Linear Partial Gromov-Wasserstein Embedding

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

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

ICLR International Conference On Learning Representations

Motivation

- **Comparing complex data across different domains requires flexibility and efficiency.** Many ML tasks—such as shape analysis, graph comparison, and cross-domain learning—need tools that can compare structured data lying on different spaces with partial correspondence.
- **Partial Gromov-Wasserstein (PGW) enables partial matching across metric spaces, but is computationally expensive.** Computing pairwise PGW distances scales quadratically with the number of objects, limiting its practicality.
- **Linear Partial Gromov-Wasserstein (LPGW) accelerates PGW while preserving its robustness.** LPGW reduces pairwise distance computation from $\mathcal{O}(K^2)$ to $\mathcal{O}(K)$ for K objects, enabling scalable learning and retrieval tasks without sacrificing the benefits of partial matching.

Contribution: We propose the **LPGW Embedding**, extending the classical linear OT (LOT) framework to the PGW setting. We prove that it gives rise to a **metric** between mm-spaces while allowing for **faster pairwise distance computations**, and we also demonstrate its viability in **experimental applications**.

Partial Gromov-Wasserstein Problem

- **Partial Gromov-Wasserstein (PGW) enables partial transport across different metric spaces.** It extends the classical Gromov-Wasserstein (GW) problem, which 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 $\mu \in \mathcal{M}_+(X), \nu \in \mathcal{M}_+(Y)$.

$$PGW_\lambda(\mathbb{X}, \mathbb{Y}) := \inf_{\gamma \in \Gamma_\leqslant(\mu,\nu)} \underbrace{\int_{(X\times Y)^2} |d_X(x,x') - d_Y(y,y')|^2 d\gamma(x,y) d\gamma(x',y')}_{GW(\mathbb{X},\mathbb{Y})} + \underbrace{\lambda\left(|\mu^{\otimes 2} - \gamma_1^{\otimes 2}| + |\nu^{\otimes 2} - \gamma_2^{\otimes 2}|\right)}_{\text{Mass creation/destruction}} \quad (1)$$

Linear Partial Gromov-Wasserstein Embedding

Linear Partial Gromov-Wasserstein (LPGW) Embedding. Let $\mathbb{S} = (S, d_S, \sigma)$ be a fixed reference space and let $\{\mathbb{X}^i\}_{i=1}^N$ be a series of mm-spaces. Let $\gamma^i \in \Gamma_\leqslant^*(\mathbb{S}, \mathbb{X}^i)$ be an optimal transport plan for $PGW(\mathbb{S}, \mathbb{X}^i)$ such that $\gamma^i = (\mathrm{id} \times T^i)_\# \gamma_S^i$ **(PGW-Monge Mapping Assumption)**.

$$\mathbb{X}^i \mapsto (k^i, \mu_t^i, \mu_c^i) := (d_S(\cdot_1, \cdot_2) - d_{X^i}(T^i(\cdot_1), T^i(\cdot_2)),\ \gamma_S^i,\ (\mu^i)^{\otimes 2} - (\gamma_{X^i}^i)^{\otimes 2}) \quad (2)$$

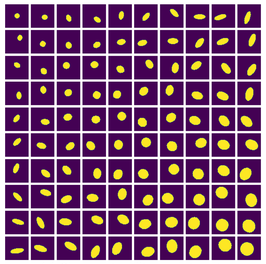
Linear Partial Gromov Wasserstein Distance:

$$LPGW(\mathbb{X}^i, \mathbb{X}^j) := \|k^i - k^j\|^2_{L(\mu_t^i \wedge \mu_t^j)} + \lambda(|(\mu_t^i)^{\otimes 2} - (\mu_t^j)^{\otimes 2}| + |\mu_c^i + \mu_c^j|) \quad (3)$$

Properties of LPGW

- Extends the linear GW (LGW) distance to the unbalanced setting.
- $LPGW$ defines a metric between mm-spaces.
- Computational Complexity: $\mathcal{O}(n^7)$ embedding construction, $\mathcal{O}(n^2)$ LPGW distance.
- Use barycentric projection when Monge maps do not exist.

Elliptical Disks Experiment



(a) Shapes in ellipse dataset.

(b) Reference spaces, $\{\mathbb{S}_1, \ldots, \mathbb{S}_9\}$.

	Points	Time (min)	MRE ↓	PCC ↑
PGW	–	46.97	–	–
$\mathbb{S}_1$	441	0.76	0.1941	0.5781
$\mathbb{S}_2$	676	3.78	0.1264	0.5738
$\mathbb{S}_3$	625	3.13	0.1431	0.5881
$\mathbb{S}_4$	52	0.08	0.2542	0.8581
$\mathbb{S}_5$	289	0.62	0.0538	0.9871
$\mathbb{S}_6$	545	1.56	0.0444	0.9930
$\mathbb{S}_7$	882	1.91	**0.0205**	**0.9952**
$\mathbb{S}_8$	882	1.98	**0.0198**	**0.9954**
$\mathbb{S}_9$	317	0.71	0.0245	0.9949

Figure 1. We evaluate the quality of the approximation of PGW by LPGW using the elliptical disks dataset given in (a). We compute the pairwise distances using PGW and compare against LPGW computed with each of the nine reference spaces in (b). We report the mean relative error (MRE) and Pearson correlation coefficient (PCC). In each shape, the color represents the mass at the corresponding location.

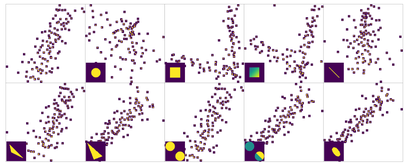


Figure 2. Multidimensional scaling (MDS) visualization of pairwise distances for PGW and LPGW based on different reference spaces.

Shape Retrieval

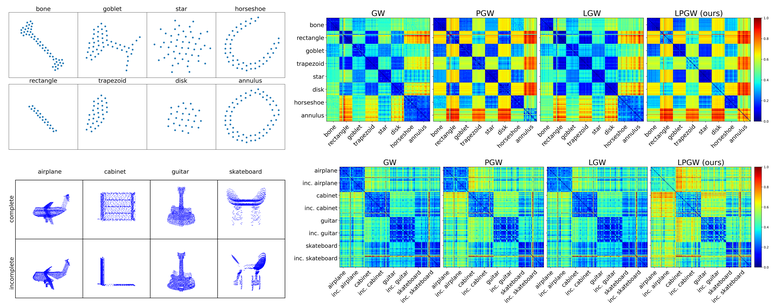


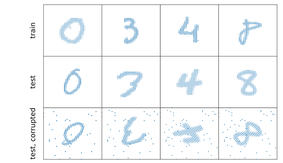
Figure 3. Shape retrieval on 2D/3D shape datasets. The first figure per row visualizes an example from each class in the dataset, and the second visualizes the pairwise distance matrices computed with GW, PGW, LGW, and LPGW (ours).

Shape Retrieval (cont.)

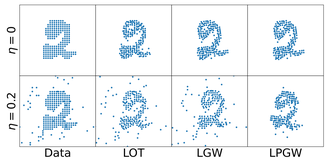
	GW	PGW	LGW	LPGW (ours)
2D Dataset				
Accuracy ↑	**98.1%**	96.2%	93.7%	**97.5%**
Time (s) ↓	43s	39s	**0.4s**	**0.5s**
3D Dataset				
Accuracy ↑	92.5%	**93.8%**	92.5%	**93.7%**
Time (m) ↓	203.0m	203.6m	**1.3m**	**1.8m**

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.

Learning with Transform-Based Embeddings



(a) MNIST 2D point cloud dataset.



(b) Reconstructed digits.

Figure 4. We create training and testing datasets by applying random rotations or horizontal flips to each digit, and we additionally add 0%, 10%, 30%, or 50% noise to testing digits. We then compute the LOT, LGW, or LPGW embeddings of each digit and train a linear regression model on training embeddings. We report the classification accuracy on the test embeddings for each model and report the time to compute all training embeddings and the time to compute all testing embeddings. In (b), we additionally visualize the reconstructed digits using each embedding method.

Data	Method	LOT	LGW	LPGW (ours)
no rotation	Accuracy ↑	89.0%	82.5%	82.5%
	Time ↓	183s+16s	405s+77s	309s+84s
$\eta = 0$	Accuracy ↑	51.2%	82.5%	82.5%
	Time ↓	183s+15s	405s+77s	309s+84s
$\eta = 0.1$	Accuracy ↑	13.4%	17.0%	81.8%
	Time ↓	183s+17s	405s+88s	309s+91s
$\eta = 0.3$	Accuracy ↑	12.5%	13.3%	75.8%
	Time ↓	183s+23s	405s+145s	309s+123s
$\eta = 0.5$	Accuracy ↑	12.5%	12.5%	72.9%
	Time ↓	183s+27s	405s+248s	309s+168s

Figure 5. Shape interpolation using GW, PGW, LGW, and LPGW. In the first column, we visualize the source shape, target shape, and reference space for LGW/LPGW. The target shape includes the addition of 30% noise.