Partial Gromov-Wasserstein Metric

Yikun Bai[1]   Rocío Díaz Martín*,[2]   **Abihith Kothapalli***[1]   Hengrong Du[3]

Xinran Liu[1]   Soheil Kolouri[1]

[1]Department of Computer Science, Vanderbilt University,
[2]Department of Mathematics, Tufts University,
[3]Department of Mathematics, University of California, Irvine

Motivation

- **Real-world data is often heterogeneous and unbalanced.** Differences in size, structure, or support—due to noise, missing data, or sampling variation—make many practical datasets difficult to compare directly. This motivates the need for robust matching methods that can handle such imperfections.
- **Classical optimal transport (OT) has limitations.** It assumes full mass matching and requires measures to lie on the same domain, making it ill-suited for comparing objects with unequal mass or differing geometries.
- **Partial Gromov-Wasserstein (PGW) offers a principled generalization of classical OT.** PGW allows for partial transport between distributions supported on different metric spaces, making it more robust for real-world, noisy, and incomplete data.

Contribution: We propose the **Partial Gromov-Wasserstein (PGW) problem** and prove that it gives rise to a **metric** between mm-spaces. We then develop **numerical solvers** for the PGW problem and demonstrate its utility in **experimental applications**.

Gromov-Wasserstein Problem

- **Gromov-Wasserstein (GW) relaxes the need for a common metric space.** It compares distributions supported on different metric spaces by aligning their intrinsic relational structures.
- Compact probability metric measure spaces (mm-space), $\mathbb{X} = (X, d_X, \mu)$, $\mathbb{Y} = (Y, d_Y, \nu)$ with $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$.

$$GW(\mathbb{X}, \mathbb{Y}) := \inf_{\gamma \in \Gamma(\mu,\nu)} \int_{(X \times Y)^2} |d_X(x,x') - d_Y(y,y')|^2 d\gamma(x,y) d\gamma(x',y'). \quad (1)$$

- GW defines a metric for probability measures in different metric spaces.
- Invariant to rigid transformations (rotation/translation).

Partial Gromov-Wasserstein Metric

Given mm-spaces $\mathbb{X} = (X, d_X, \mu)$, $\mathbb{Y} = (Y, d_Y, \nu)$, where $\mu \in \mathcal{M}_+(X), \nu \in \mathcal{M}_+(Y)$:

$$PGW_\lambda(\mathbb{X}, \mathbb{Y}) := \inf_{\gamma \in \Gamma_{\leqslant}(\mu,\nu)} \underbrace{\int_{(X \times Y)^2} |d_X(x,x') - d_Y(y,y')|^2 d\gamma(x,y) d\gamma(x',y')}_{GW(\mathbb{X}, \mathbb{Y})} + \underbrace{\lambda \left(|\mu^{\otimes 2} - \gamma_1^{\otimes 2}| + |\nu^{\otimes 2} - \gamma_2^{\otimes 2}| \right)}_{\text{Mass creation/destruction}}. \quad (2)$$

Theorem 1. $PGW_\lambda(\cdot, \cdot)$ defines a metric between mm-spaces for $\lambda > 0$.

- As $\lambda \to \infty$, $PGW_\lambda(\mathbb{X}, \mathbb{Y}) \to GW(\mathbb{X}, \mathbb{Y})$.
- As $\lambda \to 0$, we allow greater mass creation/destruction, potentially discarding important correspondences.
- As $\lambda \to \infty$, we force more alignment, which may be harmful in the presence of noise.

Frank-Wolfe Solver

In the discrete setting, let $\mu = \sum_{i=1}^n p_i \delta_{x_i}$ and $\nu = \sum_{j=1}^m q_j \delta_{y_j}$. The PGW problem becomes

$$PGW_\lambda(\mathbb{X}, \mathbb{Y}) = \min_{\gamma \in \Gamma_{\leqslant}(\mu,\nu)} \underbrace{\langle (M - 2\lambda) \circ \gamma, \gamma \rangle}_{\mathcal{L}(\gamma)}. \quad (3)$$

$M = \left[|d_X(x_i, x_i') - d_Y(y_j, y_{j'})|^2 \right]_{i,i' \in [1:n], j,j' \in [1:m]} \in \mathbb{R}^{n \times m \times n \times m}$ and $M \circ \gamma := \left[\langle M[i, j, :, :], \gamma \rangle \right]_{i,j} \in \mathbb{R}^{n \times m}$.

Frank-Wolfe Solver. For $k = 1, 2, \dots N$:

- Step 1. Gradient Computation:

$$\nabla_\gamma \mathcal{L}^{(k)} = 2(M - 2\lambda) \circ \gamma^{(k)}.$$

  Solve Partial OT Instance:

$$\tilde{\gamma}^{(k)} \leftarrow \arg\min_{\gamma \in \Gamma_{\leqslant}(\mu,\nu)} \langle \nabla_\gamma \mathcal{L}^{(k)}, \gamma \rangle.$$

- Step 2. Line Search:

$$\alpha^* \leftarrow \arg\min_{\alpha \in [0,1]} \left[\mathcal{L} \left((1 - \alpha)\gamma^{(k+1)} + \alpha \tilde{\gamma}^{(k)} \right) \right].$$

$$\gamma^{(k)} \leftarrow (1 - \alpha^*)\gamma^{(k)} + \alpha^* \tilde{\gamma}^{(k)}.$$

- Requires an initial guess, $\gamma^{(1)}$. If there is no prior knowledge, we can take

$$\gamma^{(1)} = \frac{\mu \otimes \nu}{\max(|\mu|, |\nu|)} \in \Gamma_{\leqslant}(\mu, \nu). \quad (4)$$
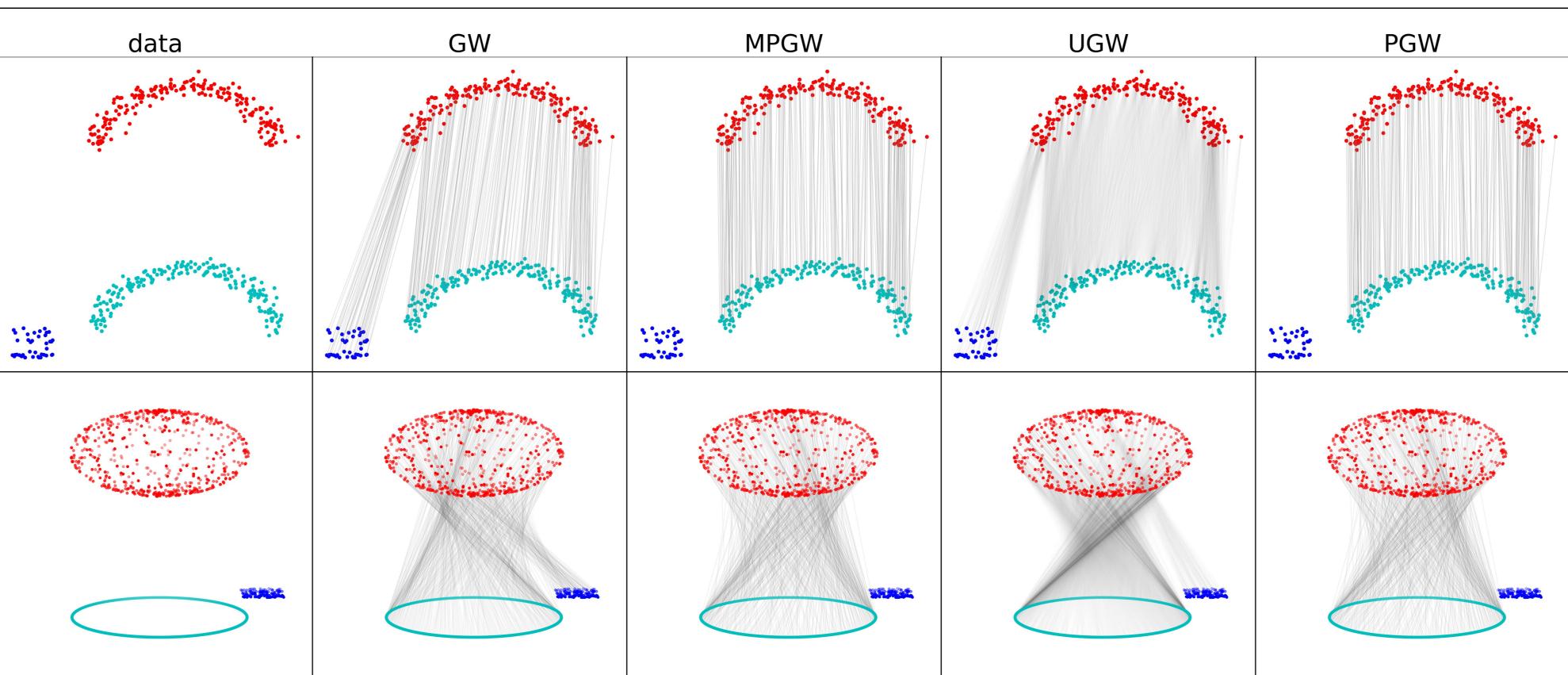
Shape Matching with Outliers

Figure 1. Shape matching between source distribution (red point cloud) and target distribution (blue point clouds). In the first row, both distributions are 2D; in the second row, the source distribution is 3D and target distribution is 2D. We compare the transport plans computed by GW, Mass-Constrained Partial Gromov-Wasserstein (MPGW), Unbalanced Gromov-Wasserstein (UGW), and PGW (ours).
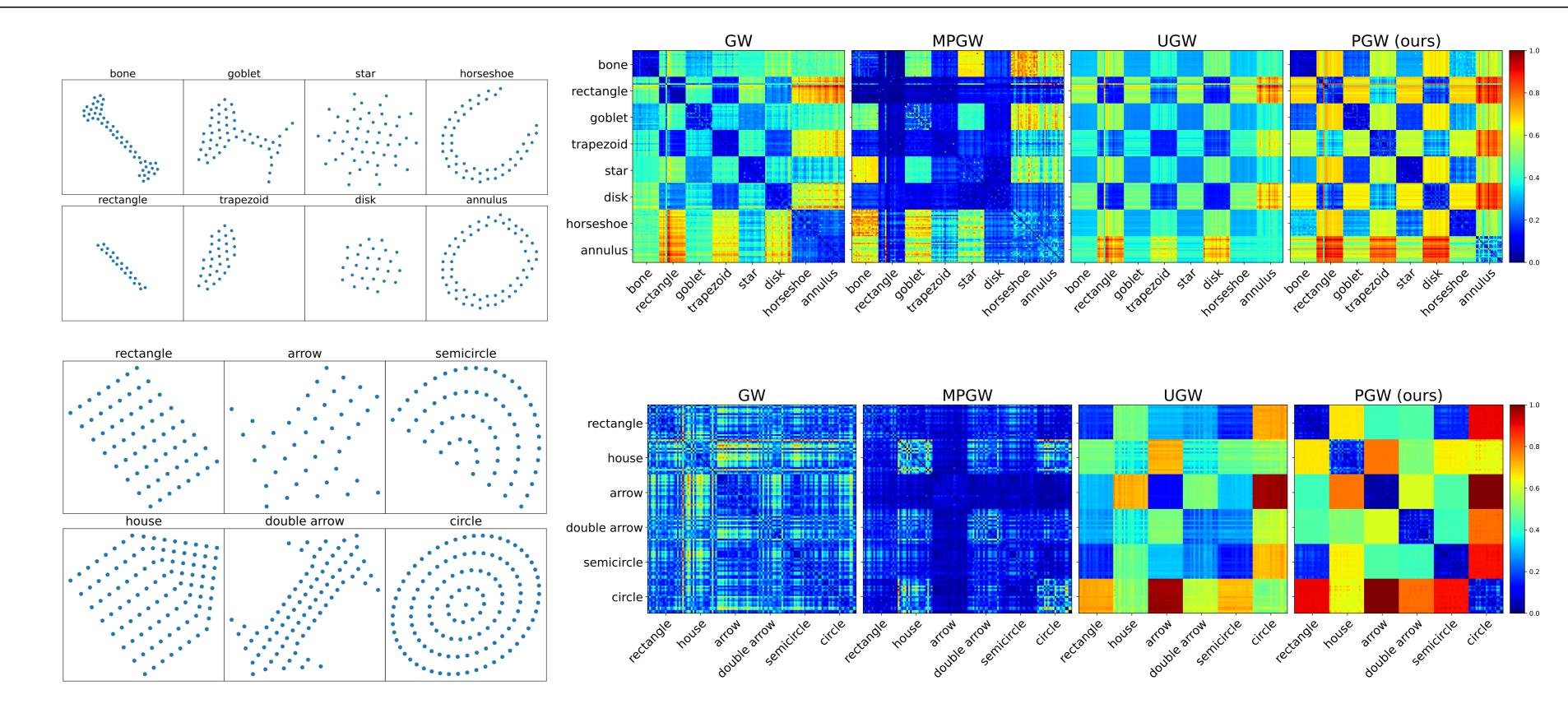
Shape Retrieval

Figure 2. Shape retrieval on two 2D shape datasets. We represent each shape as an mm-space and compute pairwise distances using GW, MPGW, UGW, and PGW (ours). In each row, the first figure visualizes an example shape from each class in the dataset, and the second visualizes the pairwise distance matrices.

Distance	Dataset I	Dataset II		Distance	Dataset I	Dataset II
GW	**0.9813**	0.8083		GW	49.02s	137.12s
MPGW	0.2375	0.2500		MPGW	49.10s	93.90s
UGW	0.89375	0.9000		UGW	1484.49s	519.91s
PGW (ours)	0.9625	**1.0000**		PGW (ours)	**35.92s**	**79.27s**
(a) Accuracy.				(b) Wall-clock time.		

Table 1. We train an SVM model with a kernel based on each of the computed pairwise distance matrices. We then report the average accuracy of the model with stratified 10-fold cross validation. We also report the wall-clock time to compute each pairwise distance matrix.
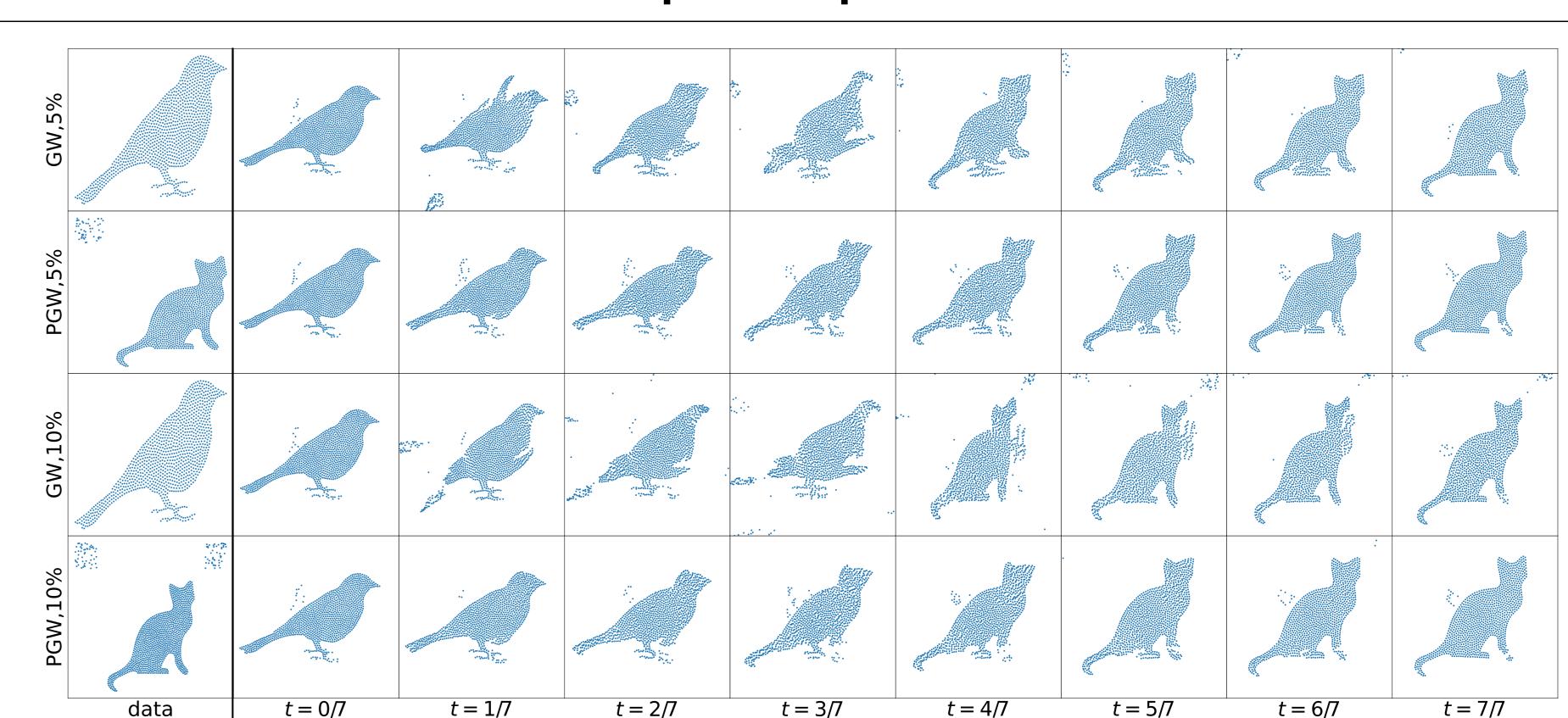
Shape Interpolation

Figure 3. Shape interpolation with PGW barycenters. In the first column, the first and second shapes are the source and target in experiment 1 (5% noise); the third and fourth shapes are the source and target in experiment 2 (10% noise).