pi + 32\sum_{\substack{1\le i\le d_x\\1\le j\le d_y}} \inf_{-\frac{M_{\mu,\nu}}{2}\le a_{ij}\le\frac{M_{\mu,\nu}}{2}} \left(a_{ij}^2 - \int a_{ij}x_i y_j d\pi\right)$$

$$= \inf_{\mathbf{A}\in\mathcal{D}_{M_{\mu,\nu}}} 32\|\mathbf{A}\|_{\mathrm{F}}^2 + \inf_{\pi\in\Pi(\mu,\nu)} \int \underbrace{\left(-4\|x\|^2\|y\|^2 - 32x^{\mathrm{T}}\mathbf{A}y\right)}_{=:\, c_{\mathbf{A}}(x,y)} d\pi = \mathrm{OT}_{c_{\mathbf{A}}}(\mu,\nu)$$

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

Theorem (Zhang-G.-Mroueh-Sriperumbudur '22)

$$\mathsf{S}_2(\mu,\nu) = \inf_{\mathbf{A}\in\mathcal{D}_{M_{\mu,\nu}}} 32\|\mathbf{A}\|_{\mathrm{F}}^2 + \mathrm{OT}_{c_{\mathbf{A}}}(\mu,\nu)$$

Sample Complexity of GW: Upper Bound

Theorem (Zhang-G.-Mroueh-Sriperumbudur '22)

Let $(\mu,\nu) \in \mathcal{P}(\mathbb{R}^{d_x}) \times \mathcal{P}(\mathbb{R}^{d_y})$ have compact support with diameter bounded by $R > 0$. Then

$$\mathbb{E}\left[\left|\mathrm{D}(\mu,\nu)^2 - \mathrm{D}(\hat{\mu}_n,\hat{\nu}_n)^2\right|\right] \lesssim_{d_x,d_y,R} \underbrace{R^4 n^{-\frac{1}{2}}}_{\text{S}_1 \text{ rate + centering bias}} + \underbrace{(1+R^4) n^{-\frac{2}{(d_x \wedge d_y)\vee 4}} (\log n)^{\mathbb{1}_{\{d_x \wedge d_y = 4\}}}}_{\text{S}_2 \text{ rate}}$$

Comments:

- **Optimality:** These rates are sharp!
- **Data dimension:** Rate depends on smaller dimension (but curse of dimensionality occurs)
- **Comparison to OT:** Rate matches best known for OT
- **One-sample:** When only μ is estimated

Sample Complexity of GW: Proof Outline

Decomposition: Split D^2 into $\mathrm{S}_1 + \mathrm{S}_2$ by centering empirical measures

$$\mathbb{E}\left[\left|\mathrm{D}(\mu,\nu)^2 - \mathrm{D}(\hat{\mu}_n,\hat{\nu}_n)^2\right|\right] \leq \mathbb{E}[|\mathrm{S}_1(\mu,\nu) - \mathrm{S}_1(\hat{\mu}_n,\hat{\nu}_n)|] + \mathbb{E}[|\mathrm{S}_2(\mu,\nu) - \mathrm{S}_2(\hat{\mu}_n,\hat{\nu}_n)|] + \frac{R^4}{\sqrt{n}}$$

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

S_2 Analysis: Hinges on dual form + regularity analysis of optimal potentials

Sample Complexity of GW: Proof Outline

S_2 Analysis: Invoke duality with radius $M = R^2$

1. **OT reduction:** $\mathbb{E}[|S_2(\mu,\nu) - S_2(\hat{\mu}_n,\hat{\nu}_n)|] \le \mathbb{E}\left[\sup_{\mathbf{A}\in\mathcal{D}_M} \left|\mathrm{OT}_{c_{\mathbf{A}}}(\mu,\nu) - \mathrm{OT}_{c_{\mathbf{A}}}(\hat{\mu}_n,\hat{\nu}_n)\right|\right]$ ⊛

2. **Dual potentials:** $\forall \mathbf{A} \in \mathcal{D}_M$, $\varphi_{\mathbf{A}}$ is concave and $\|\varphi_{\mathbf{A}}\|_{\mathrm{Lip}} \vee \|\varphi_{\mathbf{A}}\|_\infty \lesssim R^4\sqrt{d_x d_y}$ (resp. $\psi_{\mathbf{A}}$)

3. **Empirical processes:** $\mathcal{F}_R := \left\{\varphi: \mathbb{R}^{d_x} \to \mathbb{R}: \text{ concave, } \|\varphi\|_{\mathrm{Lip}} \vee \|\varphi\|_\infty \lesssim R^4\sqrt{d_x d_y}\right\}$ **&** $\mathcal{G}_R$

$$⊛ \le \mathbb{E}\left[\sup_{\varphi\in\cup_{\mathbf{A}}\mathcal{F}_{\mathbf{A}}} |(\mu-\hat{\mu}_n)\varphi|\right] + \mathbb{E}\left[\sup_{\psi\in\cup_{\mathbf{A}}\mathcal{G}_{\mathbf{A}}} |(\nu-\hat{\nu}_n)\psi|\right]$$

$$\le \mathbb{E}\left[\sup_{\varphi\in\mathcal{F}_R} |(\mu-\hat{\mu}_n)\varphi|\right] + \mathbb{E}\left[\sup_{\psi\in\mathcal{G}_R} |(\nu-\hat{\nu}_n)\psi|\right] \lesssim_{R,d_x,d_y} n^{-\frac{2}{d_x}} + n^{-\frac{2}{d_y}} \le n^{-\frac{2}{d_x \vee d_y}}$$

OT duality

Regularity

Covering bound: $\log N(\epsilon, \mathcal{F}_R, \|\cdot\|_\infty) \lesssim_d \epsilon^{-d_x/2}$

Sample Complexity of GW: Proof Outline

S_2 Analysis: Invoke duality with radius $M = R^2$

Assume $d_x < d_y$

1. **OT reduction:** $\mathbb{E}[|S_2(\mu,\nu) - S_2(\hat{\mu}_n,\hat{\nu}_n)|] \leq \mathbb{E}\left[\sup_{\mathbf{A}\in\mathcal{D}_M} \left|\mathrm{OT}_{c_{\mathbf{A}}}(\mu,\nu) - \mathrm{OT}_{c_{\mathbf{A}}}(\hat{\mu}_n,\hat{\nu}_n)\right|\right]$ ⊛

2. **Dual potentials:** $\forall \mathbf{A} \in \mathcal{D}_M$, $\varphi_{\mathbf{A}}$ is concave and $\|\varphi_{\mathbf{A}}\|_{\mathrm{Lip}} \vee \|\varphi_{\mathbf{A}}\|_\infty \lesssim R^4\sqrt{d_x d_y}$ (resp. $\psi_{\mathbf{A}}$)

3. **Empirical processes:** $\mathcal{F}_R := \left\{\varphi: \mathbb{R}^{d_x} \to \mathbb{R}: \text{ concave, } \|\varphi\|_{\mathrm{Lip}} \vee \|\varphi\|_\infty \lesssim R^4\sqrt{d_x d_y}\right\}$ & ~~$\mathcal{G}_R$~~

$(\varphi_{\mathbf{A}}, \varphi_{\mathbf{A}}^c)$ are optimal

$\varphi^c(y) := \inf_x c_{\mathbf{A}}(x,y) - \varphi(x)$

$$⊛ \leq \mathbb{E}\left[\sup_{\varphi\in\cup_{\mathbf{A}}\mathcal{F}_{\mathbf{A}}} |(\mu-\hat{\mu}_n)\varphi|\right] + \mathbb{E}\left[\sup_{\psi\in\cup_{\mathbf{A}}\mathcal{F}_{\mathbf{A}}^c} |(\nu-\hat{\nu}_n)\psi|\right]$$

$$\leq \mathbb{E}\left[\sup_{\varphi\in\mathcal{F}_R} |(\mu-\hat{\mu}_n)\varphi|\right] + \mathbb{E}\left[\sup_{\psi\in\mathcal{F}_R^c} |(\nu-\hat{\nu}_n)\psi|\right] \lesssim_{R,d_x,d_y} n^{-\frac{2}{d_x}} + n^{-\frac{2}{d_x}} \leq n^{-\frac{2}{d_x \wedge d_y}}$$

LCA principle [Hundrieser et al. '22]: $N(\epsilon,\mathcal{F},\|\cdot\|_\infty) = N(\epsilon,\mathcal{F}^c,\|\cdot\|_\infty) \implies \log N(\epsilon,\mathcal{F}_R^c,\|\cdot\|_\infty) \lesssim_d \epsilon^{-d_x/2}$

Sample Complexity of GW: Lower Bound

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

For $\mathcal{X} \subseteq \mathbb{R}^{d_x}$ and $\mathcal{Y} \subseteq \mathbb{R}^{d_y}$ with diameter at most R and any n sufficiently large, we have

$$\sup_{(\mu,\nu)\in\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})} \mathbb{E}\left[\left|\mathrm{D}(\mu,\nu)^2 - \mathrm{D}(\hat{\mu}_n,\hat{\nu}_n)^2\right|\right] \gtrsim_{d_x,d_y,R} n^{-\frac{2}{(d_x\wedge d_y)\vee 4}}$$

Proof Idea:

- **Wasserstein Procrustes Lemma:** $\mathrm{D}(\mu,\nu) \gtrsim_{\lambda_{\min}(\boldsymbol{\Sigma}_\mu),\lambda_{\min}(\boldsymbol{\Sigma}_\nu)} \inf_{\mathbf{U}\in O(d)} \mathrm{W}_2(\mu, \mathbf{U}_\# \nu)$
- **Construction:** $\mu = \mathrm{Unif}\big(B_d(0,1)\big)$ **&** $\nu = \mathrm{Unif}\big(B_d(0,2)\big)$
- **Lower bound:** $\mathbb{E}\left[\inf_{\mathbf{U}\in O(d)} \mathrm{W}_2(\hat{\mu}_n, \mathbf{U}_\#\mu)\right] \gtrsim \inf_{\mathbf{U}\in O(d)} \mathbb{E}[\mathrm{W}_2(\hat{\mu}_n, \mathbf{U}_\#\mu)] \geq \mathbb{E}[\mathrm{W}_1(\hat{\mu}_n,\mu)] \gtrsim_d n^{-\frac{1}{d}}$ [Dudley' 69]

Computation via Entropic Gromov-Wasserstein

GW is QAP: $\mathrm{D}_{p,q}\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \frac{1}{n}\sum_{i=1}^n \delta_{y_i}\right)^p = \frac{1}{n^2}\min_{\sigma\in S_n}\sum_{i,j=1}^n \left|d_{\mathcal{X}}(x_i,x_j)^q - d_{\mathcal{Y}}(y_{\sigma(i)},y_{\sigma(j)})^q\right|^p$

🚫 Quadratic assignment problem (non-convex) [Commander '05] $\Longrightarrow$ **NP complete**

Entropic Gromov-Wasserstein Distance (Peyré-Cuturi-Solomon '16)

$$\mathrm{S}_\epsilon(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \mathbb{E}_{\pi\otimes\pi}\left[\left|\|X-X'\|^2 - \|Y-Y'\|^2\right|^2\right] + \epsilon \mathrm{D}_{\mathrm{KL}}(\pi\|\mu\otimes\nu)$$

1. **Algorithms:** Heuristic methods [Peyré-Cuturi-Solomon '16], [Solomon-Peyré-Kim-Sra '16]
2. **Approximation:** $\left|\mathrm{D}(\mu,\nu)^2 - \mathrm{S}_\epsilon(\mu,\nu)\right| \lesssim_{d_x,d_y} \epsilon\log(1/\epsilon)$ [Zhang-G.-Mroueh-Sriperumbudur '23]
3. **Estimation:** $\mathbb{E}[|\mathrm{S}_\epsilon(\mu,\nu) - \mathrm{S}_\epsilon(\hat{\mu}_n,\hat{\nu}_n)|] \asymp_{d_x,d_y,\epsilon} n^{-1/2}$ [————]

From Stability Analysis to Convexity

$$\mathrm{S}_{\epsilon}(\mu,\nu) = \mathrm{S}_1(\mu,\nu) + \min_{\mathbf{A}\in\mathcal{D}_M}\Big\{ \underbrace{32\|\mathbf{A}\|_{\mathrm{F}}^2 + \mathrm{EOT}_{\epsilon,c_{\mathbf{A}}}(\mu,\nu)}_{=:\,\Phi(\mathbf{A})} \Big\}$$

Analysis:

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

Theorem (Rioux-G.-Kato '23)

1. Φ is strictly convex whenever $\epsilon > 16\sqrt{M_4(\mu)M_4(\nu)}$
2. Φ is L-smooth on $\mathcal{D}_M$ with $L \leq 64 \vee \left(32^2\epsilon^{-1}\sqrt{M_4(\mu)M_4(\nu)} - 64\right)$

Accelerated First-Order Inexact Oracle Methods

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

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

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

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

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

$\Longrightarrow$ 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 $\tau_k = \frac{2}{k+3}$

```
k ← 0
A_0 ← 0
G_0 ← D̃Φ_[A_0]
W_0 ← α_0 G_0
while stopping condition is not met do
    B_k ← (M/2) sign(A_k − L^{-1} G_k) min( (2/M) |A_k − L^{-1} G_k| , 1)
    C_k ← (M/2) sign(−L^{-1} W_k) min( (2/M) |L^{-1} W_k| , 1)
    A_{k+1} ← τ_k C_k + (1 − τ_k) B_k
    G_{k+1} ← D̃Φ_[A_{k+1}]
    W_{k+1} ← W_k + α_{k+1} G_{k+1}
    k ← k + 1
return B_k
```

Global Convergence Guarantees (Convex)

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

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

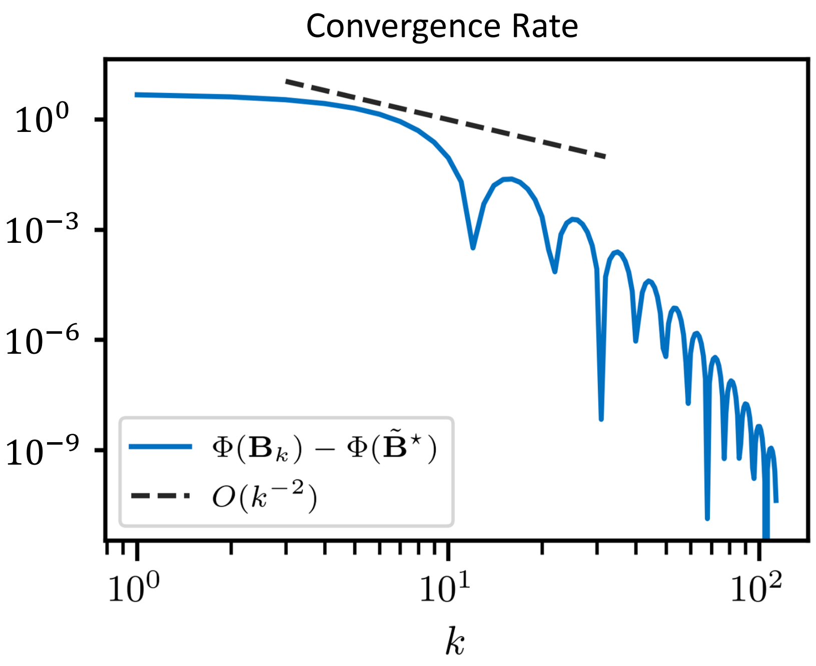
$$\Phi(\mathbf{B}_k) - \Phi(\mathbf{B}_*) \le \frac{2L\|\mathbf{B}_*\|_{\mathrm{F}}^2}{(k+1)(k+2)} + O(M\delta)$$

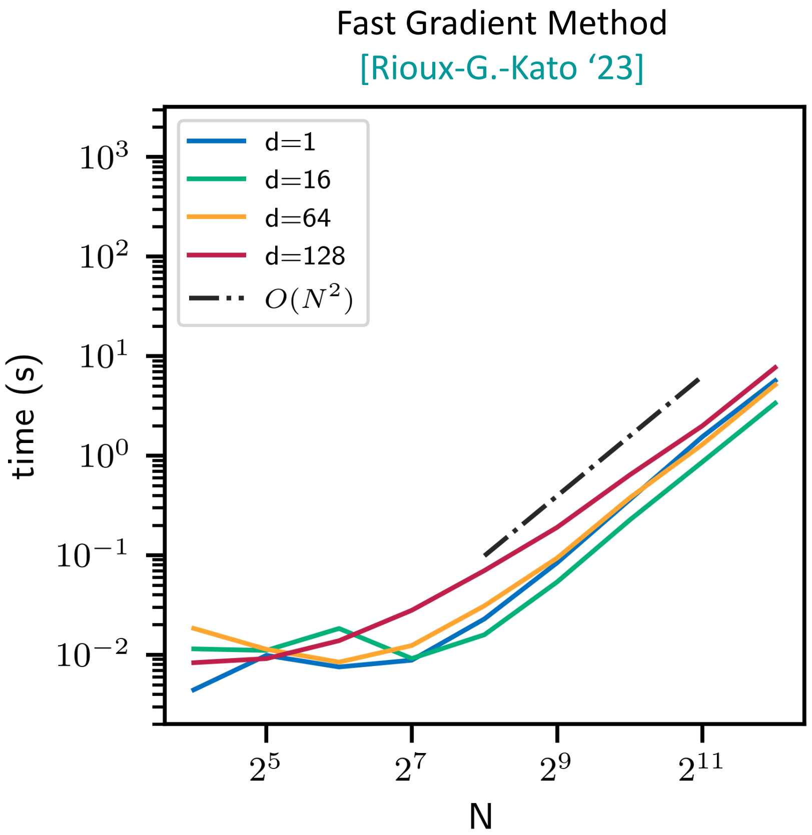
Comments:

- **Optimality:** Optimal complexity of $O\left(1/k^2\right)$ for smooth constrained opt. [Nesterov '03]
- **Non-convex regime:** Via smooth non-convex opt. with inexact oracle [Ghadimi-Lan '16]

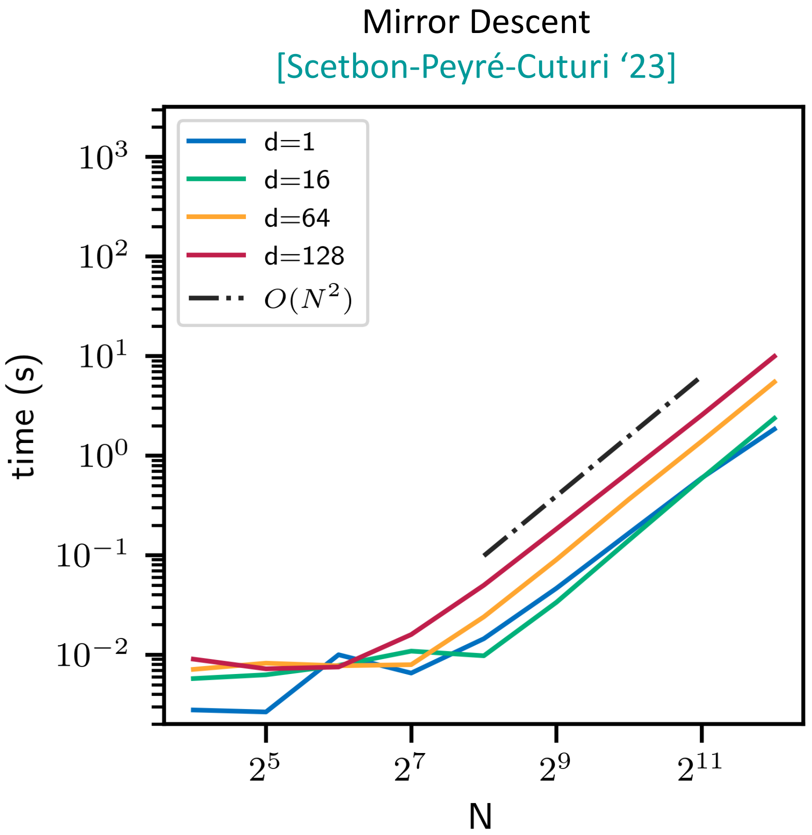
  ↳ Adapts to convexity of Φ (yields improved rates if convex)

Numerical Results





Time = iteration × Sinkhorn
$= k \times O(N^2)$

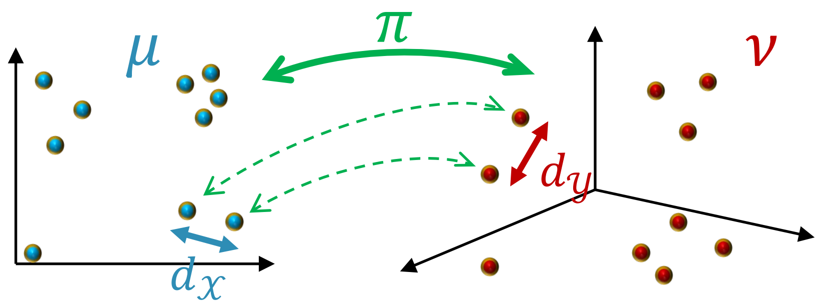


Time = iteration × cost update
$= k \times d \times O(N^2)$

Summary

Gromov-Wasserstein Distance: Quantifies discrepancy between mm spaces

- Alignment of heterogeneous datasets
- Foundational statistical & computational questions open



Contributions: Duality, empirical rates, and algorithms

- Dual form that connects to OT
- Sharp sample complexity for quadratic GW
- First algorithms w/ convergence rates for entropic GW
- Duality and empirical rates for EGW

Thank you!

[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