Gromov–Wasserstein Alignment: Statistical and Computational Advancements via Duality

Gromov–Wasserstein Alignment: Statistical and Computational Advancements via Duality

Ziv Goldfeld
Cornell University
email: goldfeld@cornell.edu

***Abstract*—The Gromov-Wasserstein (GW) distance quantifies dissimilarity between metric measure (mm) spaces and provides a natural correspondence between them. As such, it serves as a figure of merit for applications involving alignment of heterogeneous datasets, including object matching, single-cell genomics, and language models translation. While various heuristic methods for approximately evaluating the GW distance from data have been developed, formal guarantees for such approaches—both statistical and computational—remained elusive. This work closes these gaps for the quadratic GW distance between Euclidean mm spaces of different dimensions. At the core of our proofs is a novel dual representation of the GW problem as an infimum of a certain class of optimal transportation problems. The dual form enables deriving, for the first time, sharp empirical convergence rates for the GW distance by providing matching upper and lower bounds. For computational tractability, we consider the entropically regularized GW distance. We derive bounds on the entropic approximation gap, establish sufficient conditions for smoothness and convexity of the objective in the dual problem, and devise efficient algorithms with local and, under convexity, even global convergence guarantees. These advancements facilitate principled estimation and inference methods for GW alignment problems, that are efficiently computable via the said algorithms.**

I. Extended Abstract

The Gromov-Wasserstein (GW) distance quantifies discrepancy between probability distributions supported on different metric spaces by aligning them with one another. Given two metric measure (mm) spaces $(\mathcal{X}, \mathsf{d}_\mathcal{X}, \mu)$ and $(\mathcal{Y}, \mathsf{d}_\mathcal{Y}, \nu)$, the (p,q)-GW distance between them is [1], [2]

$$\mathsf{D}_{p,q}(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \left(\iint_{\mathcal{X}\times\mathcal{Y}} \Delta_q^p \, d\pi\otimes\pi \right)^{\frac{1}{p}}, \qquad (1)$$

where $\Delta_q(x,y,x',y') := \left|\mathsf{d}_\mathcal{X}(x,x')^q - \mathsf{d}_\mathcal{Y}(y,y')^q\right|$ is the distance distortion cost, $\Pi(\mu,\nu)$ is the set of all couplings between μ and ν. The GW distances thus equals the least amount of distance distortion one can achieve between the mm spaces when optimizing over all possible alignments thereof (as modeled by couplings). This approach, which is rooted in optimal transport (OT) theory, is an L^p relaxation of the Gromov-Hausdorff distance between metric spaces and enjoys various favorable properties. Among others, the GW distance (i) identifies pairs of mm spaces between which there exists an measure preserving isometry; (ii) defines a metric on the space of all mm spaces modulo the aforementioned isomorphic relation; and (iii) captures empirical convergence of mm space, i.e., when μ,ν are replaced with their empirical measures $\hat{\mu}_n,\hat{\nu}_n$ based on n samples. Although alignment schemes inspired by the GW framework have seen many applications in computer vision, machine learning, single-cell genomics, and more, existing estimation and computation methods are heuristic and lack formal sample or time complexity guarantees.

To close these gaps, we develop a duality theory for the GW distance, which linearizes this quadratic functional and ties it to the well-understood OT problem. This is done by introducing an auxiliary, matrix-valued optimization variable $\mathbf{A}\in\mathbb{R}^{d_x\times d_y}$ that enables linearizing the dependence on the coupling. We then interchange the optimization over $\mathbf{A}$ and π and identify the inner problem as classical OT problem with respect to a cost function $c_\mathbf{A}$ that depends on $\mathbf{A}$. This representation allows us to lift tool from statistical OT to derive, for the first time, the sample complexity of the empirical plug-in estimator of the GW distance. The derived two-sample rate is $n^{-2/\max\{\min\{d_x,d_y\},4\}}$ (up to a log factor when $\min\{d_x,d_y\}=4$), which matches the corresponding rates for empirical OT. We then provide matching lower bounds, thereby establishing sharpness of the derived rates.

From a computational standpoint, evaluation of the GW distance requires solving a quadratic assignment problem, which is known to be NP-complete. A popular, computationally tractable proxy is the entropic GW (EGW) problem, which regularizes the distance distortion cost from (1) by the Kullback-Leibler divergence penalty $\epsilon\mathsf{D}_{\mathsf{KL}}(\pi\|\mu\otimes\nu)$; $\epsilon>0$ is the regularization parameter. We show that the entropic approximation gap is at most $O(\epsilon\log(1/\epsilon))$, whereby EGW can approximate GW to an arbitrary precision. We then note that our GW duality naturally extends to the EGW distance, up to replacing the OT cost in the variational problem with its entropic OT (EOT) counterpart. Leveraging the connection to EOT, we derive smoothness and convexity properties of the objective in this variational problem, which enable computing it via accelerated gradient descent. Gradients are evaluated by employing Sinkhorn's algorithm to solve the EOT problem, which we model as an inexact oracle and account for it in our analysis. This results in the first efficient algorithms for solving the EGW problem that are subject to formal guarantees in both the convex and non-convex regimes. There results enable principled estimation and computation of GW alignment.

References

[1] F. Mémoli, "Gromov-Wasserstein distances and the metric approach to object matching," *Found. Comput. Math.*, vol. 11, no. 4, 2011.
[2] K.-T. Sturm, "The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces," *arXiv:1208.0434*, 2012.