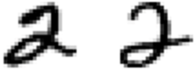
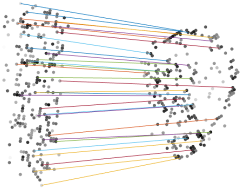
Globally solving the Gromov-Wasserstein problem for point clouds in low dimensional Euclidean spaces

Neural Information Processing Systems 2023

Martin Ryner, Jan Kronkqvist, Johan Karlsson
Division of Numerical Analysis, Optimization and Systems Theory
Department of Mathematics
KTH Royal Institute of Technology, Stockholm, Sweden
martinrr@kth.se

Motivation





Highlights

1. An algorithm that accurately compute Gromov-Wasserstein discrepancy for objects in embedded in low dimensional spaces
2. We prove that the proposed algorithm converges to a global optimal solution.
3. We show that the algorithm produces an optimality certificate in each iteration, in the form of upper- and lower bounds, which informs us of the potential suboptimality if the algorithm is terminated early.

The discrete Gromov-Wasserstein[1] distance ($p = 2$)

$$\min_{\Gamma \in P} \frac{1}{2} \sum_{i,i',j,j'=1}^{n} (d_{\mathcal{X}}(x_i, x_{i'}) - d_{\mathcal{Y}}(y_j, y_{j'}))^2 \Gamma_{i,j} \Gamma_{i',j'}$$

P the set of permutation matrices

$$\min_{\Gamma \in P} \quad -\langle C_x \Gamma, \Gamma C_y \rangle + \frac{1}{2}(\langle C_x, C_x \rangle + \langle C_y, C_y \rangle)$$

$$C_x = [d_{\mathcal{X}}(x_i, x_{i'})]_{i,i'=1}^{n}, C_y = [d_{\mathcal{Y}}(y_j, y_{j'})]_{j,j'=1}^{n}$$

[1] Mémoli, Facundu. "Gromov–Wasserstein Distances and the Metric Approach to Object Matching." Found Comput Math (2011) 11:417–487

Low rank Gromov-Wasserstein discrepancy and QAP

Discrepancy as squared Euclidean distance

$$x_i \in \mathbb{R}^{\ell_x}, y_j \in \mathbb{R}^{\ell_y}, X = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}, Y = \begin{pmatrix} y_1 & y_2 & \dots & y_n \end{pmatrix}$$
$$C_x = (\|x_i - x_j\|_2^2)_{i,j=1}^n = \mathbf{1} m_x^T - 2X^T X + m_x \mathbf{1}^T,$$
$$C_y = (\|y_i - y_j\|_2^2)_{i,j=1}^n = \mathbf{1} m_y^T - 2Y^T Y + m_y \mathbf{1}^T$$

where $m_x = (\|x_1\|^2, \|x_2\|^2, \dots, \|x_n\|^2)^T$, $m_y = (\|y_1\|^2, \|y_2\|^2, \dots, \|y_n\|^2)^T$ then

$$\langle C_x \Gamma, \Gamma C_y \rangle = \langle 2X\Gamma Y^T, 2X\Gamma Y^T \rangle + \langle L, \Gamma \rangle + 2\mathbf{1}^T m_y \mathbf{1}^T m_x,$$

where $L = 2n m_x m_y^T - 4 m_x \mathbf{1}^T Y^T Y - 4 X^T X \mathbf{1} m_y^T$.

$$\min_{\Gamma \in P} \quad -\|2X\Gamma Y^T\|_F^2 - \langle L, \Gamma \rangle + c_0 \tag{1}$$

Relaxation in two steps

Compact set of doubly stochastic matrices

$$\min_{\Gamma \in \overline{P}} \quad -\|2X\Gamma Y^T\|_F^2 - \langle L, \Gamma \rangle + c_0, \tag{2}$$

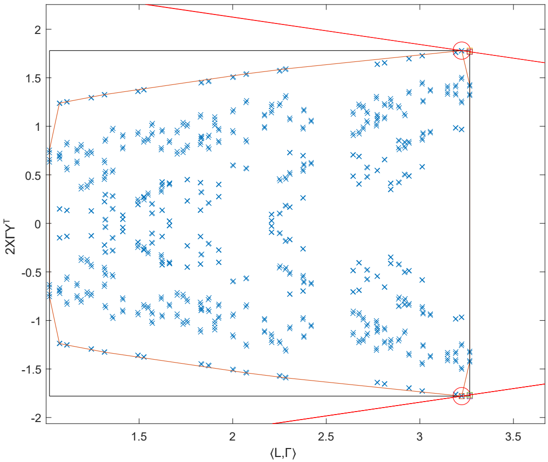
$\overline{P}$ the set of doubly stochastic matrices

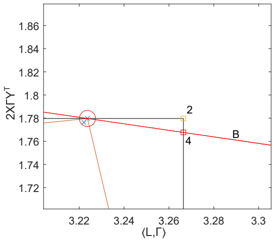
Approximation using a set linear constraints

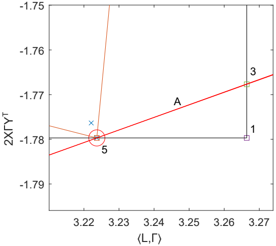
$$\min_{W \in \mathbb{R}^{\ell_x \times \ell_y}, w \in \mathbb{R}} \quad -\|W\|_F^2 - w + c_0 \tag{3a}$$

$$\text{subject to} \quad \langle Z_r, W \rangle + \alpha_r w \leq \beta_r, \quad \text{for} \quad r = 1, \dots, N. \tag{3b}$$

Procedure





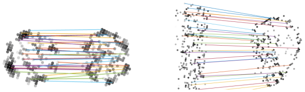


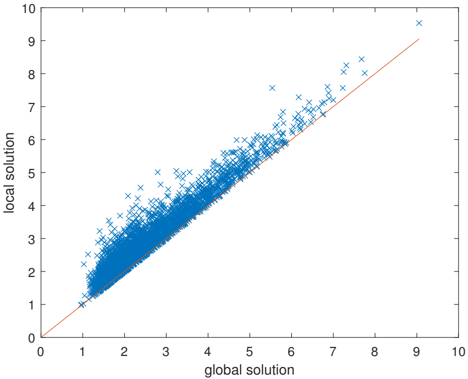
Numerical results

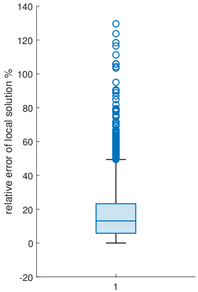
Type	n	ℓ_x, ℓ_y	Rel. error	Proposed [s] Extreme point / B&B	MILP1 [s] [2]	(2) B&B [s]
$\mathcal{U}$	10	2,2	10^{-8}	0.14 (0.07-0.3) / 21 (6-47)	39 (11-58)	0.15 (0.14-0.16)
$\mathcal{U}$	100	2,2	10^{-8}	0.48 (0.3-0.7) / 86 (52-107)	-	25 (19-39)
$\mathcal{U}$	500	2,2	10^{-8}	11 (9-16) / 408 (269-511)	-	-
$\mathcal{U}$	1000	2,2	10^{-8}	69 (54-85) / 576 (389-1059)	-	-
$\mathcal{U}$	2000	2,2	10^{-8}	460 (313-653) / -	-	-
$\mathcal{U}$	10	2,3	10^{-8}	1.8 (1.2-2.4) / 133 (45-296)	105 (49-147)	2.4(1.8-3.4)
$\mathcal{U}$	100	2,3	10^{-8}	278 (99-813) / -	-	172 (133-221)
$\mathcal{U}$	500	2,3	10^{-8}	9568 / -	-	-
$\mathcal{N}_1$	10	2,3	10^{-8}	0.51 (0.39-0.65) / 708 (233-1184)	146 (66-227)	3 (2.6-4.0)
$\mathcal{N}_1$	100	2,3	10^{-8}	86 (20-275) / -	-	95 (73-116)
$\mathcal{N}_1$	500	2,3	10^{-5}	5310!/ -	-	-
$\mathcal{N}_2$	10	3,3	10^{-2}	1.8 (0.7-3.2) / 142 (73-210)	117 (71-163)	0.2(0.1-0.3)
$\mathcal{N}_2$	100	3,3	10^{-2}	36 (22-55)/ -	-	45(36-65)
$\mathcal{N}_2$	500	3,3	10^{-2}	436 (228-862) / -	-	-
$\mathcal{N}_3$	10	3,3	10^{-2}	1.2 (0.5-2.3) / 22 (11-43)	72 (43-94)	0.2(0.1-0.3)
$\mathcal{N}_3$	100	3,3	10^{-2}	7 (5-8)/ 91 (76-111)	-	10 (9-12)
$\mathcal{N}_3$	2000	3,3	10^{-2}	93 (91-100) / 578 (429-691)	-	-

[2] Wang, Yang, et al. "The Rank-One Quadratic Assignment Problem." INFORMS Journal on Computing 33.3 (2021): 979-996.

Comparison with local search[3]







[3] Peyré, Gabriel, Marco Cuturi, and Justin Solomon. "Gromov-wasserstein averaging of kernel and distance matrices." International conference on machine learning. PMLR, 2016.

Thanks!