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

Meyer SCETBON, Gabriel PEYRÉ, Marco CUTURI

Overview

Problem: How to obtain an efficient procedure to compute an approximation of the Gromov Wasserstein cost?

Contributions:

- We show first that a low-rank factorization (or approximation) of the two input cost matrices that define GW, one for each measure, can be exploited to lower the complexity of the entropic GW problem from cubic to quadratic.
- We show next, independently, that by imposing a low-nonnegative rank on the couplings involved in the GW problem we obtain a solver only requiring $\mathcal{O}(n^2)$ operations with no prior assumption on input cost matrices.
- Finally, we show that both low-rank assumptions (on costs and couplings) can be combined to shave yet another factor and reach a linear $\mathcal{O}(n)$ GW approximation.
- We show *experimentally* the efficiency of our approach.

Entropic Gromov Wasserstein

Discrete Distributions: $\mu = \sum_{i=1}^n a_i\,\delta_{x_i}$ and $\nu = \sum_{j=1}^m b_j\,\delta_{y_j}$

Cost matrices: $A = [d_{\mathcal{X}}(x_i, x_j)]_{1\leq i,j\leq n}$, $B = [d_{\mathcal{Y}}(y_i, y_j)]_{1\leq i,j\leq m}$

Definition of Entropic Gromov Wasserstein:

$$GW_\varepsilon((a, A), (b, B)) := \min_{\substack{P \in \mathbb{R}_+^{n\times m}\\ P\mathbf{1}_m = a,\, P^T\mathbf{1}_n = b}} \mathcal{E}_{A,B}(P) - \varepsilon \mathrm{H}(P)$$

$$\text{where } \mathcal{E}_{A,B}(P) := \sum_{i,j,i',j'} |A_{i,i'} - B_{j,j'}|^2 P_{i,j} P_{i',j'}$$

Shannon entropy

Mirror Descent Scheme:

Init: $a, A, b, B, \varepsilon, P$

for $t = 0, \dots, T$:

$C = -4APB$ ⟵ Update the cost: $\mathcal{O}(nm(n+m))$

$K = \exp(-C/\varepsilon)$ ⟵ Update the kernel: $\mathcal{O}(nm)$

$P = arg \min_{P\geq 0,\, P\mathbf{1}_m = a,\, P^T\mathbf{1}_n = b} \mathrm{KL}(P, K)$ ⟵ Solve the entropic OT: $\mathcal{O}(nm)$

Low-rank Costs

Idea: Replace A by $\tilde{A} = A_1 A_2^T$ where $(A_1, A_2) \in (\mathbb{R}_*^+)^{n\times d} \times (\mathbb{R}_*^+)^{n\times d}$

Replace B by $\tilde{B} = B_1 B_2^T$ where $(B_1, B_2) \in (\mathbb{R}_*^+)^{n\times d'} \times (\mathbb{R}_*^+)^{n\times d'}$

⟶ Updating the cost $C = -4A_1A_2^T P B_1 B_2^T$ requires now $nm(d + d') + dd'(n + m)$ operations.

Examples:

- SE distance: $A = \left[\|x_i - x_j\|_2^2\right]_{i,j} = A_1 A_2^T$ with $z = (X^{\odot 2})^T \mathbf{1}_d$,

$A_1 = [z, \mathbf{1}_n, -2X^T] \in \mathbb{R}^{n\times(d+2)}$, $A_2 = [\mathbf{1}_n, z, X^T] \in \mathbb{R}^{n\times(d+2)}$

computable in $\mathcal{O}(n)$

- General Distance Matrix: $\|A - A_1A_2^T\|_F^2 \leq \|A - C_d\|_F^2 + \gamma\|A\|_F^2$

Low-Rank Couplings

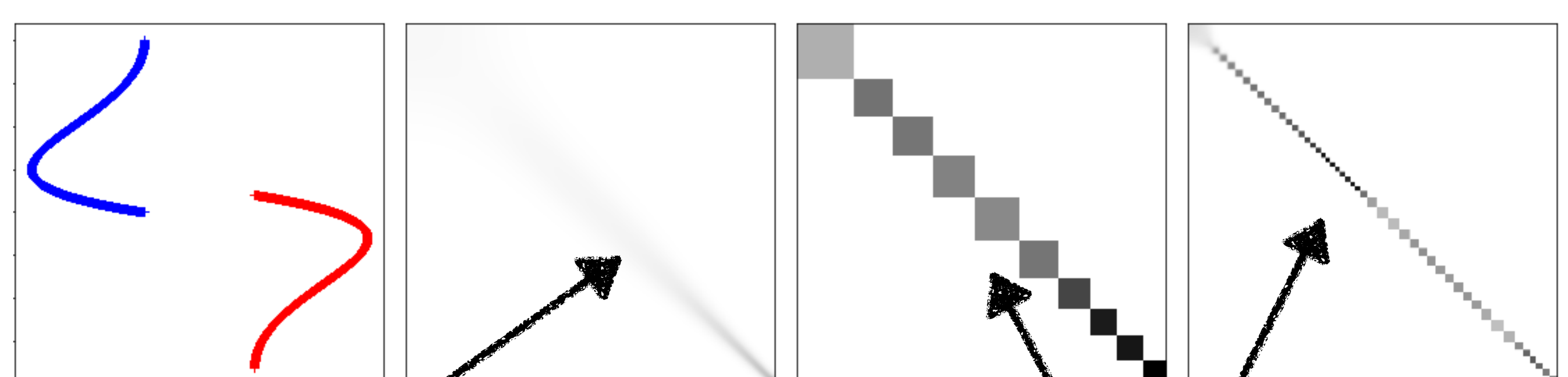
NN-rank: $\mathrm{rk}_+(M) := \min\left\{ q \,|\, M = \sum_{i=1}^q R_i, \forall i, \mathrm{rk}(R_i) = 1, R_i \geq 0 \right\}$

Low-NN rank couplings:

$$\Pi_{a,b}(r) := \left\{ P \in \mathbb{R}_+^{n\times m} \text{ s.t. } P\mathbf{1}_m = a,\ P^T\mathbf{1}_n = b \text{ and } \mathrm{rk}_+(P) \leq r \right\}$$

Definition of Low-Rank Gromov Wasserstein:

$$\text{GW-LR}_r((a, A), (b, B)) := \min_{P\in\Pi_{a,b}(r)} \mathcal{E}_{A,B}(P)$$



Reparametrization of GW-LR:

$$\text{GW-LR}_r((a, A), (b, B)) = \min_{(Q,R,g)\in\mathcal{C}_1(a,b,r)\cap\mathcal{C}_2(r)} \mathcal{E}_{A,B}(Q\mathrm{Diag}(1/g)R^T)$$

$$\mathcal{C}_1(a, b, r) := \left\{ (Q, R, g) \in \mathbb{R}_+^{n\times r} \times \mathbb{R}_+^{m\times r} \times (\mathbb{R}_+^*)^r \text{ s.t. } Q\mathbf{1}_r = a, R\mathbf{1}_r = b \right\}$$

$$\mathcal{C}_2(r) := \left\{ (Q, R, g) \in \mathbb{R}_+^{n\times r} \times \mathbb{R}_+^{m\times r} \times (\mathbb{R}_+)^r \text{ s.t. } Q^T\mathbf{1}_n = R^T\mathbf{1}_m = g \right\}$$

Mirror-Descent Scheme

Algorithm 2: Low-Rank GW

Inputs: a, A, B, b, r, Q, R, g

for $t = 1, \dots$ **do**

$C_1 \leftarrow -AQ\,\mathrm{diag}(1/g)$
$C_2 \leftarrow R^T B$
} Update the costs: $(n^2 + m^2)r$

$K^{(1)} \leftarrow Q \odot \exp(4\gamma C_1 C_2 R\,\mathrm{diag}(1/g))$
$K^{(2)} \leftarrow R \odot \exp(4\gamma (C_1C_2)^T Q\,\mathrm{diag}(1/g))$
$\omega \leftarrow \mathcal{D}(Q^T C_1 C_2 R)$
} Update the kernels: $\mathcal{O}((n+m)r^2)$

$K^{(3)} \leftarrow g \odot \exp(-4\gamma\omega/g^2)$

$Q, R, g \leftarrow \operatorname{argmin}_{\zeta\in\mathcal{C}_1(a,b,r)\cap\mathcal{C}_2(r)} \mathrm{KL}(\zeta, \mathbf{K})$ Solve the convex Barycenter problem with Dykstra: $\mathcal{O}((n+m)r)$

end

Return: $\mathcal{E}$

Double Low-Rank GW

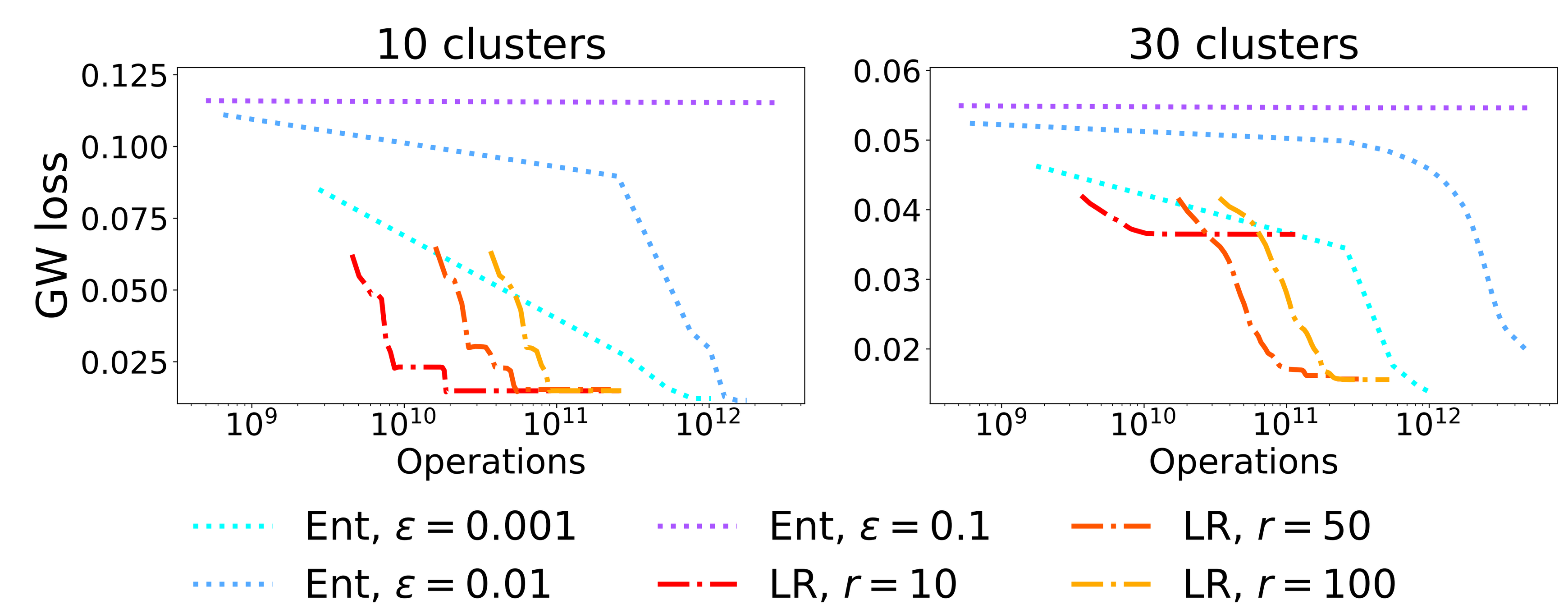
The only steps which remain quadratic are the updates of the costs C_1 and C_2.

⟶ By replacing A by $\tilde{A} = A_1 A_2^T$ and B by $\tilde{B} = B_1 B_2^T$:

$C_1 = -A_1 A_2^T Q \mathrm{Diag}(1/g)$ and $C_2 = -R^T B_2 B_1^T$

$\mathcal{O}(nrd + mrd')$

Experiments



Comparison of the time-accuracy tradeoff between our method and the Entropic GW. We plot $n = 5000$ samples from two isotropic Gaussian Blobs in 10 and 15-D. We observe that our method obtains similar GW loss, while being orders of magnitude faster.