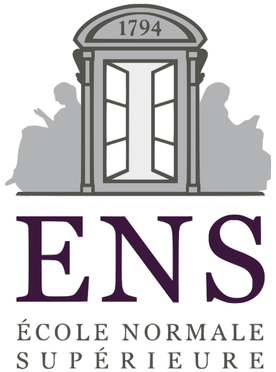
Linear-Time Gromov Wasserstein Distances using Low Rank Couplings and Costs

M. Scetbon G. Peyré M. Cuturi

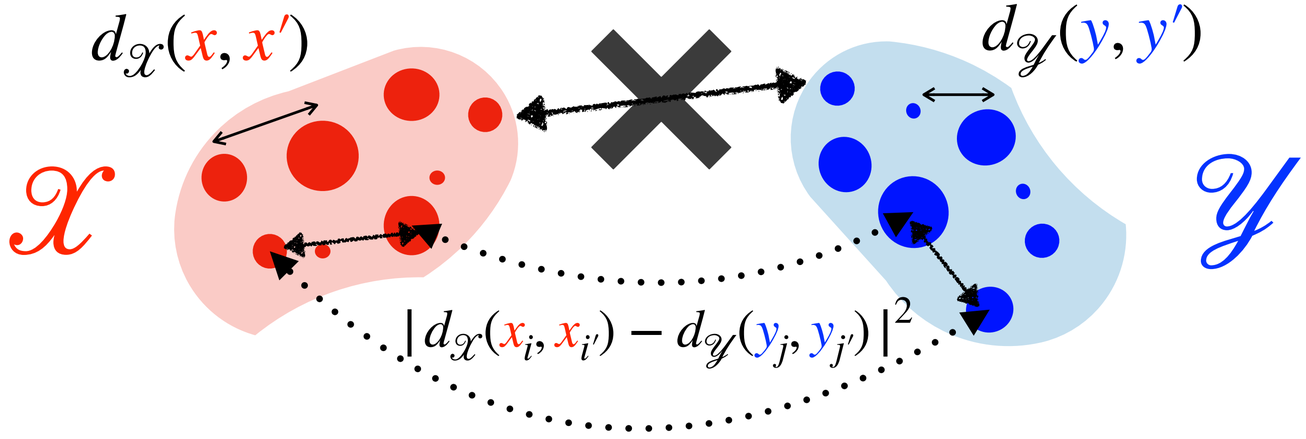


Thirty-ninth International Conference on Machine Learning

Gromov-Wasserstein

How to align points across two incomparable point clouds?

Not Comparable directly



Discrete Gromov-Wasserstein:

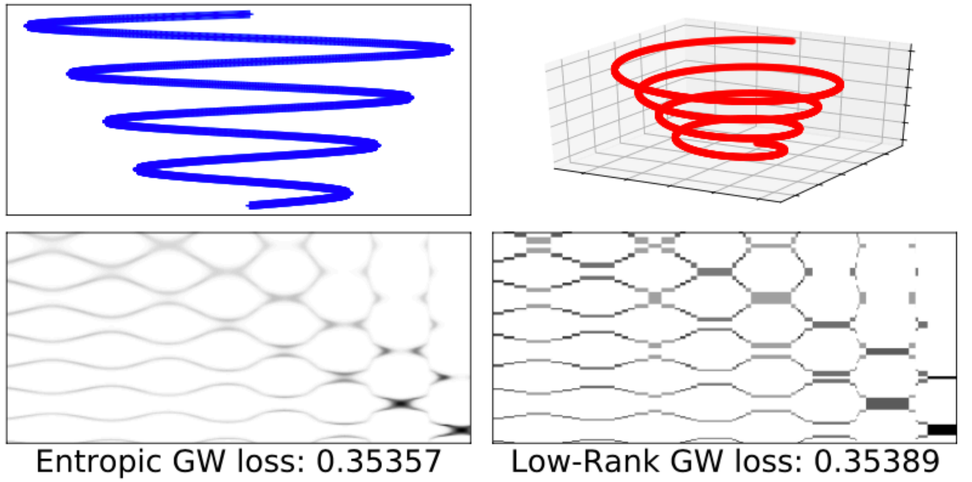
- Discrete Distributions: $\mu = \sum_{i=1}^{n} a_i \, \delta_{x_i} \in \mathcal{M}_1^+(\mathcal{X})$, $\nu = \sum_{j=1}^{m} b_j \, \delta_{y_j} \in \mathcal{M}_1^+(\mathcal{Y})$
- Set of Couplings: $\Pi_{a,b} = \left\{ P \in \mathbb{R}_+^{n \times m} \text{ s.t. } P\mathbf{1}_m = a, P^T\mathbf{1}_n = b \right\}$
- Cost Matrices: $A = (d_{\mathcal{X}}(x_i, x_{i'}))_{i,i'}$, $B = (d_{\mathcal{Y}}(y_j, y_{j'}))_{j,j'}$

$$GW_c((a, A), (b, B)) = \min_{P \in \Pi_{a,b}} \mathcal{Q}_{A,B}(P) := \sum_{i,i',j,j'} |A_{i,i'} - B_{j,j'}|^2 P_{i,j} P_{i',j'}$$

Hard to solve in practice:

- GW is a ***non convex*** quadratic problem
- GW is ***NP-hard*** in general

→ Need for a fast solver approximating the GW cost



Prior Art: Entropic Regularization

Shannon Entropy: $\mathrm{H}(P) = -\sum_{i,j} P_{i,j}(\log(P_{i,j}) - 1)$

$$GW_{c,\epsilon}((a,A),(b,B)) = \min_{P \in \Pi_{a,b}} \mathcal{Q}_{A,B}(P) - \epsilon H(P)$$

Mirror Descent Scheme:

At each iteration:

- Update the cost matrix: $C = -4APB \quad \in \mathcal{O}(nm(n+m))$
- Update the kernel matrix: $K = e^{-C/\epsilon} \quad \in \mathcal{O}(nm)$
- Solve Entropic OT using Sinkhorn: $P = \arg\min_{P \in \Pi_{a,b}} \mathrm{KL}(P||K) \quad \in \mathcal{O}(nm)$

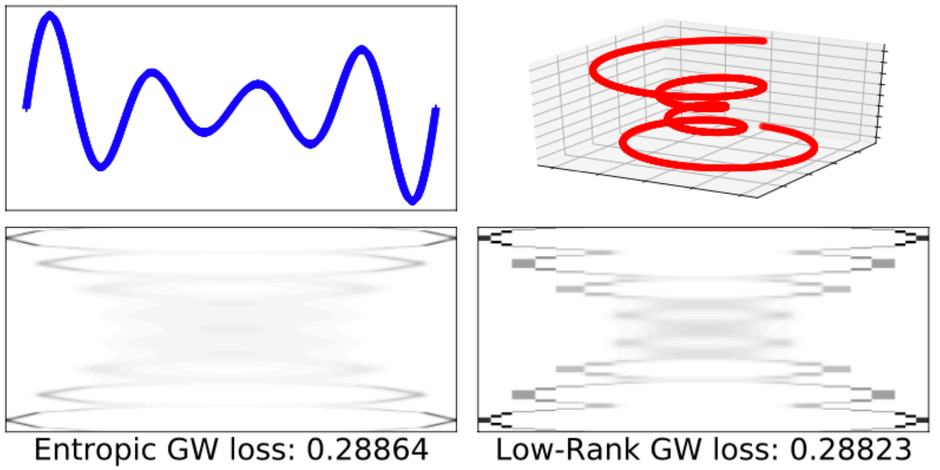
} Total Complexity per iteration: $\mathcal{O}(nm(n+m))$

In this work, we propose *instead* to directly constraint the coupling to admit a low-NN rank

Low-Rank Gromov-Wasserstein

NN rank : $\mathrm{rk}_+(M) := \min \left\{ q \,|\, M = \sum_{i=1}^{q} R_i, \forall i, \mathrm{rk}(R_i) = 1, R_i \geq 0 \right\}$

Low-NN Rank Couplings: $\Pi_{a,b}(r) := \left\{ P \in \Pi_{a,b} \text{ s.t. } \mathrm{rk}_+(P) \leq r \right\}$



Definition of Low-rank Gromov-Wasserstein

$$\mathrm{LGW}_{c,r}((a, A), (b, B)) := \min_{P \in \Pi_{a,b}(r)} \mathcal{Q}_{A,B}(P)$$

Characterization of Low-NN Rank Couplings:

$$\Pi_{a,b}(r) = \{ P \in \mathbb{R}_+^{n \times m} \,|\, P = Q \mathrm{Diag}(1/g) R^T,\ Q \in \Pi_{a,g}, R \in \Pi_{b,g},\ g > 0 \text{ and } g \in \Delta_r \}$$

Reparametrization of LGW

$$\mathrm{LGW}_r((a,A),(b,B)) = \min_{(Q,R,g)\in\mathscr{C}_1(a,b,r)\cap\mathscr{C}_2(r)} \mathscr{Q}_{A,B}(Q\mathrm{Diag}(1/g)R^T)$$

where
$$\begin{cases} \mathscr{C}_1(a,b,r) := \left\{(Q,R,g)\in\mathbb{R}_+^{n\times r}\times\mathbb{R}_+^{m\times r}\times(\mathbb{R}_+^*)^r \text{ s.t. } Q\mathbf{1}_r = a, R\mathbf{1}_r = b\right\} \\ \mathscr{C}_2(r) := \left\{(Q,R,g)\in\mathbb{R}_+^{n\times r}\times\mathbb{R}_+^{m\times r}\times(\mathbb{R}_+)^r \text{ s.t. } Q^T\mathbf{1}_n = R^T\mathbf{1}_m = g\right\} \end{cases}$$

Mirror Descent Scheme:

At each iteration:

$\in \mathcal{O}((n^2+m^2)r)$

- Update the cost matrices: $C_1 = -AQD_{1/g}$, $C_2 = R^TB$, $C_3 = \mathscr{D}(Q^TC_1C_2R)$ $\in \mathcal{O}((n+m)r^2)$

  where $\mathscr{D}$ is the operator extracting the diagonal of a squared matrix

- Update the kernel matrices: $K_1 = Q \odot e^{4\gamma C_1C_2RD_{1/g}}$, $K_2 = R \odot e^{4\gamma C_2^TC_1^TQD_{1/g}}$, $K_3 = g \odot e^{-4\gamma C_3/g^2}$

  where D_a is the operator transforming a vector a into a diagonal matrix $\in \mathcal{O}((n+m)r)$

- Solve the convex Barycenter problem using Dykstra: $(Q,R,g) = \arg\min_{(Q,R,g)\in\mathscr{C}_1(a,b,r)\cap\mathscr{C}_2(r)} \mathrm{KL}((Q,R,g)||(K_1,K_2,K_3))$

→ Total Complexity per iteration: $\mathcal{O}((n^2+m^2)r) \ll \mathcal{O}((n+m)nm)$ as soon as $r \ll \min(n,m)$

From a Quadratic Solver to a Linear Solver

Remark: The only operations which remains quadratic in the MD scheme described before is the updates of the cost matrices C_1 and C_2.

→ By assuming that A and B admit low-rank structures, we obtain a linear time algorithm with respect to the number of samples.

Low-rank cost matrices:

If $A = A_1 A_2^T$ and $B = B_1 B_2^T$ with $A_1, A_2 \in \mathbb{R}^{n \times d}$ and $B_1, B_2 \in \mathbb{R}^{m \times d'}$ with $d, d' \ll \min(n, m)$ then updating the cost matrices can be done in linear time:

$$C_1 = -A_1 A_2^T Q D_{1/g} \in \mathcal{O}(nrd) \quad \text{and} \quad C_2 = R^T B_1 B_2^T \in \mathcal{O}(mrd')$$

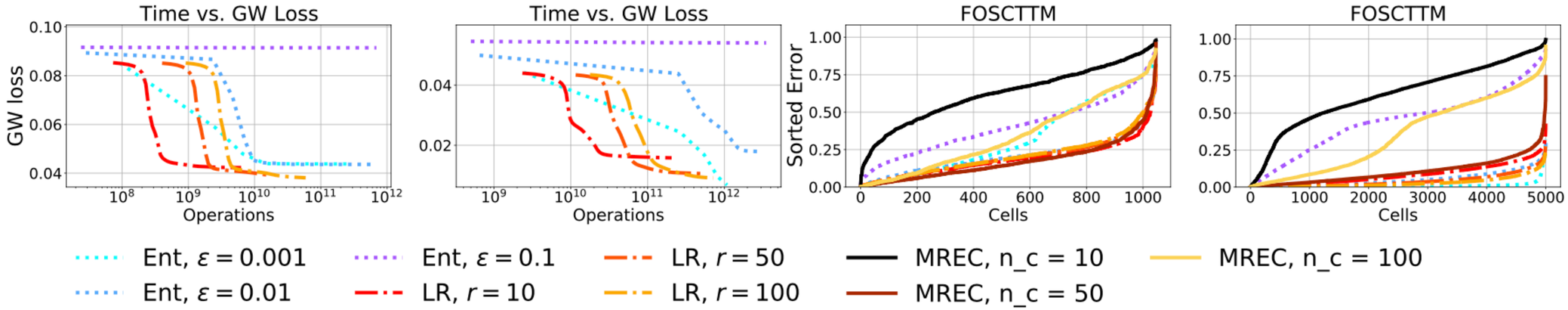
→ Total Complexity per iteration: $\mathcal{O}(r(nd + md')) \ll \mathcal{O}((n^2 + m^2)r)$ as soon as $d, d' \ll \min(n, m)$

Example: The squared Euclidean distance, or more generally any distance matrix.

Other results

We provide a quantitive bound and show the ***non-asymptotic stationary convergence*** of our algorithm. Roughly speaking, our algorithm requires $\mathcal{O}(1/\delta)$ iterations for a precision of δ.

Experiments



Problem: We consider the single-cell alignment problem where we have access to two representations of the same cells. These representations are not directly comparable, and therefore we apply GW to recover the true matching.

Thank you

Results:

- We observe that our method is able to obtain similar GW cost (and ever better) while being order of magnitude faster than the Entropic approach.

- In addition, the quality of the coupling (measured by the FOSCTTM) is comparable to the one obtained by the entropic method.