

QDOT: AN EFFICIENT QUANTILE-WEIGHTED DISTANCE METRIC FOR GEOMETRIC COMPARISON VIA OPTIMAL TRANSPORT

Anonymous authors

Paper under double-blind review

ABSTRACT

Measuring the discrepancy between data distributions in heterogeneous metric spaces is a fundamental challenge. Existing methods, typically based on geometric structures, address this by embedding distributions into a shared space. However, these approaches face fundamental limitations, including the loss of geometric information, computationally intractable representations, and inability to preserve essential structural features. In this work, we introduce the Quantile-weighted Distance Optimal Transport (QDOT), a novel and efficient metric for geometric comparison. QDOT constructs a family of isometry-invariant distance representations by leveraging distance quantiles as structural weights in Euclidean space, thereby preserving essential geometric characteristics and enabling optimal transport coupling within a common space. We prove that, under mild conditions, QDOT is a well-defined metric with a convergence rate no slower than the classical Wasserstein distance. Moreover, we present an integral version that computes the loss in complexity of $\mathcal{O}(n \log n)$. Extensive experiments demonstrate that our methods achieves strong performance across diverse applications, including cross-space comparison, transfer learning, and molecule generation, while also achieving state-of-the-art results on several key metrics.

1 INTRODUCTION

The Wasserstein distance is a powerful metric for comparing probability distributions defined on a same metric space. It has found widespread adoption in a diverse range of machine learning tasks, such as generative models (Arjovsky et al., 2017; De Bortoli et al., 2021; Tong et al., 2024), language models (Kusner et al., 2015; Melnyk et al., 2024), multimodal learning (Xu & Chen, 2023; Alatkari & Wang, 2023; Shi et al., 2024) and reinforcement learning (Klink et al., 2022; Asadulaev et al., 2024). However, a notable limitation of the Wasserstein distance is its sensitivity to the spatial separation of the distributions’ supports, meaning that two distributions with similar shapes but far-apart supports can still have a large Wasserstein distance. Furthermore, the applicability of optimal transport is predicated on a pre-defined ground metric between the supports of the distributions. This requirement renders it intractable for comparing distributions in heterogeneous spaces where such a metric is not readily available.

To address these challenges, metrics based on shape features have emerged (Gromov, 1981; Sturm, 2006; Mémoli, 2011). These methods leverage the distribution of distances within a metric space to compare distributions. Among these, the Gromov-Wasserstein (GW) distance (Sturm, 2006; Mémoli, 2011) stands out as a canonical example. It resolves the issue of comparing distributions in disparate spaces by seeking an optimal coupling of points in a shared, albeit implicitly defined, metric space. However, this formulation is a non-convex optimization problem, the structure of the latent space is not explicitly constructed, and its computation is prohibitively expensive, hindering its use on practical datasets. A related approach, EMD under Transformation Sets (EMD^G) (Cohen & Guibas, 1999), extends the GW concept to Euclidean spaces of the same dimension by finding an optimal orthogonal transformation. However, it is not applicable for cross-space comparisons and is susceptible to converging to local optima. Faster approximations, such as the Invariant Sliced Gromov-Wasserstein (RISGW) (Titouan et al., 2019), achieve a favorable $\mathcal{O}(n \log n)$ complexity, but at the cost of sacrificing theoretical guarantees, such as key metric properties.

Another line of work focuses on extracting shape-invariant features (Belongie et al., 2002; Sun et al., 2019; Yang et al., 2016). These methods map distributions from different spaces into a common feature space for comparison. However, they often suffer from unavoidable information loss due to the unidirectional nature of the feature mapping. To capture richer information, subsequent approaches have employed deep neural networks (Chen et al., 2019; Kim et al., 2020) or transformer based representations (Fuchs et al., 2020; Yu et al., 2023, TBR) for comparison. A key limitation is that these models are typically trained for specific tasks and thus lack generalizability.

To the best of our knowledge, the existing geometric comparison methods can be unified under a common paradigm: they map distributions into a shared space for comparison. However, the mappings employed by these methods often suffer from critical issues: (1) they lack an explicit form; (2) they lead to significant information loss; or (3) they introduce non-geometric artifacts that can corrupt the comparison.

To overcome these challenges, we introduce QDOT, a novel framework that constructs an explicit, geometry-based representation derived from quantile-weighted distances. This principled construction guarantees the integrity of the intrinsic distance information. As illustrated in Figure 1, QDOT strikes a highly effective balance between theoretical properties, generalization capability, time complexity, and computational resource usage, addressing the key trade-offs that limit existing methods. The key contributions of QDOT can be summarized as follows:

Theoretical Guarantees. We prove that QDOT and its integral version constitute a metric on the space of isometry classes under certain conditions. Furthermore, we establish that their sample convergence rate is at least as fast as that of the Wasserstein distance.

Computational Efficiency. We propose a highly efficient algorithm. The computation of our representations has a complexity of only $\mathcal{O}(n \log n)$. The subsequent comparison step for the full QDOT requires an additional $\mathcal{O}(n^2 \log n)$ for a standard Wasserstein distance calculation. In contrast, its integral variant, IQDOT, leverages the closed-form solution of one-dimensional OT to achieve an overall complexity of just $\mathcal{O}(n \log n)$. Consequently, our family of methods offers a highly efficient, quasi-linear time solution.

Versatility and Strong Performance. Our method achieved strong performance across a diverse range of experiments. In point cloud comparison tasks, it accurately captured the geometric dissimilarities between distributions and produced high-quality alignments, showcasing its powerful cross-space capabilities. Furthermore, leveraging its computational efficiency, our method achieved excellent results in a large-scale transfer learning scenario. Finally, when integrated as a loss function for a molecular generation model, the QDOT loss significantly enhanced model generalization and achieved state-of-the-art performance on multiple key metrics.

2 PRELIMINARY

To establish a rigorous framework, we begin with the fundamental definitions. For any metric space (X, d_X) , we can define its Borel σ -algebra, denoted $\mathcal{B}(X)$. For any probability measure μ_X defined on the measurable space $(X, \mathcal{B}(X))$, its support, written as $\text{supp}(\mu_X)$, is the smallest closed set $C \subseteq X$ such that $\mu_X(C) = 1$. For a measurable map $f : X \rightarrow Y$, the *push-forward measure* $f_{\#}\mu_X$ on $(Y, \mathcal{B}(Y))$ is defined by $f_{\#}\mu_X(A) := \mu_X(f^{-1}(A))$ for any set $A \in \mathcal{B}(Y)$. For two measures μ_X and μ_Y in spaces X and Y , a joint probability π on the product space $X \times Y$ is a *coupling* if its marginals satisfy $(\text{proj}_X)_{\#}\pi = \mu_X$ and $(\text{proj}_Y)_{\#}\pi = \mu_Y$. When μ_X and μ_Y are defined on the same metric space (Ω, d) , they can be compared using the Wasserstein distance (Villani et al., 2008), which is defined as:

$$\mathcal{W}_p(\mu_X, \mu_Y) := \left(\inf_{\pi \in \Pi(\mu_X, \mu_Y)} \int_{\Omega \times \Omega} d(x, y)^p d\pi(x, y) \right)^{\frac{1}{p}}. \quad (1)$$

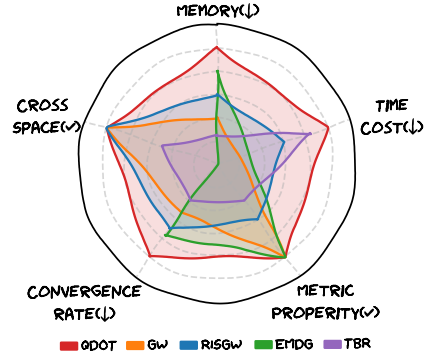


Figure 1: Comparison of various geometric metrics. Lower is better for memory, time cost, and convergence rate. We also evaluate metric properties and cross-space capabilities.

The primary challenge arises when comparing distributions defined on different metric spaces. To handle this scenario, the measurable metric space (Mémoli, 2011, mm-space) is defined as the triplet $\mathcal{X} = (X, d_X, \mu_X)$. For two distinct mm-spaces, $\mathcal{X} = (X, d_X, \mu_X)$ and $\mathcal{Y} = (Y, d_Y, \mu_Y)$, Sturm (2006) proposed the Gromov-Wasserstein distance, extending the concept from equation 1:

$$\mathfrak{G}_p(\mathcal{X}, \mathcal{Y}) := \inf_Z \mathcal{W}_p(f(\mathcal{X}), g(\mathcal{Y})), \quad (2)$$

where the maps $f : X \rightarrow Z$ and $g : Y \rightarrow Z$ are isometric embeddings, satisfying $d_Z(f(x), f(x')) = d_X(x, x')$, $\forall x, x' \in \text{supp}(\mu_X)$ and $d_Z(g(y), g(y')) = d_Y(y, y')$, $\forall y, y' \in \text{supp}(\mu_Y)$. Intuitively, the definition in equation 2 can be understood as finding an optimal "loss-less" projection of the two disparate spaces X and Y into a common latent space Z , within which their Wasserstein distance can be computed. A significant contribution of the Gromov-Wasserstein distance is that it provides a complete metric on the space of mm-spaces. This metric is formally defined as follows.

Definition 1 (Metric on Isometry Classes of mm-spaces) For two mm-spaces $\mathcal{X} = (X, d_X, \mu_X)$ and $\mathcal{Y} = (Y, d_Y, \mu_Y)$, a map $f : X \rightarrow Y$ is called an isometry if it is a surjection satisfying $d_Y(f(x), f(x')) = d_X(x, x')$, $\forall x, x' \in \text{supp}(\mu_X)$. Two mm-spaces are considered isometric if such an isometry exists between them. A function \mathcal{L} that measures the dissimilarity between two mm-spaces is a metric on the isometry classes of mm-spaces if it satisfies: (1) **Identity of Indiscernibles**: $\mathcal{L}(\mathcal{X}, \mathcal{Y}) \geq 0$, and $\mathcal{L}(\mathcal{X}, \mathcal{Y}) = 0$ iff \mathcal{X} and \mathcal{Y} are isometric; (2) **Symmetry**: $\mathcal{L}(\mathcal{X}, \mathcal{Y}) = \mathcal{L}(\mathcal{Y}, \mathcal{X})$; (3) **Triangle Inequality**: $\mathcal{L}(\mathcal{X}, \mathcal{Y}) \leq \mathcal{L}(\mathcal{X}, \mathcal{Z}) + \mathcal{L}(\mathcal{Y}, \mathcal{Z})$.

Inspired by the structure of the Gromov-Wasserstein distance in equation 2, we summarized those methods for comparing two mm-spaces can be expressed in a general form:

$$\mathcal{L}_p(\mathcal{X}, \mathcal{Y}) = \mathcal{W}_p(f(\mathcal{X}), g(\mathcal{Y})) \quad (3)$$

where $f : X \rightarrow Z$ and $g : Y \rightarrow Z$ are mapping functions into a common space Z . However, constructing such maps f and g that yield a valid metric satisfying Definition 1 imposes two key requirements. First, the maps must be **isometry-invariant**. That is, if mm-spaces \mathcal{X}_1 and \mathcal{X}_2 are isometric, their representations must be identically distributed, i.e., $f_{\#}\mu_{\mathcal{X}_1} = f_{\#}\mu_{\mathcal{X}_2}$. Second, to satisfy the identity of indiscernibles property, the representation must be **information-preserving**; it must uniquely encode the metric structure of the original space such that non-isometric spaces map to distinct distributions.

However, the requirement of isometry invariance often entails an unavoidable loss of structural information, and existing methods conforming to the structure of equation 3 fail to meet both requirements simultaneously, thereby falling short of constituting a well-defined metric.

Our work introduces a novel framework designed to explicitly resolve this conflict. Our work introduces a novel framework designed to explicitly resolve this conflict. To preserve information, we draw inspiration from the principle of trilateration. As illustrated in Figure 2, this principle dictates that the location of a target point can be uniquely determined from its distances to a sufficient number of known anchor points. This implies that the distance information can be fully characterized simply by the distances to specific anchor points. Another key problem is how to choose the anchors such that they are isometry-invariant. To address this issue, we introduce the Quantile-weighted Distance Mean (QDM), which utilizes a family of isometry-invariant weights to compute distinct means serving as the distribution's anchors, thereby ensuring isometry invariance between the anchors of different distributions. Consequently, the collection of distances from all points in the support to these QDMs forms a new representation that effectively captures the intrinsic distance information of the entire distribution. In the following sections, we will formally introduce how to determine such isometry-invariant weights and the corresponding QDMs, and detail their theoretical properties.

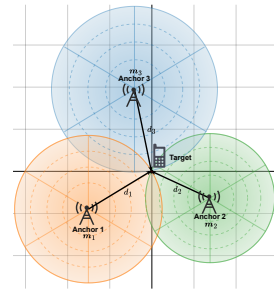


Figure 2: Illustration of trilateration

3 PROPOSED METHOD

For a mm-space $\mathcal{X} = (X, d_X, \mu_X)$, its Barycenter b_X is defined as $\arg\min_{x \in X} \mathbb{E}_{\mu_X} (d_X^2(x, X))$.

Note that, since it minimizes the aggregate squared distance to the distribution, the barycenter is isometry-invariant; consequently, the distribution of distances with respect to the Barycenter is also isometry-invariant. Specifically, if \mathcal{X} and \mathcal{Y} are isometric, then the corresponding distributions of distances to the Barycenter, denoted as $d_X(\cdot, b_X)_{\# \mu_X}$ and $d_Y(\cdot, b_Y)_{\# \mu_Y}$, must be identical. In Euclidean space, the Barycenter has an explicit solution corresponding to the mean of the distribution. Therefore, in the subsequent analysis, for two mm-spaces \mathcal{X} and \mathcal{Y} in Euclidean space, we assume without loss of generality that they are centered, and we utilize the norm as the distance to the Barycenter.

3.1 QUANTILE DISTANCE-WEIGHTED OPTIMAL TRANSPORT

Given a mm-space \mathcal{X} , since the distribution of its norm $\|\cdot\|_{2\# \mu_X}$ is isometry-invariant, to derive a family of distinct isometry-invariant Quantile Weights and QDMs, we employ Gaussian kernels centered at distinct quantiles of $\|\cdot\|_{2\# \mu_X}$ to perform weighting as follows:

Definition 2 (Quantile-weighted Distance Mean (QDM)) Given a centered mm-space $\mathcal{X} = (X, d_X, \mu_X)$, where $X \subseteq \mathbb{R}^p$ and the metric is the standard Euclidean distance, $d_X(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\|_2$, we can derive the distribution of its norms, denoted by $\mu_{\|X\|_2} = \|\cdot\|_{2\# \mu_X}$. Let $F_{\|X\|_2}$ denote the cumulative distribution function (CDF) of $\mu_{\|X\|_2}$, and let $F_{\|X\|_2}^{-1}$ be its associated quantile function. For any quantile level $q \in (0, 1)$, we define the Quantile Distance Weight function $w : X \times (0, 1) \rightarrow \mathbb{R}$ as

$$w^X(\mathbf{x}, q) := e^{-\sigma(\|\mathbf{x}\|_2 - F_{\|X\|_2}^{-1}(q))^2}, \quad (4)$$

where σ is a bandwidth parameter. Based on these weights, the corresponding QDM $\mathbf{m}^X : (0, 1) \rightarrow \mathbb{R}^p$ is defined as the weighted mean:

$$\mathbf{m}^X(q) := \frac{\mathbb{E}_{\mu_X}[w^X(X, q)X]}{\mathbb{E}_{\mu_X}[w^X(X, q)]}. \quad (5)$$

Intuitively, for a given quantile level q , the weight function $w^X(\mathbf{x}, q)$ is concentrated on points whose norms are close to the q -th quantile of the norm distribution. Therefore, the QDM can be served as a canonical, isometry-invariant anchor point. We use it to define the Quantile Distance-weighted Mean Distance (QDMD), which is a function $\phi : X \times (0, 1) \rightarrow \mathbb{R}$ that measures the distance from a point to the QDM:

$$\phi^X(\mathbf{x}, q) = d_X(\mathbf{x}, \mathbf{m}^X(q)). \quad (6)$$

Given a quantile level vector $\mathbf{q} = (q_1, \dots, q_k) \in (0, 1)^k$. The corresponding QDMD maps each point \mathbf{x} to a feature vector in \mathbb{R}^{k+1} :

$$\phi^X(\mathbf{x}, \mathbf{q}) := [\phi_0^X(\mathbf{x}), \phi^X(\mathbf{x}, q_1), \phi^X(\mathbf{x}, q_2), \dots, \phi^X(\mathbf{x}, q_k)],$$

where $\phi_0^X(\mathbf{x}) = \|\mathbf{x}\|_2$ is the original norm, included as a fundamental reference distance to the origin. This mapping transforms the measure μ_X into a push-forward measure on \mathbb{R}^{k+1} . Based on this transformation, we define the Quantile-weighted Distance Optimal Transport as follows:

Definition 3 (Quantile-weighted Distance Optimal Transport (QDOT)) Let $\mathcal{X} = (X, d_X, \mu_X)$ and $\mathcal{Y} = (Y, d_Y, \mu_Y)$ be two mm-spaces, where $X \subseteq \mathbb{R}^d$ and $Y \subseteq \mathbb{R}^s$, and their metrics d_X and d_Y are the standard Euclidean distances. For a given vector of quantile levels $\mathbf{q} \in (0, 1)^k$, we compute their corresponding QDMD representations, ϕ^X and ϕ^Y . The QDOT distance is then defined as the Wasserstein distance between the resulting push-forward measures:

$$\mathcal{QD}_p(\mathcal{X}, \mathcal{Y}) = \mathcal{W}_p(\phi_{\# \mu_X}^X, \phi_{\# \mu_Y}^Y). \quad (7)$$

The overall procedure of the QDOT framework is illustrated in Figure 7.

3.2 THEORETICAL RESULTS

We will now establish the key theoretical properties of our proposed QDOT framework, demonstrating that it is a well-defined metric with favorable sample convergence guarantees.

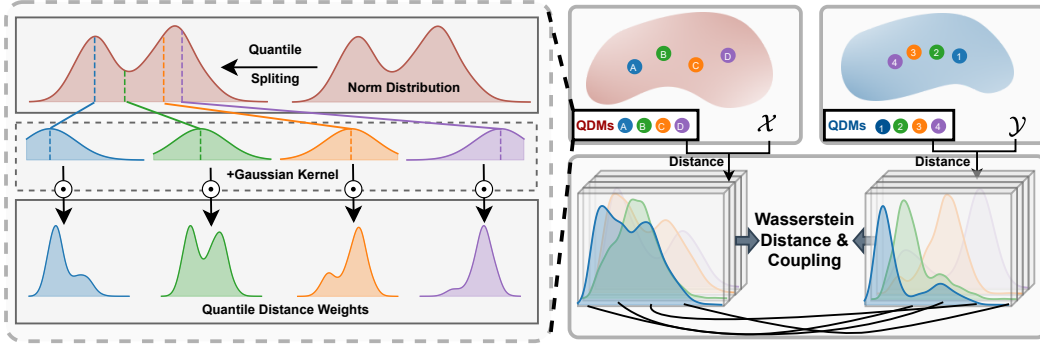


Figure 3: Illustration of the QDOT framework. The method first computes canonical QDM anchors for each distribution, then generates QDMD vector representations based on distances to these anchors. The final QDOT distance is the Wasserstein distance between these two resulting representation distributions.

Theorem 1 (Metric Property on Isometry Classes) Let $\mathcal{X} = (X, d_X, \mu_X)$ and $\mathcal{Y} = (Y, d_Y, \mu_Y)$ be two mm-spaces embedded in Euclidean spaces. For a given quantile level vector $\mathbf{q} \in (0, 1)^k$, suppose their corresponding QDMs satisfy the dimensionality condition $\dim(\{\mathbf{m}^X(q_i)\}_{1 \leq i \leq k}) = \dim(X)$ or $\dim(\{\mathbf{m}^Y(q_i)\}_{1 \leq i \leq k}) = \dim(Y)$. Then, the p -QDOT distance, \mathcal{QD}_p , defines a metric on the isometry classes of these mm-spaces.

The condition on dimensionality is inspired by the principle of trilateration (Thomas & Ros, 2005) and is generally satisfied in practical scenarios. The proof for Theorem 1 is provided in Appendix B.1. In addition to its metric properties, we also prove that QDOT has an empirical convergence rate.

Theorem 2 (Empirical Convergence Rate) Let $\mathcal{X} = (X, d_X, \mu_X)$ be an mm-space where $X \subseteq \mathbb{R}^d$. Let $\{x_i\}_{i=1}^n$ be i.i.d. samples drawn from μ_X , and let the empirical measure be defined as $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$. If we denote the corresponding empirical mm-space as $\mathcal{X}_n = (\{x_1, \dots, x_n\}, d_X, \mu_n)$, then we have $\mathcal{QD}_p(\mathcal{X}, \mathcal{X}_n) \rightarrow 0$, as $n \rightarrow \infty$.

Moreover, under additional regularity conditions, namely (1) the distribution has a finite fourth moment and (2) its CDF is continuous at the chosen quantile levels and has a strictly positive density, the convergence rate of \mathcal{QD}_p remains consistent with the standard rate for the Wasserstein distance (Fournier & Guillin, 2015). Specifically, for the common case where $d \geq 2$, the expected error is bounded by:

$$\mathbb{E}[\mathcal{QD}_p(\mathcal{X}, \mathcal{X}_n)] = O(n^{-1/d}).$$

The finite moment requirement is a standard condition inherited from Fournier & Guillin (2015), while the continuity condition is a classical assumption that ensures the consistency of empirical quantiles. Both of these conditions are mild and hold for the vast majority of practical data distributions. For a detailed discussion and rates under other conditions, we refer the reader to Fournier & Guillin (2015) and our proof in Appendix B.2.

Based on the preceding theoretical results, we can summarize several key properties of the QDOT.

Corollary 1 Let $\mathcal{X} = (X, d_X, \mu_X)$ and $\mathcal{Y} = (Y, d_Y, \mu_Y)$ be two mm-spaces embedded in Euclidean spaces, where $X \subseteq \mathbb{R}^d$ and $Y \subseteq \mathbb{R}^s$. The QDOT distance exhibits the following properties:

- (1) **Location Invariance.** For any translation vector $z \in \mathbb{R}^d$, let the translated mm-space be denoted by $\mathcal{X}_z = (X + z, d_X, \mu_{X+z})$, we have $\mathcal{QD}_p(\mathcal{X}_z, \mathcal{Y}) = \mathcal{QD}_p(\mathcal{X}, \mathcal{Y})$.
- (2) **Rotation and Reflection Invariance.** For any orthogonal transformation $R : \mathbb{R}^d \rightarrow \mathbb{R}^d$, let $\mathcal{X}_R = (R(X), d_{R(X)}, R_{\#}\mu_X)$, we have $\mathcal{QD}_p(\mathcal{X}_R, \mathcal{Y}) = \mathcal{QD}_p(\mathcal{X}, \mathcal{Y})$.
- (3) **Numerical Convergence.** Let \mathcal{X}_n and \mathcal{Y}_n be the empirical mm-spaces constructed from n i.i.d. samples drawn from \mathcal{X} and \mathcal{Y} , respectively. The empirical QDOT distance converges

as $n \rightarrow \infty$: $\mathcal{QD}_p(\mathcal{X}_n, \mathcal{Y}_n) \rightarrow \mathcal{QD}_p(\mathcal{X}, \mathcal{Y})$. Furthermore, under the condition in 2, its expectation convergence rate is given by $\mathbb{E}|\mathcal{QD}_p(\mathcal{X}_n, \mathcal{Y}_n) - \mathcal{QD}_p(\mathcal{X}, \mathcal{Y})| = \mathcal{O}(n^{-\frac{1}{\max(d,s)}})$.

3.3 NUMERICAL IMPLEMENTATION

In the numerical setting, we consider discrete probability measures. Let Δ^{n-1} denote the $(n-1)$ -simplex. Given two sets of samples, $X_n = \{\mathbf{x}_i\}_{1 \leq i \leq n}$ in \mathbb{R}^d and $Y_m = \{\mathbf{y}_j\}_{1 \leq j \leq m}$ in \mathbb{R}^s , we represent the sample sets as data matrices $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^\top \in \mathbb{R}^{n \times d}$ and $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m)^\top \in \mathbb{R}^{m \times s}$. The corresponding probability vectors are denoted as $\mathbf{p}^{\mathbf{X}} \in \Delta^{n-1}$ and $\mathbf{p}^{\mathbf{Y}} \in \Delta^{m-1}$, and their associated discrete measures are $\mu_{\mathbf{X}} = \sum_{i=1}^n p_i^{\mathbf{X}} \delta_{\mathbf{x}_i}$ and $\mu_{\mathbf{Y}} = \sum_{j=1}^m p_j^{\mathbf{Y}} \delta_{\mathbf{y}_j}$, respectively. For a given quantile level vector $\mathbf{q} \in (0, 1)^k$, the procedure for computing the QDOT distance is detailed in Algorithm 1.

Algorithm 1 QDOT

Require: $\mathbf{X}, \mathbf{Y}, \mathbf{p}^{\mathbf{X}}, \mathbf{p}^{\mathbf{Y}}, \mathbf{q}$

- 1: Initialize the data matrices \mathbf{X} and \mathbf{Y} .
- 2: Compute the sample norms of \mathbf{X} and \mathbf{Y} : $\phi_0^{\mathbf{X}} \leftarrow (\|\mathbf{x}_i\|_2)_{1 \leq i \leq n}$, $\phi_0^{\mathbf{Y}} \leftarrow (\|\mathbf{y}_j\|_2)_{1 \leq j \leq m}$.
- 3: Compute the \mathbf{q} -quantiles of the sample norms: $\triangleright \mathcal{O}(n \log n + m \log m)$
- 4: $\mathbf{r}^{\mathbf{X}} \leftarrow (F_{\|\mathbf{X}\|_2}^{-1}(q_1), \dots, F_{\|\mathbf{X}\|_2}^{-1}(q_k))$, $\mathbf{r}^{\mathbf{Y}} \leftarrow (F_{\|\mathbf{Y}\|_2}^{-1}(q_1), \dots, F_{\|\mathbf{Y}\|_2}^{-1}(q_k))$.
- 5: **for** $i \leftarrow 1$ **to** k **do** $\triangleright \mathcal{O}(knd + kms)$
- 6: Compute the quantile weights:

$$\mathbf{w}_{ij}^{\mathbf{X}} \leftarrow \frac{p_j^{\mathbf{X}} \exp\{-\sigma(\mathbf{d}_j^{\mathbf{X}} - \mathbf{r}_i^{\mathbf{X}})^2\}}{\sum_{j'=1}^n p_{j'}^{\mathbf{X}} \exp\{-\sigma(\mathbf{d}_{j'}^{\mathbf{X}} - \mathbf{r}_i^{\mathbf{X}})^2\}}, \quad \mathbf{w}_{ij}^{\mathbf{Y}} \leftarrow \frac{p_j^{\mathbf{Y}} \exp\{-\sigma(\mathbf{d}_j^{\mathbf{Y}} - \mathbf{r}_i^{\mathbf{Y}})^2\}}{\sum_{j'=1}^m p_{j'}^{\mathbf{Y}} \exp\{-\sigma(\mathbf{d}_{j'}^{\mathbf{Y}} - \mathbf{r}_i^{\mathbf{Y}})^2\}}$$
- 7: Compute the q_i -quantile means: $\mathbf{m}_i^{\mathbf{X}} \leftarrow \mathbf{X}^\top \mathbf{w}_i^{\mathbf{X}}$, $\mathbf{m}_i^{\mathbf{Y}} \leftarrow \mathbf{Y}^\top \mathbf{w}_i^{\mathbf{Y}}$
- 8: Compute the distances to quantile means:

$$\phi_i^{\mathbf{X}} \leftarrow (\|\mathbf{x}_j - \mathbf{m}_i^{\mathbf{X}}\|_2)_{1 \leq j \leq n}, \quad \phi_i^{\mathbf{Y}} \leftarrow (\|\mathbf{y}_j - \mathbf{m}_i^{\mathbf{Y}}\|_2)_{1 \leq j \leq m}$$
- 9: **end for**
- 10: Concatenate representations: $\Phi^{\mathbf{X}} \leftarrow [\phi_0^{\mathbf{X}}, \phi_1^{\mathbf{X}}, \dots, \phi_k^{\mathbf{X}}]$, $\Phi^{\mathbf{Y}} \leftarrow [\phi_0^{\mathbf{Y}}, \phi_1^{\mathbf{Y}}, \dots, \phi_k^{\mathbf{Y}}]$
- 11: \mathcal{W}_p Computation: $\mathcal{QD}_p, \Pi_{\mathcal{QD}_p} \leftarrow \mathcal{W}_p((\Phi^{\mathbf{X}})_{\#} \mu_{\mathbf{X}}, (\Phi^{\mathbf{Y}})_{\#} \mu_{\mathbf{Y}}) \triangleright \mathcal{O}(n^2 \log n + m^2 \log m)$
- 12: **return** $\mathcal{QD}_p, \Pi_{\mathcal{QD}_p}$

Computational Cost. For simplicity, we assume $m \leq n$ and $s \leq d$. The initial norm computation requires $\mathcal{O}(nd)$ operations. Computing the quantiles takes $\mathcal{O}(n \log n)$ time. The main loop for computing the QWs, QDMs, and QDMDs has a total cost of $\mathcal{O}(knd)$. It is noteworthy that the final representations are generated in nearly linear time with respect to n . The final step of calculating the Wasserstein distance has a complexity of $\mathcal{O}(n^2 \log n)$ by Sinkhorn Algorithm (Cuturi, 2013) or Earth Moving Distance (Rubner et al., 2000).

To mitigate the high complexity of the final step, we also propose an Integral-QDOT approach.

3.4 INTEGRAL-QDOT

To address the high computational cost of the standard Wasserstein distance, methods based on slicing (Bonneel et al., 2015; Deshpande et al., 2019) and the closed-form solution of one-dimensional Optimal Transport have become increasingly popular. We observe that the QDMD representation, $\phi^{\mathbf{X}}(x, q)$, defined in the previous section, is a scalar value for any given quantile level q . This structure naturally inspires an alternative approach: instead of comparing the multi-dimensional representations in \mathbb{R}^{k+1} , we can compare the one-dimensional distributions of the QDMD scalars for each q and then aggregate the results. This leads to the Integral-QDOT (IQDOT) approach.

Definition 4 (Integral-QDOT) Given two mm-spaces $\mathcal{X} = (X, d_X, \mu_X)$ and $\mathcal{Y} = (Y, d_Y, \mu_Y)$, where $X \subseteq \mathbb{R}^d$, $Y \subseteq \mathbb{R}^s$, and d_X, d_Y are the standard Euclidean distances. For any quantile level $q \in (0, 1)$, we can obtain their corresponding scalar QDMD representations, $(\phi^{\mathbf{X}}(\cdot, q))_{\#} \mu_X$ and $(\phi^{\mathbf{Y}}(\cdot, q))_{\#} \mu_Y$. The Integral-QDOT distance is then defined as the L_p -norm of the 1-D Wasserstein

distances between these push-forward measures, integrated over all $q \in (0, 1)$:

$$\mathcal{IQD}_p(\mathcal{X}, \mathcal{Y}) := \inf_{\pi \in \Pi(\mu_X, \mu_Y)} \left(\int_{(0,1)} \int_{X \times Y} |\phi^X(x, q) - \phi^Y(y, q)|^p d\pi(x, y) dq \right)^{1/p}. \quad (8)$$

We further establish that IQDOT also constitutes a well-defined metric.

Theorem 3 (Metric Property of IQDOT) *Let $\mathcal{X} = (X, d_X, \mu_X)$ or $\mathcal{Y} = (Y, d_Y, \mu_Y)$ be two mm-spaces embedded in Euclidean spaces. If the set of QDMs satisfies the dimensionality condition $\dim(\{\mathbf{m}^X(q)\}_{q \in (0,1)}) = \dim(X)$ and $\dim(\{\mathbf{m}^Y(q)\}_{q \in (0,1)}) = \dim(Y)$, then \mathcal{IQD}_p defines a metric on the isometry classes of these mm-spaces.*

As for the numerical implementation of IQDOT, we consider computing the mean of the one-dimensional Wasserstein distances for the QDMD representations corresponding to k quantiles. Since the 1d Wasserstein distance has a closed-form solution, the computational cost of the final step in Algorithm 1 is reduced to $\mathcal{O}(kn \log n)$, thereby achieving highly efficient computation. [The specific implementation of IQDOT is detailed in Algorithm 2 in Appendix C.2.](#)

4 EXPERIMENTS

We next present a series of experiments designed to empirically validate several key properties of our proposed framework. The evaluation aims to demonstrate QDOT’s: (1) effectiveness in cross-space tasks; (2) fast computational efficiency; (3) transferability and versatility in comparing diverse distributions; and (4) strong performance in complex models. All CPU-based experiments were conducted on an Intel(R) Xeon(R) Platinum 8280 CPU @ 2.70GHz with 256GB RAM. All GPU-based experiments were performed on a single NVIDIA RTX 4090 GPU with 24GB VRAM.

4.1 CROSS SPACE TASKS

To evaluate the cross-space alignment capability and the metric accuracy of our proposed method, we conduct experiments on the camel-gallop data (Sumner & Popović, 2004). This data consists of a reference 3D point cloud model and a corresponding 48-frame sequence of a galloping camel. For our tests, we subsample the point clouds to 10,000 points. The experimental task is to match each frame of the 3D sequence against projections of the static reference model onto three distinct 2D subspaces. We assess alignment quality using the Transformed Mean Squared Error (TMSE) and the Inlier Ratio (IR), formally defined in Appendix D.2. TMSE measures the alignment cost in the original 3D space, while IR quantifies the percentage of correctly matched points within a given tolerance. Our comparisons are structured as follows: First, we evaluate our QDOT by Sinkhorn algorithm, against Entropic Gromov-Wasserstein (EGW) (Peyré et al., 2016), testing both methods with regularization parameters $\lambda = 0.1$ and $\lambda = 0.01$. Second, to assess performance with sparse couplings, we compare QDOT by EMD against the classical GW method.

Table 1: Cross Space Results on the camel-gallop dataset

Methods	Transformed MSE ↓					Inlier Ratio (%) ↑					Time ($\times 10^2 s$)
	3D	2D _{1st}	2D _{2nd}	2D _{3rd}	Avg.	3D	2D _{1st}	2D _{2nd}	2D _{3rd}	Avg.	
EGW _($\lambda=0.1$)	0.37	0.37	0.39	0.40	0.38	41.71	41.90	39.66	38.69	40.49	3.03
QDOT_(Sink-0.1)	0.32	0.32	0.32	0.33	0.32	48.75	48.25	48.84	47.03	48.22	0.13
EGW _($\lambda=0.01$)	0.21	0.23	0.22	0.33	0.25	72.12	64.92	69.51	50.89	64.36	21.51
QDOT_(Sink-0.01)	0.22	0.24	0.23	0.25	0.24	71.15	63.75	68.48	63.20	66.65	0.74
GW	0.26	0.25	0.30	0.35	0.29	61.29	63.04	56.78	49.16	58.06	14.63
QDOT_(EMD)	0.25	0.27	0.25	0.27	0.26	63.35	58.60	63.87	60.02	61.46	0.38

As shown in Table 1, the couplings produced by QDOT achieve comparable or superior results to those from GW and EGW. Notably, this performance is attained with a computational cost that is up to 30 times lower, underscoring the effectiveness and efficiency of our algorithm for cross-space alignment. Furthermore, to evaluate the ability of cross-space metrics, we plotted the dissimilarity trends over the point-cloud sequence, and included IQDOT, SGW, and RISGW for comparison.

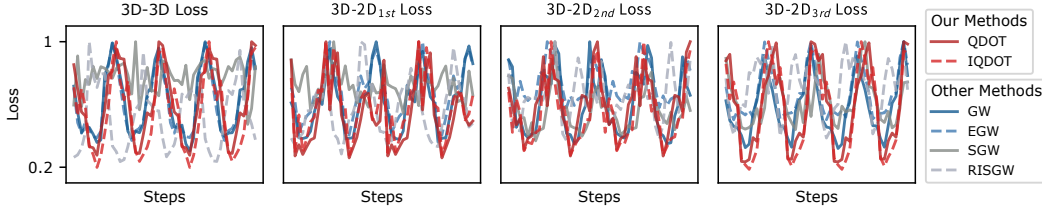


Figure 4: Loss trends between the camel-gallop sequence and the four reference in different spaces. It is evident that QDOT, IQDOT, GW, and EGW successfully capture the four periodic cycles of the camel’s gallop, exhibiting consistent fluctuation amplitudes and trends across all reference spaces. In contrast, SGW and RISGW fail to reveal a clear, meaningful pattern.

The results, visualized in Figure 4, demonstrate that methods possessing well-defined metric properties consistently produce effective and meaningful dissimilarity curves. In stark contrast, SGW and RISGW, which lack these theoretical guarantees, fail to reveal this underlying geometric pattern, underscoring the importance of sound metric properties.

4.2 TIME COST

To benchmark the practical runtime performance of our methods, we conduct an experiment using randomly generated 2D point clouds, with the number of support points ranging from 10^2 to 10^6 . We compare our methods against GW, EGW, SGW, and RISGW. For this test, all methods are configured with 50 projections (or quantiles); IQDOT is additionally implemented using 5 quantiles, and QDOT uses an EMD solver. The results are visualized in Figure 5. The plot clearly shows that for smaller sample sizes, the QDOT family of methods is faster than all baselines. In large-scale scenarios involving tens of thousands of points, IQDOT achieves a remarkable speedup, running up to thousands of times faster than the classic GW method. QDOT also maintains a significant advantage. The overall trends in the log-log plot confirm our complexity analysis: IQDOT exhibits a quasi-linear time complexity, whereas QDOT scales quadratically.

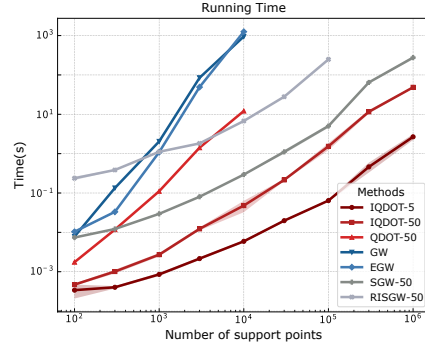


Figure 5: Runtime comparison of various methods. All experiments were conducted in a CPU environment.

4.3 TRANSFER LEARNING

Leveraging the robust cross-domain behavior of QDOT, we propose a parameter-free transfer learning approach for point cloud classification. We evaluate it on ModelNet40 (Wu et al., 2015) and ShapeNetPart (Fan et al., 2017), focusing on the seven shared classes (3,072 samples from ModelNet and 15,402 from ShapeNet). This setting entails tens of millions of pairwise comparisons, making computationally intensive approaches infeasible; we therefore restrict our baselines to linear-time methods. To test rotational robustness, random vertical-axis rotations are applied, and experiments are conducted with point clouds subsampled to $n = 1024$ and $n = 2048$. The task treats one dataset as the source and the other as the target: each target cloud is classified via 1-Nearest Neighbor using distances to all source clouds computed by our metric. Results are reported below.

As shown in Table 2, IQDOT demonstrates more consistent performance compared to SGW. In the ModelNet-to-ShapeNet transfer task, IQDOT achieves superior overall results. Furthermore, its performance in the ShapeNet-to-ModelNet direction is exceptionally strong, improving the average accuracy by 35% over the SGW method. Notably, IQDOT completed this large-scale task in just a few hours, highlighting its remarkable efficiency.

Table 2: Point Cloud Transfer Learning Classification Accuracy (%)

Methods (Mo→Sh)	airplane	car	chair	guitar	lamp	laptop	table	Avg.	Time(h)
SGW-1024	<u>96.69</u>	99.10	60.67	0.00	<u>71.29</u>	<u>98.66</u>	<u>94.25</u>	79.77	13.11
SGW-2048	96.98	99.10	60.24	0.00	71.62	98.66	94.47	79.83	21.84
IQDOT-1024	94.86	84.18	<u>83.82</u>	96.56	51.13	96.45	84.95	<u>83.89</u>	1.59
IQDOT-2048	95.35	87.86	85.01	<u>96.44</u>	52.55	98.00	87.15	85.42	<u>5.98</u>
Methods (Sh→Mo)	airplane	car	chair	guitar	lamp	laptop	table	Avg.	Time(h)
SGW-1024	89.66	37.03	43.68	0.00	80.55	20.11	<u>99.59</u>	59.66	13.11
SGW-2048	91.18	37.03	44.08	0.00	83.33	20.11	99.79	60.31	21.84
IQDOT-1024	98.89	<u>87.20</u>	<u>91.10</u>	98.43	81.94	<u>97.04</u>	96.74	<u>93.97</u>	1.59
IQDOT-2048	98.89	91.91	93.22	<u>98.03</u>	<u>77.77</u>	97.63	97.56	95.05	<u>5.98</u>

4.4 MOLECULE GENERATION

Molecular generation is a central challenge in drug discovery and molecular science. Recent advances show that diffusion models with equivariant neural architectures have become the dominant paradigm (Hooeboom et al., 2022; Xu et al., 2023; Song et al., 2024; Feng et al., 2025). Since molecular properties are invariant to absolute position and orientation, prior works attempted to enforce stability through explicit structural alignment during training (Song et al., 2023; Hassan et al., 2024), but such procedures incur high computational cost. We instead integrate the **IQDOT** distance directly into the diffusion loss, providing an alignment-free learning signal that is inherently invariant and emphasizes geometric fidelity. This encourages physically stable and generalizable molecule generation without the overhead of explicit alignment.

We evaluate this approach on QM9 (Ramakrishnan et al., 2014) by adapting the classic model EDM (Hooeboom et al., 2022) and a recently model UniGEM (Feng et al., 2025), with **IQDOT**-augmented loss functions. Both are trained for 3000 epochs, and the results are reported below.

Table 3: Molecule Generation Results on QM9 Dataset

Methods	Atom sta(%)	Mol sta(%)	Valid(%)	V * U(%)
Data	99.0	95.2	97.7	97.7
EDM	98.70 \pm 0.01	86.56 \pm 0.27	93.73 \pm 0.12	92.04 \pm 0.03
→MSE _{0.1}	98.81 \pm 0.04	87.63 \pm 0.21	94.54 \pm 0.15	92.81\pm0.20
→MSE _{0.3}	98.68 \pm 0.07	86.04 \pm 0.74	93.83 \pm 0.43	92.39 \pm 0.55
→ IQDOT _{0.1}	99.15 \pm 0.01	90.91 \pm 0.21	96.75 \pm 0.33	89.06 \pm 0.19
→ IQDOT _{0.3}	99.35\pm0.06	93.29\pm0.81	97.80\pm0.22	81.21 \pm 2.03
UniGEM	98.90 \pm 0.03	89.40 \pm 0.02	94.58 \pm 0.07	92.75 \pm 0.11
→MSE _{0.1}	98.91 \pm 0.04	89.00 \pm 0.22	95.00 \pm 0.16	93.13 \pm 0.10
→MSE _{0.3}	99.00 \pm 0.07	89.48 \pm 0.70	95.16 \pm 0.16	93.34 \pm 0.18
→ IQDOT _{0.1}	99.24 \pm 0.03	92.73 \pm 0.09	96.85 \pm 0.11	93.42\pm0.23
→ IQDOT _{0.3}	99.44\pm0.01	95.23\pm0.18	97.94\pm0.06	83.88 \pm 0.30

¹ The notation "**IQDOT**_{0.1}" indicates that the error function $\mathcal{L} \leftarrow 0.1\mathcal{L}_{\text{IQDOT}} + 0.9\mathcal{L}_{\text{MSE}}$; "MSE_{0.1}" indicates that $\mathcal{L} \leftarrow 0.9\mathcal{L}_{\text{MSE}}$.

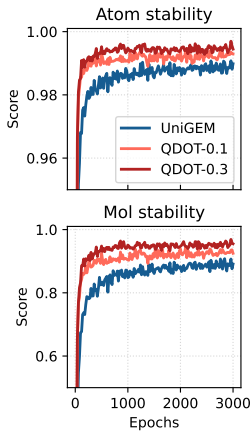


Figure 6: Stability trends across epochs.

As shown in Table 3, our **IQDOT** loss significantly enhances both EDM and UniGEM baselines. For UniGEM, a mere 10% QDOT weight (**IQDOT**_{0.1}) improves all four metrics, while a 30% weight (**IQDOT**_{0.3}) establishes a new state-of-the-art, boosting Atom and Molecule Stability to 99.44% and 95.23%, respectively. An ablation study confirms these gains stem from **IQDOT**'s geometric guidance, not merely from re-weighting the original loss. Furthermore, as illustrated in Figure 6, **IQDOT** accelerates training, halving the convergence time for UniGEM to just 1000 epochs and thereby improving training efficiency and stability.

To assess the fine-tuning capability of the IQDOT loss, we conducted an additional experiment on GEOM-Drug (Axelrod & Gomez-Bombarelli, 2022). Starting from a UniGEM model pre-trained for 13 epochs (Feng et al., 2025), we introduced IQDOT into the loss and fine-tuned for 3 more epochs with a learning rate of 10^{-4} . The results in Table 4 show that while continued training with the original MSE loss yields no further gains, incorporating IQDOT leads to substantial improvements in Atom Stability and Molecule Stability, with only a negligible drop in validity.

Table 4: Fine-tuning Results on GEOM-Drugs

Methods	Atom sta(%)	Mol sta(%)	Valid(%)
UniGEM	84.84	1.20	98.29
→MSE	84.39	1.13	99.01
→IQDOT _{0.1}	87.94	6.04	<u>98.46</u>
→IQDOT _{0.2}	91.87	<u>17.14</u>	97.86
→IQDOT _{0.3}	<u>91.73</u>	22.05	94.62

5 CONCLUSION

This work introduces QDOT, a novel geometric metric. By constructing isometry-invariant anchors from distance quantiles and deriving a lossless distance representation through trilateration theory, QDOT establishes a rigorous metric on isometry classes of mm-spaces. Experimental results demonstrate its effectiveness across cross-space alignment and comparison, computational efficiency, transfer learning, and molecular generation tasks. Future directions include: (1) extending QDOT to hyperbolic and spherical spaces; (2) developing more general approaches for constructing isometry-invariant anchors and representations; and (3) applying QDOT-based alignment techniques and QDOT as a loss function in complex models.

ETHICS STATEMENT

All datasets utilized in this study are publicly available, and their use raises no ethical concerns.

REPRODUCIBILITY STATEMENT

We are committed to ensuring the reproducibility of our research. The source code for all experiments is provided in the supplementary materials, which can be used to reproduce the results presented in this paper. For the molecule generation experiment, we have included the evaluate checkpoint of QDOT-0.3. This checkpoint can also be reproduced using the provided training code. Furthermore, we have explicitly stated all key assumptions for the theorems presented. The sources and preprocessing scripts for all datasets used in this work are also provided.

REFERENCES

- Sayali Anil Alatar and Daifeng Wang. CMOT: cross-modality optimal transport for multimodal inference. *Genome Biology*, 24(1):163, 2023.
- Jason Altschuler, Francis Bach, Alessandro Rudi, and Jonathan Niles-Weed. Massively scalable sinkhorn distances via the nyström method. *Advances in neural information processing systems*, 32, 2019.
- Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pp. 214–223. PMLR, 2017.
- Arip Asadulaev, Rostislav Korst, Aleksandr Korotin, Vage Egiazarian, Andrey Filchenkov, and Evgeny Burnaev. Rethinking optimal transport in offline reinforcement learning. *Advances in Neural Information Processing Systems*, 37:123592–123607, 2024.
- Simon Axelrod and Rafael Gomez-Bombarelli. Geom, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific Data*, 9(1):185, 2022.
- Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. *IEEE transactions on pattern analysis and machine intelligence*, 24(4):509–522, 2002.
- Nicolas Bonneel, Michiel Van De Panne, Sylvain Paris, and Wolfgang Heidrich. Displacement interpolation using lagrangian mass transport. In *Proceedings of the 2011 SIGGRAPH Asia conference*, pp. 1–12, 2011.
- Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and Radon Wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51:22–45, 2015.
- Chao Chen, Guanbin Li, Ruijia Xu, Tianshui Chen, Meng Wang, and Liang Lin. Clusternet: Deep hierarchical cluster network with rigorously rotation-invariant representation for point cloud analysis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4994–5002, 2019.
- Samir Chowdhury, David Miller, and Tom Needham. Quantized gromov-wasserstein. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 811–827. Springer, 2021.
- Scott Cohen and L Guibas. The earth mover’s distance under transformation sets. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pp. 1076–1083. IEEE, 1999.
- Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.
- Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. *Advances in neural information processing systems*, 34:17695–17709, 2021.

- Ishan Deshpande, Yuan-Ting Hu, Ruoyu Sun, Ayis Pyrros, Nasir Siddiqui, Sanmi Koyejo, Zhizhen Zhao, David Forsyth, and Alexander G Schwing. Max-sliced Wasserstein distance and its use for GANs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10648–10656, 2019.
- Haoqiang Fan, Hao Su, and Leonidas J Guibas. A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 605–613, 2017.
- Shikun Feng, Yuyan Ni, Lu yan, Zhi-Ming Ma, Wei-Ying Ma, and Yanyan Lan. UniGEM: A unified approach to generation and property prediction for molecules. In *The Thirteenth International Conference on Learning Representations*, 2025. URL <https://openreview.net/forum?id=Lb91pXwZMR>.
- Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, et al. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021.
- Nicolas Fournier and Arnaud Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability theory and related fields*, 162(3):707–738, 2015.
- Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. SE (3)-transformers: 3d rotation equivariant attention networks. *Advances in neural information processing systems*, 33:1970–1981, 2020.
- Michael Gromov. Groups of polynomial growth and expanding maps (with an appendix by Jacques Tits). *Publications Mathématiques de l’IHÉS*, 53:53–78, 1981.
- Majdi Hassan, Nikhil Shenoy, Jungyoon Lee, Hannes Stärk, Stephan Thaler, and Dominique Beaini. Et-flow: Equivariant flow-matching for molecular conformer generation. *Advances in Neural Information Processing Systems*, 37:128798–128824, 2024.
- Emiel Hoogeboom, Victor Garcia Satorras, Clement Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, pp. 8867–8887. PMLR, 2022.
- Tanguy Kerdoncuff, Remi Emonet, and Marc Sebban. Sampled gromov wasserstein. *Machine Learning*, 110(8):2151–2186, 2021.
- Seohyun Kim, Jaeyoo Park, and Bohyung Han. Rotation-invariant local-to-global representation learning for 3d point cloud. *Advances in Neural Information Processing Systems*, 33:8174–8185, 2020.
- Pascal Klink, Haoyi Yang, Carlo D’Eramo, Jan Peters, and Joni Pajarinen. Curriculum reinforcement learning via constrained optimal transport. In *International Conference on Machine Learning*, pp. 11341–11358. PMLR, 2022.
- Matt Kusner, Yu Sun, Nicholas Kolkin, and Kilian Weinberger. From word embeddings to document distances. In *International conference on machine learning*, pp. 957–966. PMLR, 2015.
- Tam Le, Makoto Yamada, Kenji Fukumizu, and Marco Cuturi. Tree-sliced variants of wasserstein distances. *Advances in neural information processing systems*, 32, 2019.
- Mengyu Li, Jun Yu, Tao Li, and Cheng Meng. Importance sparsification for sinkhorn algorithm. *Journal of Machine Learning Research*, 24(247):1–44, 2023a.
- Mengyu Li, Jun Yu, Hongteng Xu, and Cheng Meng. Efficient approximation of Gromov-Wasserstein distance using importance sparsification. *Journal of Computational and Graphical Statistics*, 32(4):1512–1523, 2023b.
- Tao Li, Cheng Meng, Hongteng Xu, and Jun Yu. Hilbert curve projection distance for distribution comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(7):4993–5007, 2024.

- Tao Li, Cheng Meng, Hongteng Xu, and Jun Zhu. Efficient variants of wasserstein distance in hyperbolic space via space-filling curve projection. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.
- Igor Melnyk, Youssef Mroueh, Brian Belgodere, Mattia Rigotti, Apoorva Nitsure, Mikhail Yurochkin, Kristjan Greenewald, Jiri Navratil, and Jarret Ross. Distributional preference alignment of llms via optimal transport. *Advances in Neural Information Processing Systems*, 37: 104412–104442, 2024.
- Facundo Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11:417–487, 2011.
- Gabriel Peyré, Marco Cuturi, and Justin Solomon. Gromov-Wasserstein averaging of kernel and distance matrices. In *International conference on machine learning*, pp. 2664–2672. PMLR, 2016.
- Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.
- Raghuathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):1–7, 2014.
- Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover’s distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121, 2000.
- Meyer Scetbon, Gabriel Peyré, and Marco Cuturi. Linear-time Gromov Wasserstein distances using low rank couplings and costs. In *International Conference on Machine Learning*, pp. 19347–19365. PMLR, 2022.
- RJ Serfling. Approximation theorems of mathematical statistics. *Wiley series in probability and mathematical statistics Show all parts in this series*, 1980.
- Liangliang Shi, Jack Fan, and Junchi Yan. Ot-clip: Understanding and generalizing clip via optimal transport. In *Forty-first International Conference on Machine Learning*, 2024.
- Yuxuan Song, Jingjing Gong, Minkai Xu, Ziyao Cao, Yanyan Lan, Stefano Ermon, Hao Zhou, and Wei-Ying Ma. Equivariant flow matching with hybrid probability transport for 3d molecule generation. *Advances in Neural Information Processing Systems*, 36:549–568, 2023.
- Yuxuan Song, Jingjing Gong, Hao Zhou, Mingyue Zheng, Jingjing Liu, and Wei-Ying Ma. Unified generative modeling of 3d molecules with bayesian flow networks. In *The Twelfth International Conference on Learning Representations*, 2024.
- Karl-Theodor Sturm. On the geometry of metric measure spaces. *Acta Mathematica*, 196(1):65–131, 2006.
- Karl-Theodor Sturm. *The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces*, volume 290. American Mathematical Society, 2023.
- Robert W Sumner and Jovan Popović. Deformation transfer for triangle meshes. *ACM Transactions on graphics (TOG)*, 23(3):399–405, 2004.
- Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. A concise and provably informative multi-scale signature based on heat diffusion. *Computer Graphics Forum*, 28(5):1383–1392, 2019.
- Federico Thomas and Lluís Ros. Revisiting trilateration for robot localization. *IEEE Transactions on robotics*, 21(1):93–101, 2005.
- Vayer Titouan, Rémi Flamary, Nicolas Courty, Romain Tavenard, and Laetitia Chapel. Sliced Gromov-Wasserstein. *Advances in Neural Information Processing Systems*, 32, 2019.
- Alexander Tong, Kilian FATRAS, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrod Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL <https://openreview.net/forum?id=CD9Snc73AW>. Expert Certification.

- Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2008.
- Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1912–1920, 2015.
- Hongteng Xu, Dixin Luo, Hongyuan Zha, and Lawrence Carin Duke. Gromov-Wasserstein learning for graph matching and node embedding. In *International conference on machine learning*, pp. 6932–6941. PMLR, 2019.
- Minkai Xu, Alexander S Powers, Ron O Dror, Stefano Ermon, and Jure Leskovec. Geometric latent diffusion models for 3d molecule generation. In *International Conference on Machine Learning*, pp. 38592–38610. PMLR, 2023.
- Yingxue Xu and Hao Chen. Multimodal optimal transport-based co-attention transformer with global structure consistency for survival prediction. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 21241–21251, 2023.
- Jianyu Yang, Hongxing Wang, Junsong Yuan, Youfu Li, and Jianyang Liu. Invariant multi-scale descriptor for shape representation, matching and retrieval. *Computer Vision and Image Understanding*, 145:43–58, 2016.
- Hao Yu, Zheng Qin, Ji Hou, Mahdi Saleh, Dongsheng Li, Benjamin Busam, and Slobodan Ilic. Rotation-invariant transformer for point cloud matching. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5384–5393, 2023.

A ADDITIONAL BACKGROUND

Here we introduce additional background concepts that are related to our work.

One-Dimensional Wasserstein Distance. A key special case of the Wasserstein distance defined in equation 1 occurs in one dimension, where it admits a convenient closed-form solution. For two one-dimensional distributions μ_X and μ_Y , let F_X and F_Y be their respective cumulative distribution functions, and let F_X^{-1} and F_Y^{-1} be the corresponding quantile functions. The p -Wasserstein distance is then given by:

$$\mathcal{W}_p(\mu_X, \mu_Y) = \left(\int_0^1 |F_X^{-1}(u) - F_Y^{-1}(u)|^p du \right)^{\frac{1}{p}}. \quad (9)$$

For discrete distributions, the empirical quantile functions are readily obtained by sorting, making the computation of the 1D Wasserstein distance explicit and efficient.

Gromov-Wasserstein Distance (Mémoli, 2011). While the original formulation of the Gromov-Wasserstein distance in equation 2 is theoretically elegant, its computation is NP-hard. Consequently, an alternative and more commonly used formulation was introduced. Given two mm-spaces $\mathcal{X} = (X, d_X, \mu_X)$ and $\mathcal{Y} = (Y, d_Y, \mu_Y)$, for $p \geq 1$, the p -Gromov-Wasserstein distance is defined as:

$$\mathcal{GW}_p(\mathcal{X}, \mathcal{Y}) := \left(\inf_{\pi \in \Pi(\mu_X, \mu_Y)} \int_{X \times Y} \int_{X \times Y} |d_X(x, x') - d_Y(y, y')|^p d\pi(x, y) d\pi(x', y') \right)^{\frac{1}{p}}. \quad (10)$$

Mémoli (2011) proved that the formulation in equation 10 is bi-Hölder equivalent to the one in equation 2. Moreover, equation 2 constitutes a metric in the sense of Definition 1.

Computational Costs. We now briefly review the computational complexity of these OT-based methods for discrete distributions supported on n points. Computing the standard Wasserstein distance in equation 1 using a classic EMD solver has a worst-case complexity of $\mathcal{O}(n^3)$ and an average-case complexity of $\mathcal{O}(n^2)$ (Bonneel et al., 2011). Alternatively, solving the entropy-regularized OT problem with the Sinkhorn algorithm (Cuturi, 2013) has a complexity of up to $\mathcal{O}(n^2 \log n)$. While some targeted algorithms may achieve faster runtimes (Altschuler et al., 2019; Li et al., 2023a), this often comes at the cost of sacrificing metric properties. Computing the Gromov-Wasserstein distance in equation 10 involves solving a non-convex quadratic program, which has a complexity of at least $\mathcal{O}(n^3 \log n)$ (Xu et al., 2019). Faster algorithms, such as those employing entropic regularization (Peyré et al., 2016; Scetbon et al., 2022; Li et al., 2023b), typically represent a trade-off between speed and accuracy. As for the one-dimensional Wasserstein distance in equation 9, the computation is significantly faster, requiring only $\mathcal{O}(n \log n)$ time via sorting. This efficiency has motivated many methods that project higher-dimensional distributions onto one-dimensional lines for comparison (Bonneel et al., 2015; Deshpande et al., 2019; Le et al., 2019; Li et al., 2024; 2025). However, for the Gromov-Wasserstein distance, the development of sliced variants is less mature, with Sliced Gromov-Wasserstein (SGW) (Titouan et al., 2019) being the most prominent example that has achieved empirical success.

B PROOFS

B.1 PROOF OF THEOREM 1

The proof for the symmetry and triangle inequality properties is straightforward, as they are directly inherited from the Wasserstein distance \mathcal{W}_p . We focus on proving the **Identity of Indiscernibles**, which is the most involved part.

To do so, we first recall an equivalent characterization of isometry from (Sturm, 2023, Lemma 1.10):

- (i) \mathcal{X}, \mathcal{Y} are isometric.

- (ii) There exists a coupling $\pi \in \Pi(\mu_X, \mu_Y)$ such that $d_X(x_0, x_1) = d_Y(y_0, y_1)$ holds for $\pi \otimes \pi$ -almost every pair of points $((x_0, y_0), (x_1, y_1))$.

The "if" part is straightforward: if \mathcal{X} and \mathcal{Y} are isometric, their QDMD representations will be identically distributed by construction, making their QDOT distance zero. We now prove the "only if" part: if $\mathcal{QD}_p(\mathcal{X}, \mathcal{Y}) = 0$, then \mathcal{X} and \mathcal{Y} are isometric by satisfying condition (ii).

Proof. By definition, the condition $\mathcal{QD}_p(\mathcal{X}, \mathcal{Y}) = 0$ implies that there exists an optimal coupling π between the pushforward measures on the representation space, such that $\phi^X(x) = \phi^Y(y)$ holds for all (x, y) in the support of the corresponding coupling on $X \times Y$.

For a given quantile level vector $\mathbf{q} \in (0, 1)^k$, let us denote the corresponding QDMs as $\mathbf{m}_i^X = \mathbf{m}^X(q_i)$ and $\mathbf{m}_i^Y = \mathbf{m}^Y(q_i)$ for $i = 1, 2, \dots, k$. The equality of the representations implies:

$$\begin{cases} \|\mathbf{x} - \mathbf{m}_i^X\|_2 = \|\mathbf{y} - \mathbf{m}_i^Y\|_2, \\ \|\mathbf{x}\|_2 = \|\mathbf{y}\|_2, \end{cases} \quad \forall (\mathbf{x}, \mathbf{y}) \in \text{supp}(\pi).$$

From $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2$ and the fact that isometric spaces have identical norm distributions (and thus identical quantile functions), it directly follows that the weights are equal for coupled points:

$$w^X(\mathbf{x}, q_i) = w^Y(\mathbf{y}, q_i), \quad \forall (\mathbf{x}, \mathbf{y}) \in \text{supp}(\pi), i = 1, 2, \dots, k.$$

For any $(\mathbf{x}, \mathbf{y}) \in \text{supp}(\pi)$, expanding the squared distance equality yields:

$$\begin{aligned} \|\mathbf{x} - \mathbf{m}_i^X\|_2^2 &= \|\mathbf{y} - \mathbf{m}_i^Y\|_2^2 \\ \Rightarrow \mathbf{x}^\top \mathbf{x} + \mathbf{m}_i^{X\top} \mathbf{m}_i^X - 2\mathbf{x}^\top \mathbf{m}_i^X &= \mathbf{y}^\top \mathbf{y} + \mathbf{m}_i^{Y\top} \mathbf{m}_i^Y - 2\mathbf{y}^\top \mathbf{m}_i^Y \\ \Rightarrow \mathbf{m}_i^{X\top} \mathbf{m}_i^X - 2\mathbf{x}^\top \mathbf{m}_i^X &= \mathbf{m}_i^{Y\top} \mathbf{m}_i^Y - 2\mathbf{y}^\top \mathbf{m}_i^Y, \end{aligned}$$

where the last step uses $\|\mathbf{x}\|_2^2 = \|\mathbf{y}\|_2^2$. Integrating this equality against the weight function $w(\cdot, q_j)$ over the respective spaces gives:

$$\begin{aligned} &\int_X w^X(q_j, \mathbf{x})(\mathbf{m}_i^{X\top} \mathbf{m}_i^X - 2\mathbf{x}^\top \mathbf{m}_i^X) d\mu_X(\mathbf{x}) \\ &= \int_Y w^Y(q_j, \mathbf{y})(\mathbf{m}_i^{Y\top} \mathbf{m}_i^Y - 2\mathbf{y}^\top \mathbf{m}_i^Y) d\mu_Y(\mathbf{y}) \\ &= \mathbb{E}_X(w^X(q_j, X))\mathbf{m}_i^{X\top} \mathbf{m}_i^X - 2\mathbb{E}_X(w^X(q_j, X)X) \\ &= \mathbb{E}_Y(w^Y(q_j, Y))\mathbf{m}_i^{Y\top} \mathbf{m}_i^Y - 2\mathbb{E}_Y(w^Y(q_j, Y)Y). \end{aligned}$$

By the definition of QDM in Eq. equation 5, this simplifies to:

$$\mathbf{m}_i^{X\top} \mathbf{m}_i^X - 2\mathbf{m}_i^{X\top} \mathbf{m}_j^X = \mathbf{m}_i^{Y\top} \mathbf{m}_i^Y - 2\mathbf{m}_i^{Y\top} \mathbf{m}_j^Y.$$

By setting $i = j$, we find $\|\mathbf{m}_i^X\|_2^2 = \|\mathbf{m}_i^Y\|_2^2$, which implies $\|\mathbf{m}_i^X\|_2 = \|\mathbf{m}_i^Y\|_2$. Substituting this back, we obtain the key result that the Gram matrices of the QDM sets are identical: $\mathbf{m}_i^{X\top} \mathbf{m}_j^X = \mathbf{m}_i^{Y\top} \mathbf{m}_j^Y$.

Now, for any $(\mathbf{x}, \mathbf{y}) \in \text{supp}(\pi)$, the relation $\|\mathbf{x} - \mathbf{m}_i^X\|_2^2 = \|\mathbf{y} - \mathbf{m}_i^Y\|_2^2$ simplifies further:

$$\begin{aligned} \mathbf{x}^\top \mathbf{x} + \mathbf{m}_i^{X\top} \mathbf{m}_i^X - 2\mathbf{x}^\top \mathbf{m}_i^X &= \mathbf{y}^\top \mathbf{y} + \mathbf{m}_i^{Y\top} \mathbf{m}_i^Y - 2\mathbf{y}^\top \mathbf{m}_i^Y \\ \Rightarrow \mathbf{x}^\top \mathbf{m}_i^X &= \mathbf{y}^\top \mathbf{m}_i^Y. \end{aligned}$$

Due to the dimensionality condition $\dim(\{\mathbf{m}^X(q_i)\}_{1 \leq i \leq k}) = \dim(X)$, any point $\mathbf{x} \in \text{supp}(\mu_X)$ can be written as a linear combination $\mathbf{x} = \sum_{i=1}^k \alpha_i \mathbf{m}_i^X$ for some coefficients $\alpha_1, \dots, \alpha_k$. We can show that its coupled counterpart \mathbf{y} must be the same linear combination of the corresponding

QDMs in Y :

$$\begin{aligned}
& \|\mathbf{y} - \sum_{i=1}^k \alpha_i \mathbf{m}_i^Y\|_2^2 \\
&= \|\mathbf{y}\|_2^2 + \left\| \sum_{i=1}^k \alpha_i \mathbf{m}_i^Y \right\|_2^2 - 2 \sum_{i=1}^k \alpha_i \mathbf{y}^\top \mathbf{m}_i^Y \\
&= \|\mathbf{y}\|_2^2 + \sum_{i=1}^k \sum_{j=1}^k \alpha_i \alpha_j \mathbf{m}_i^{Y\top} \mathbf{m}_j^Y - 2 \sum_{i=1}^k \alpha_i \mathbf{x}^\top \mathbf{m}_i^X \\
&= \|\mathbf{y}\|_2^2 + \sum_{i=1}^k \sum_{j=1}^k \alpha_i \alpha_j \mathbf{m}_i^{X\top} \mathbf{m}_j^X - 2 \mathbf{x}^\top \sum_{i=1}^k \alpha_i \mathbf{m}_i^X \\
&= \|\mathbf{y}\|_2^2 + \mathbf{x}^\top \mathbf{x} - 2 \mathbf{x}^\top \mathbf{x} = 0,
\end{aligned}$$

where the last equality uses $\|\mathbf{x}\|_2^2 = \|\mathbf{y}\|_2^2$. This implies $\mathbf{y} = \sum_{i=1}^k \alpha_i \mathbf{m}_i^Y$.

Finally, for any two pairs $(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2) \in \text{supp}(\pi)$, let $\mathbf{x}_1 = \sum_{i=1}^k \alpha_i \mathbf{m}_i^X$ and $\mathbf{x}_2 = \sum_{j=1}^k \beta_j \mathbf{m}_j^X$. It follows that their distances are preserved:

$$\begin{aligned}
& \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 - \|\mathbf{y}_1 - \mathbf{y}_2\|_2^2 \\
&= (\|\mathbf{x}_1\|_2^2 - \|\mathbf{y}_1\|_2^2) + (\|\mathbf{x}_2\|_2^2 - \|\mathbf{y}_2\|_2^2) - (2\mathbf{x}_1^\top \mathbf{x}_2 - 2\mathbf{y}_1^\top \mathbf{y}_2) \\
&= -2 \sum_{i=1}^k \sum_{j=1}^k \alpha_i \beta_j (\mathbf{m}_i^{X\top} \mathbf{m}_j^X - \mathbf{m}_i^{Y\top} \mathbf{m}_j^Y) = 0.
\end{aligned}$$

Thus, $d_X(\mathbf{x}_1, \mathbf{x}_2) = d_Y(\mathbf{y}_1, \mathbf{y}_2)$ holds for $\pi \otimes \pi$ -a.e. pairs, fulfilling condition (ii) and completing the proof. \square

B.2 PROOF OF THEOREM 2

This section provides the proof for Theorem 2. We begin by establishing the necessary notation and preliminary lemmas. Let $r_q = F_{\|X\|_2}^{-1}(q)$ be the true q -th quantile of the norm distribution, and let $\hat{r}_q = \hat{F}_{\|\mathbb{X}_n\|_2}^{-1}(q)$ be its empirical counterpart estimated from the sample set $\mathbb{X}_n = \{x_1, \dots, x_n\}$. For brevity, we denote the weight functions as $w_i^X(x) = w^X(x, q_i)$ and its empirical version as $\hat{w}_i^{\mathbb{X}_n}(x) = e^{-\sigma(\|x\|_2 - \hat{r}_q)^2}$.

Our proof relies on the well-established convergence rates for the Wasserstein distance, summarized in the following lemma.

Lemma 1 (Convergence Rate of Wasserstein Distance (Fournier & Guillin, 2015)) *For a distribution μ defined on a vector measure space $(X, \mathcal{B}(X))$ and its empirical version μ_n , if the q -th moment $\mathcal{M}_q(\mu)$ is finite, i.e., $\mathcal{M}_q(\mu) = \int_X \|x\|^q d\mu(x) < \infty$, then the following holds:*

$$\begin{aligned}
& \mathbb{E}(\mathcal{W}_p(\mu_n, \mu)) \leq C M_q^{p/q}(\mu) \\
& \times \begin{cases} n^{-1/2} + n^{-(q-p)/q} & \text{if } p > d/2 \text{ and } q \neq 2p \\ n^{-1/2} \log(1+n) + n^{-(q-p)/q} & \text{if } p = d/2 \text{ and } q \neq 2p \\ n^{-p/d} + n^{-(q-p)/q} & \text{if } p \in (0, d/2) \text{ and } q \neq d/(d-p) \end{cases}
\end{aligned}$$

As discussed in (Fournier & Guillin, 2015), for a sufficiently large moment order q , the term $n^{-(q-p)/q}$ becomes negligible compared to the leading term. For the common case of $p = 2$, this holds for $q \geq 4$. We will assume this condition holds in our subsequent analysis.

Next, we state the standard convergence rate for empirical quantiles.

Lemma 2 (Convergence Rate of Quantiles) *Let F_X be the CDF of a distribution μ_X , and assume it is continuous on r_q and has a strictly positive density. Then, the mean squared error of the empirical quantile converges as follows:*

$$\mathbb{E}|r_q - \hat{r}_q|^2 = O(n^{-1}).$$

This is a classical result in asymptotic theory, which can be found in (Serfling, 1980, Chap 2.3).

Furthermore, our proof requires the following bounds. The first lemma ensures that the denominator in the QDM definition is well-behaved.

Lemma 3 *If $\mathbb{E}\|X\|_2 < \infty$, then for any given quantile level q_i , the expected weight is bounded away from zero, i.e., $\frac{1}{\mathbb{E}_{\mu_X}(w_i^X(X))} < \infty$.*

Proof. Since $\mathbb{E}_{\mu_X}(\|X\|_2) < \infty$, for any given $p_0 < 1$, there exists an M_0 such that $\Pr(\|X\|_2 \leq M_0) \geq p_0$. Therefore, we have:

$$\begin{aligned} \mathbb{E}_{\mu_X}(w_i^X(X)) &\geq \int_{\|x\|_2 < M_0} e^{-\sigma(x-r_{q_i})^2} d\mu_X(x) \geq \int_{\|x\|_2 < M_0} e^{-\sigma \max\{x, r_{q_i}\}^2} d\mu_X(x) \\ &\geq \int_{\|x\|_2 < M_0} e^{-\sigma \max\{M_0, r_{q_i}\}^2} d\mu_X(x) \geq p_0 e^{-\sigma \max\{M_0, r_{q_i}\}^2}. \end{aligned}$$

This implies that the expectation is strictly positive, and thus $\frac{1}{\mathbb{E}_{\mu_X}(w_i^X(X))} < \infty$. \square

Lemma 4 *For any quantile level q_i , let the normalized empirical weights be $\hat{p}_j = \frac{\hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_j)}{\sum_{j'=1}^n \hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_{j'})}$. Then for any power $1 \leq \alpha < \infty$, we have the following bound:*

$$\sum_{j=1}^n \mathbb{E}(\hat{p}_j^\alpha)^{1/\alpha} < \infty.$$

Proof. Let $M_\alpha(\hat{w}_i) = \mathbb{E}(\hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_j)^\alpha)$. Since $\hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_j) < 1$, it is clear that $M_\alpha(\hat{w}_i) < 1$. Define the sample mean of the weights as $\bar{w}(r) = \frac{1}{n} \sum_{j=1}^n e^{-\sigma(\|\mathbf{x}_j\|_2 - r)^2}$. We then have $\mathbb{E}(\bar{w}(\hat{r}_{q_i})) = M_1(\hat{w}_i)$. Now, consider the following probability:

$$\mathbb{P}(\bar{w}_i(\hat{r}_{q_i}) < \frac{1}{2} M_1(\hat{w}_i)) \leq \underbrace{\mathbb{P}(\bar{w}_i(r_{q_i}) < \frac{3}{4} M_1(\hat{w}_i))}_{\mathbb{P}_1} + \underbrace{\mathbb{P}(|\bar{w}_i(r_{q_i}) - \bar{w}_i(\hat{r}_{q_i})| > \frac{1}{4} M_1(\hat{w}_i))}_{\mathbb{P}_2}.$$

For the first term, \mathbb{P}_1 , we note that $\mathbb{E}(\bar{w}_i(r_{q_i})) = M_1(\hat{w}_i)$ and the terms in the sum are i.i.d. Applying Hoeffding's inequality yields:

$$\begin{aligned} \mathbb{P}_1 &= \mathbb{P}(\mathbb{E}(\bar{w}_i(r_{q_i})) - \bar{w}_i(r_{q_i}) > \frac{n}{4} M_1(\hat{w}_i)) \\ &= \mathbb{P}\left(\sum_{j=1}^n \mathbb{E}(w_i^X(\mathbf{x}_j)) - \sum_{j=1}^n w_i^X(\mathbf{x}_j) > \frac{n}{4} M_1(\hat{w}_i)\right) \\ &\leq \exp(-\frac{n}{8} M_1^2(\hat{w}_i)). \end{aligned}$$

For the second term, \mathbb{P}_2 , we first bound the difference by applying the Mean Value Theorem:

$$\begin{aligned} |\bar{w}_i(r_{q_i}) - \bar{w}_i(\hat{r}_{q_i})| &\leq \frac{1}{n} \sum_{j=1}^n |e^{-\sigma(\|\mathbf{x}_j\|_2 - r_{q_i})^2} - e^{-\sigma(\|\mathbf{x}_j\|_2 - \hat{r}_{q_i})^2}| \\ &\leq \frac{1}{n} \sum_{j=1}^n |2\sigma(\|\mathbf{x}_j\|_2 - r_{j0}) e^{-\sigma(\|\mathbf{x}_j\|_2 - r_{j0})^2}| |r_{q_i} - \hat{r}_{q_i}| \quad (11) \\ &\leq \sqrt{2\sigma} e^{-1/2} |r_{q_i} - \hat{r}_{q_i}| \quad (\text{since } \sup_{t \geq 0} 2\sigma t e^{-\sigma t^2} \leq \sqrt{2\sigma} e^{-1/2}). \end{aligned}$$

Consequently, we have:

$$\mathbb{P}_2 \leq \mathbb{P}\left(|r_{q_i} - \hat{r}_{q_i}| > \sqrt{\frac{e}{32\sigma}} M_1(\hat{w}_i)\right) \leq 2 \exp(-2n\delta),$$

where $\delta = \min\{F_X(r_{p_i} + \sqrt{\frac{e}{32\sigma}} M_1(\hat{w}_i)) - p_i, p_i - F_X(r_{p_i} - \sqrt{\frac{e}{32\sigma}} M_1(\hat{w}_i))\} > 0$. The final inequality follows from a standard result on the concentration of empirical quantiles, see, e.g.,

(Serfling, 1980, Theorem 2.3.2). By the symmetry of the i.i.d. samples, $\mathbb{E}(\hat{p}_j^\alpha) = \mathbb{E}(\hat{p}_l^\alpha)$ for all $j, l \in \{1, \dots, n\}$. Thus we can write:

$$\begin{aligned}
\sum_{j=1}^n \mathbb{E}(\hat{p}_j^\alpha)^{1/\alpha} &= n \mathbb{E}(\hat{p}_1^\alpha)^{1/\alpha} \\
&= \mathbb{E}\left(\frac{\hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_1)^\alpha}{\bar{w}_i(\hat{r}_{q_i})^\alpha}\right)^{1/\alpha} \\
&\leq \mathbb{E}_{\{\bar{w}_i(\hat{r}_{q_i}) \geq \frac{1}{2}M_1(\hat{w}_i)\}}\left(\frac{\hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_1)^\alpha}{\bar{w}_i(\hat{r}_{q_i})^\alpha}\right)^{1/\alpha} + n\mathbb{P}(\bar{w}_i(\hat{r}_{q_i}) < \frac{1}{2}M_1(\hat{w}_i)) \\
&\leq \frac{2\mathbb{E}(\hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_1)^\alpha)^{1/\alpha}}{M_1(\hat{w}_i)} + n \exp(-\frac{n}{8}M_1^2(\hat{w}_i)) + 2n \exp(-2n\delta) \\
&\leq \frac{2M_\alpha(\hat{w}_i)^{1/\alpha}}{M_1(\hat{w}_i)} + \frac{8e^{-1}}{M_1^2(\hat{w}_i)} + \frac{e^{-1}}{\delta} < \infty.
\end{aligned}$$

The final line is finite, leveraging the result from Lemma 3 which ensures $M_1(\cdot)$ is bounded away from zero. This completes the proof. \square

Lemma 5 *If the fourth moment of the norm is finite, i.e., $\mathbb{E}\|X\|_2^4 < \infty$, then the expected squared norm of the empirical QDM is also finite: $\mathbb{E}\|\hat{m}_i^{\mathbb{X}_n}\|_2^2 < \infty$.*

Proof. The proof follows from applying the Cauchy-Schwarz or Hölder's inequality and leveraging the result from Lemma 4.

$$\begin{aligned}
\mathbb{E}\|\hat{m}_i^{\mathbb{X}_n}\|_2^2 &= \sum_{j=1}^n \sum_{l=1}^n \mathbb{E}(\hat{p}_j \hat{p}_l \mathbf{x}_j^\top \mathbf{x}_l) \\
&\leq \sum_{j=1}^n \sum_{l=1}^n \mathbb{E}(\hat{p}_j^4)^{1/4} \mathbb{E}(\hat{p}_l^4)^{1/4} \mathbb{E}((\mathbf{x}_j^\top \mathbf{x}_l)^2)^{1/2} \\
&\leq \sum_{j=1}^n \mathbb{E}(\hat{p}_j^4)^{1/2} \mathbb{E}(\|\mathbf{x}_j\|_2^4)^{1/2} + \sum_{j \neq l} \mathbb{E}(\hat{p}_j^4)^{1/4} \mathbb{E}(\hat{p}_l^4)^{1/4} \mathbb{E}(\|\mathbf{x}_j\|_2^2 \|\mathbf{x}_l\|_2^2)^{1/2} \\
&= \mathbb{E}(\|X\|_2^4)^{1/2} \sum_{j=1}^n \mathbb{E}(\hat{p}_j^4)^{1/2} + \mathbb{E}(\|X\|_2^2) \left(\sum_{j=1}^n \mathbb{E}(\hat{p}_j^4)^{1/2}\right)^2 < \infty.
\end{aligned}$$

\square

Lemma 6 *If $\mathbb{E}\|X\|_2^4 < \infty$, then the expectation of the product of the empirical QDM norm, a sample norm, and the quantile error converges at the rate of $O(n^{-1/2})$: $\mathbb{E}(\|\hat{m}_i^{\mathbb{X}_n}\|_2 \|\mathbf{x}_1\|_2 |r_{q_i} - \hat{r}_{q_i}|) = O(n^{-1/2})$.*

Proof. This proof also relies on the Cauchy-Schwarz inequality and the results from Lemma 2 and Lemma 4.

$$\begin{aligned}
\mathbb{E}(\|\hat{m}_i^{\mathbb{X}_n}\|_2 \|\mathbf{x}_1\|_2 |r_{q_i} - \hat{r}_{q_i}|) &\leq \mathbb{E}(\hat{q}_1 \|\mathbf{x}_1\|_2^2 |r_{q_i} - \hat{r}_{q_i}|) + \sum_{j=2}^n \mathbb{E}(\hat{q}_j \|\mathbf{x}_1\|_2 \|\mathbf{x}_j\|_2 |r_{q_i} - \hat{r}_{q_i}|) \\
&\leq \mathbb{E}(\|X\|_2^4)^{1/2} \mathbb{E}(|r_{q_i} - \hat{r}_{q_i}|^2)^{1/2} \\
&\quad + \sum_{j=2}^n \mathbb{E}(\hat{q}_j^4)^{1/4} \mathbb{E}(\|\mathbf{x}_1\|_2^4 \|\mathbf{x}_j\|_2^4)^{1/4} \mathbb{E}(|r_{q_i} - \hat{r}_{q_i}|^2)^{1/2} \\
&= \mathbb{E}(|r_{q_i} - \hat{r}_{q_i}|^2)^{1/2} \mathbb{E}(\|X\|_2^4)^{1/2} \left(1 + \sum_{j=2}^n \mathbb{E}(\hat{q}_j^4)^{1/4}\right).
\end{aligned}$$

From Lemma 2, we know that $\mathbb{E}(|r_{q_i} - \hat{r}_{q_i}|^2)^{1/2} = O(n^{-1/2})$. Since $\mathbb{E}(\|X\|_2^4) < \infty$ is assumed, and Lemma 4 ensures that $\sum_{j=2}^n \mathbb{E}(\hat{q}_j^4)^{1/4} < \infty$, the result follows. \square

Proof of Theorem 2. First, we use a property derived from the triangle inequality:

$$\begin{aligned} |d(\mathbf{x}, \mathbf{x}') - d(\mathbf{y}, \mathbf{y}')| &\leq |d(\mathbf{x}, \mathbf{x}') - d(\mathbf{x}, \mathbf{y}')| + |d(\mathbf{x}, \mathbf{y}') - d(\mathbf{y}, \mathbf{y}')| \\ &\leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{x}', \mathbf{y}'). \end{aligned}$$

This allows us to bound the QDOT distance.

$$\begin{aligned} \mathcal{QD}_p(\mathcal{X}, \mathcal{Y})^p &= \inf_{\pi \in \Pi(\mu_X, \mu_Y)} \int \|\phi^X(\mathbf{x}, \mathbf{q}) - \phi^Y(\mathbf{y}, \mathbf{q})\|_p^p d\pi(\mathbf{x}, \mathbf{y}) \\ &\leq \inf_{\pi \in \Pi(\mu_X, \mu_Y)} \int \sum_{i=0}^k |\phi_i^X - \phi_i^Y|^p d\pi(\mathbf{x}, \mathbf{y}) \\ &\leq \inf_{\pi \in \Pi(\mu_X, \mu_Y)} \int \sum_{i=0}^k |d(\mathbf{x}, \mathbf{y}) + d(\mathbf{m}_i^X, \mathbf{m}_i^Y)|^p d\pi(\mathbf{x}, \mathbf{y}) \\ &\leq \inf_{\pi \in \Pi(\mu_X, \mu_Y)} \int (k+1)2^{p-1} d(\mathbf{x}, \mathbf{y})^p + 2^{p-1} \sum_{i=0}^k d(\mathbf{m}_i^X, \mathbf{m}_i^Y)^p d\pi(\mathbf{x}, \mathbf{y}) \quad (12) \\ &= (k+1)2^{p-1} \inf_{\pi \in \Pi(\mu_X, \mu_Y)} \int d(\mathbf{x}, \mathbf{y})^p d\pi(\mathbf{x}, \mathbf{y}) + 2^{p-1} \sum_{i=0}^k d(\mathbf{m}_i^X, \mathbf{m}_i^Y)^p \\ &= \underbrace{C_0 \mathcal{W}_p(\mu_X, \mu_Y)^p}_{\text{Wasserstein Discrepancy}} + \underbrace{C_1 \sum_{i=1}^k d(\mathbf{m}_i^X, \mathbf{m}_i^Y)^p}_{\text{QDM Discrepancy}} \end{aligned}$$

where $\mathbf{m}_0^X = \mathbf{0}$, $\mathbf{m}_0^Y = \mathbf{0}$, and the constants $C_0 = (k+1)2^{p-1}$ and $C_1 = 2^{p-1}$ are finite. From (Fournier & Guillin, 2015, Theorem 1), we have the bound for the Wasserstein Discrepancy between the true measure and its empirical version:

$$\mathbb{E}(\mathcal{W}_p(\mu_X, \mu_n)^p) \leq Cn^{-p/d}.$$

Next, we analyze the QDM Discrepancy term, $d(\mathbf{m}_i^X, \hat{\mathbf{m}}_i^{\mathbb{X}_n})$.

$$\begin{aligned} d(\mathbf{m}_i^X, \hat{\mathbf{m}}_i^{\mathbb{X}_n}) &= \left\| \frac{\frac{1}{n} \sum_{j=1}^n \hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_j) \mathbf{x}_j}{\frac{1}{n} \sum_{j=1}^n \hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_j)} - \frac{\mathbb{E}_{\mu_X}(w_i^X(X)X)}{\mathbb{E}_{\mu_X}(w_i^X(X))} \right\|_2 \\ &= \left\| \frac{\frac{1}{n} \sum_{j=1}^n \hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_j) \mathbf{x}_j \mathbb{E}_{\mu_X}(w_i^X(X)) - \mathbb{E}_{\mu_X}(w_i^X(X)X) \frac{1}{n} \sum_{j=1}^n \hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_j)}{\frac{1}{n} \sum_{j=1}^n \hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_j) \mathbb{E}_{\mu_X}(w_i^X(X))} \right\|_2 \\ &\leq C_2 \underbrace{\left| \mathbb{E}_{\mu_X}(w_i^X(X)) - \frac{1}{n} \sum_{j=1}^n w_i^X(\mathbf{x}_j) \right|}_{\text{Term (I)}} \\ &\quad + C_2 \underbrace{\left\| \hat{\mathbf{m}}_i^{\mathbb{X}_n} \right\|_2 \left\| \frac{1}{n} \sum_{j=1}^n \hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_j) \mathbf{x}_j - \mathbb{E}_{\mu_X}(w_i^X(X)X) \right\|_2}_{\text{Term (II)}}. \quad (13) \end{aligned}$$

Here, $C_2 = \frac{1}{\mathbb{E}_{\mu_X}(w_i^X(X))}$. By Lemma 3, under the condition that $q \geq 4$, C_2 is a finite constant. We now focus on the two main terms. Term (I) can be decomposed as:

$$\text{Term (I)} \leq \underbrace{\left| \mathbb{E}_{\mu_X}(w_i^X(X)) - \frac{1}{n} \sum_{j=1}^n w_i^X(\mathbf{x}_j) \right|}_{\text{Term (I.1)}} + \underbrace{\left| \frac{1}{n} \sum_{j=1}^n w_i^X(\mathbf{x}_j) - \frac{1}{n} \sum_{j=1}^n \hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_j) \right|}_{\text{Term (I.2)}}. \quad (14)$$

For Term (I.1), standard results for the mean of i.i.d. variables give:

$$\mathbb{E}(\text{Term (I.1)}) \leq \sqrt{\mathbb{E} \left(\mathbb{E}_{\mu_X}(w_i^X(X)) - \frac{1}{n} \sum_{j=1}^n w_i^X(\mathbf{x}_j) \right)^2} = \sqrt{\frac{1}{n} \text{Var}(w_i^X(\mathbf{x}_1))} = O(n^{-1/2}).$$

Using the bound from equation 11, we have:

$$\text{Term (I.2)} \leq \sqrt{2\sigma}e^{-1/2} |r_{q_i} - \hat{r}_{q_i}|.$$

Since $\mathbb{E}|r_{q_i} - \hat{r}_{q_i}| = O(n^{-1/2})$, it follows that $\mathbb{E}(\text{Term (I)}) = O(n^{-1/2})$. For Term (II), we use a similar decomposition:

$$\begin{aligned} \text{Term (II)} &\leq \underbrace{\|\hat{m}_i^{\mathbb{X}_n}\|_2 \left\| \frac{1}{n} \sum_{j=1}^n \hat{w}_i^{\mathbb{X}_n}(x_j) x_j - \frac{1}{n} \sum_{j=1}^n w_i^X(x_j) x_j \right\|_2}_{\text{Term (II.1)}} \\ &\quad + \underbrace{\|\hat{m}_i^{\mathbb{X}_n}\|_2 \left\| \sum_{j=1}^n w_i^X(x_j) x_j - \mathbb{E}_{\mu_X}(w_i^X(X)X) \right\|_2}_{\text{Term (II.2)}}. \end{aligned}$$

For Term (II.1), we have:

$$\begin{aligned} \text{Term (II.1)} &= \|\hat{m}_i^{\mathbb{X}_n}\|_2 \cdot \left\| \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j (e^{-\sigma(\|\mathbf{x}_j\|_2 - r_{q_i})} - e^{-\sigma(\|\mathbf{x}_j\|_2 - \hat{r}_{q_i})}) \right\|_2 \\ &\leq \frac{1}{n} \|\hat{m}_i^{\mathbb{X}_n}\|_2 \sum_{j=1}^n \left(\|\mathbf{x}_j\|_2 |e^{-\sigma(\|\mathbf{x}_j\|_2 - r_{q_i})} - e^{-\sigma(\|\mathbf{x}_j\|_2 - \hat{r}_{q_i})}| \right) \\ &\leq \frac{\sqrt{2\sigma}e^{-1/2}}{n} \sum_{j=1}^n \|\hat{m}_i^{\mathbb{X}_n}\|_2 \|\mathbf{x}_j\|_2 |r_{q_i} - \hat{r}_{q_i}|. \end{aligned}$$

From Lemma 6, we know $\mathbb{E}(\|\hat{m}_i^{\mathbb{X}_n}\|_2 \|\mathbf{x}_j\|_2 |r_{q_i} - \hat{r}_{q_i}|) = O(n^{-1/2})$, which implies $\mathbb{E}(\text{Term (II.1)}) = O(n^{-1/2})$. For Term (II.2), since $w_i^X(x) \leq 1$, we have $\mathbb{E}\|w(X)X\|_2^2 < \infty$. Using the Cauchy-Schwarz inequality, we get:

$$\begin{aligned} \mathbb{E}(\text{Term (II.2)}) &\leq \sqrt{\mathbb{E}\|\hat{m}_i^{\mathbb{X}_n}\|_2^2 \cdot \mathbb{E}\left(\mathbb{E}_{\mu_X} \|w_i^X(X)X - \frac{1}{n} \sum_{j=1}^n w_i^X(\mathbf{x}_j) \mathbf{x}_j\|_2^2\right)} \\ &= \sqrt{\mathbb{E}\|\hat{m}_i^{\mathbb{X}_n}\|_2^2} \sqrt{\frac{1}{n} \text{tr Cov}(w_i^X(\mathbf{x}_1) \mathbf{x}_1)} = O(n^{-1/2}). \end{aligned}$$

Thus, we also have $\mathbb{E}(\text{Term (II)}) = O(n^{-1/2})$. In summary, we have shown that $\mathbb{E}(d(\mathbf{m}_i^X, \hat{\mathbf{m}}_i^{\mathbb{X}_n})) = O(n^{-1/2})$, which means the convergence rate for the QDM Discrepancy is $O(n^{-p/2})$. For typical cases where $d \geq 2$, this rate is faster than the Wasserstein rate, so the overall convergence is dominated by the Wasserstein Discrepancy. Therefore, we obtain the final rate:

$$\mathbb{E}(\mathcal{QD}_p(\mathcal{X}, \mathcal{X}_n)) = O(n^{-1/d}).$$

□

B.3 PROOF OF THEOREM 3

The proof for the Identity of Indiscernibles is analogous to that of Theorem 1, and symmetry holds trivially by definition. Therefore, we focus on proving the Triangle Inequality. We begin by introducing the following well-known Gluing Lemma.

Lemma 7 (Gluing Lemma) *For three mm-spaces $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ with corresponding measures μ_X, μ_Y, μ_Z , and given couplings $\pi_{X,Z} \in \Pi(\mu_X, \mu_Z)$ and $\pi_{Y,Z} \in \Pi(\mu_Y, \mu_Z)$, there exists a joint coupling $\pi_{X,Y,Z} \in \Pi(\mu_X, \mu_Y, \mu_Z)$ such that its marginals satisfy*

$$(\text{proj}_{(X,Z)})_{\#} \pi_{X,Y,Z} = \pi_{X,Z}, \quad (\text{proj}_{(Y,Z)})_{\#} \pi_{X,Y,Z} = \pi_{Y,Z},$$

and its (X,Y) -marginal, $(\text{proj}_{(X,Y)})_{\#} \pi_{X,Y,Z}$, is a coupling in $\Pi(\mu_X, \mu_Y)$.

A proof of this lemma can be found in (Villani et al., 2008, Chap 1).

Proof of Triangle Inequality. Consider three mm-spaces $\mathcal{X} = (X, d_X, \mu_X)$, $\mathcal{Y} = (Y, d_Y, \mu_Y)$, and $\mathcal{Z} = (Z, d_Z, \mu_Z)$. Let $\pi_{X,Y}^* \in \Pi(\mu_X, \mu_Y)$, $\pi_{X,Z}^* \in \Pi(\mu_X, \mu_Z)$, and $\pi_{Y,Z}^* \in \Pi(\mu_Y, \mu_Z)$ denote the respective pairwise optimal couplings for the IQDOT distance. According to Lemma 7, we can find a joint coupling $\pi_{X,Y,Z} \in \Pi(\mu_X, \mu_Y, \mu_Z)$ such that

$$(\text{proj}_{(X,Z)})_{\#} \pi_{X,Y,Z} = \pi_{X,Z}^*, \quad (\text{proj}_{(Y,Z)})_{\#} \pi_{X,Y,Z} = \pi_{Y,Z}^*.$$

Let $\pi_{X,Y} := (\text{proj}_{(X,Y)})_{\#} \pi_{X,Y,Z}$. By construction, $\pi_{X,Y} \in \Pi(\mu_X, \mu_Y)$. We can now bound the IQDOT distance as follows:

$$\begin{aligned} \mathcal{IQD}_p(\mathcal{X}, \mathcal{Y}) &= \mathbb{E}_{(0,1) \times \pi_{X,Y}^*} (|\phi^X(x, q) - \phi^Y(y, q)|^p)^{1/p} \\ &\leq \mathbb{E}_{(0,1) \times \pi_{X,Y}} (|\phi^X(x, q) - \phi^Y(y, q)|^p)^{1/p} \\ &\leq \mathbb{E}_{(0,1) \times \pi_{X,Y,Z}} (|\phi^X(x, q) - \phi^Z(z, q)| + |\phi^Z(z, q) - \phi^Y(y, q)|^p)^{1/p} \\ &\leq \mathbb{E}_{(0,1) \times \pi_{X,Z}^*} (|\phi^X(x, q) - \phi^Z(z, q)|^p)^{1/p} + \mathbb{E}_{(0,1) \times \pi_{Y,Z}^*} (|\phi^Z(z, q) - \phi^Y(y, q)|^p)^{1/p} \\ &= \mathcal{IQD}_p(\mathcal{X}, \mathcal{Z}) + \mathcal{IQD}_p(\mathcal{Z}, \mathcal{Y}), \end{aligned}$$

where the last inequality follows from Minkowski's inequality. This completes the proof. \square

B.4 RELATIONSHIP BETWEEN QDOT AND IQDOT

It is evident that both QDOT and IQDOT utilize the same QDMD representations in their numerical implementations. Since IQDOT is formulated as a pairwise 1-D Wasserstein distance, the inequality $\mathcal{IQD} \leq \mathcal{QD}$ consistently holds in practice. Consequently, the convergence $\mathcal{QD}(\mu_n, \mu) \rightarrow 0$ implies $\mathcal{IQD}(\mu_n, \mu) \rightarrow 0$. Furthermore, as both constitute distinct well-defined metrics, under the requisite conditions, they share the equivalence property regarding in distinguish ability: $\mathcal{QD}(\mu, \nu) = 0 \Leftrightarrow \mathcal{IQD}(\mu, \nu) = 0$.

B.5 DISCUSSIONS ON THE SYMMETRIC SHAPES.

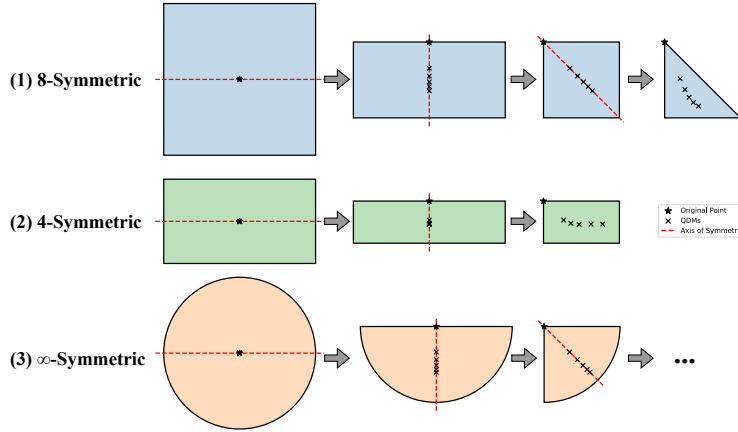


Figure 7: Illustrations of simple symmetric shapes. Note that for non-radially symmetric shapes, utilizing a finite number of QDMs suffices to satisfy the dimension conditions in Theorem 1 (i.e., ensuring that they are not collinear). For radially symmetric shapes, infinite subdivision implies a degeneration into a radial distribution.

C IMPLEMENTATION DETAILS

C.1 DETAILS OF ALGORITHM 1

The following provides further details on some of the operations and parameter choices in Algorithm 1.

Data Initialization. For two unknown distributions situated in Euclidean spaces, preprocessing the data is necessary. The objective is to obtain the distribution of distances relative to the barycenter, which is isometry-invariant. However, this initialization step is not necessary for intra-space comparisons or in cases where a correspondence between the origins is already known. The intuition of scalization aligns with the numerical implementation of Gromov-Wasserstein (Flamary et al., 2021), serving as a common technique for metric space alignment.

Quantile Level Vector q . The quantile levels in the vector q are theoretically chosen from the open interval $(0, 1)$. In practice, we select these levels by taking equispaced points within a truncated interval $[\delta, 1 - \delta]$. The default value for δ is 0.1, which avoids numerical instability at the extreme tails of the distribution. For larger datasets where the empirical quantiles are more stable, a smaller δ may be used to satisfy the theoretical conditions more closely.

Choice of the Number of Quantiles k . A large number of quantiles, k , is often not required to achieve strong performance. For low-dimensional data (e.g., 3D point clouds), the trilateration condition is readily satisfied, and a small value such as $k = 5$ can be sufficient. For high-dimensional data, which often exhibits low-rank structure, a value of k smaller than the ambient dimension can also be effective. While a larger k can lead to improved numerical stability, we find that $k = 50$ provides a robust default choice across most applications.

Bandwidth Parameter σ . The choice of the bandwidth parameter σ in the Gaussian kernel is related to the number of support points, n . A very small σ will cause the weights to concentrate on a single point, while a very large σ will lead to nearly uniform weights across all samples. Empirically, we recommend choosing σ such that the maximum weight is on the order of the average weight, i.e., $\max_i \{w_i\} \sim n^{-1/2}$. In practice, a default value of $\sigma = 10$ provides robust and effective performance across most of our experiments.

We analyze the robustness of our method to these parameters through a toy example detailed in Appendix D.1.

C.2 ALGORITHM OF IQDOT

For numerical inputs, given by discrete sample matrices $\mathbf{X} \in \mathbb{R}^{n \times d}$ and $\mathbf{Y} \in \mathbb{R}^{m \times q}$ with corresponding probability vectors $\mathbf{p}^X \in \Delta^{n-1}$ and $\mathbf{p}^Y \in \Delta^{m-1}$, we now detail the numerical implementation of the IQDOT, as introduced in Definition 4. The procedure is summarized in Algorithm 2.

Algorithm 2 Intergal-QDOT

Require: $\mathbf{X}, \mathbf{Y}, \mathbf{p}^X, \mathbf{p}^Y, q$

- 1: Initialize the data matrices \mathbf{X} and \mathbf{Y} .
 - 2: Compute the sample norms of \mathbf{X} and \mathbf{Y} : $\phi_0^X \leftarrow (\|\mathbf{x}_i\|)_{1 \leq i \leq n}$, $\phi_0^Y \leftarrow (\|\mathbf{y}_j\|)_{1 \leq j \leq m}$.
 - 3: Compute the q -quantiles of the sample norms: $\triangleright \mathcal{O}(n \log n + m \log m)$
 $\mathbf{r}^X \leftarrow (F_{\|\mathbf{X}\|}^{-1}(q_1), \dots, F_{\|\mathbf{X}\|}^{-1}(q_k))$, $\mathbf{r}^Y \leftarrow (F_{\|\mathbf{Y}\|}^{-1}(q_1), \dots, F_{\|\mathbf{Y}\|}^{-1}(q_k))$.
 - 4: **for** $i \leftarrow 1$ **to** k **do** $\triangleright \mathcal{O}(knd + kms)$
 - 5: Compute the quantile weights:

$$\mathbf{w}_{ij}^X \leftarrow \frac{p_j^X \exp\{-\sigma(d_j^X - \mathbf{r}_i^X)^2\}}{\sum_{j'=1}^n p_{j'}^X \exp\{-\sigma(d_{j'}^X - \mathbf{r}_i^X)^2\}}, \quad \mathbf{w}_{ij}^Y \leftarrow \frac{p_j^Y \exp\{-\sigma(d_j^Y - \mathbf{r}_i^Y)^2\}}{\sum_{j'=1}^m p_{j'}^Y \exp\{-\sigma(d_{j'}^Y - \mathbf{r}_i^Y)^2\}}$$
 - 6: Compute the q_i -quantile means: $\mathbf{m}_i^X \leftarrow \mathbf{X}^\top \mathbf{w}_i^X$, $\mathbf{m}_i^Y \leftarrow \mathbf{Y}^\top \mathbf{w}_i^Y$
 - 7: Compute the distances to quantile means:
 $\phi_i^X \leftarrow (\|\mathbf{x}_j - \mathbf{m}_i^X\|)_{1 \leq j \leq n}$, $\phi_i^Y \leftarrow (\|\mathbf{y}_j - \mathbf{m}_i^Y\|)_{1 \leq j \leq m}$
 - 8: Calculate the 1D- \mathcal{W}_p loss: $\mathcal{L}_i \leftarrow \mathcal{W}_p((\phi_i^X)_\# \mu_X, (\phi_i^Y)_\# \mu_Y)$ $\triangleright \mathcal{O}(n \log n + m \log m)$
 - 9: **end for**
 - 10: Set the final loss: $\mathcal{IQD}_p \leftarrow \frac{1}{k} \sum_{i=1}^k \mathcal{L}_i$
 - 11: **return** \mathcal{IQD}_p
-

Computational Cost. Similar to the analysis for QDOT, the pre-computation of QDMs and QDMs is highly efficient. The final one-dimensional Wasserstein distances can also be computed in $\mathcal{O}(n \log n)$ time by North-West corner rule (Peyré et al., 2019). Consequently, the overall time complexity of the IQDOT algorithm is $\mathcal{O}(n \log n)$, assuming k and the data dimensions are small constants.

D ADDITIONAL EXPERIMENT RESULTS

D.1 TOY EXAMPLE AND PARAMETER ANALYSIS

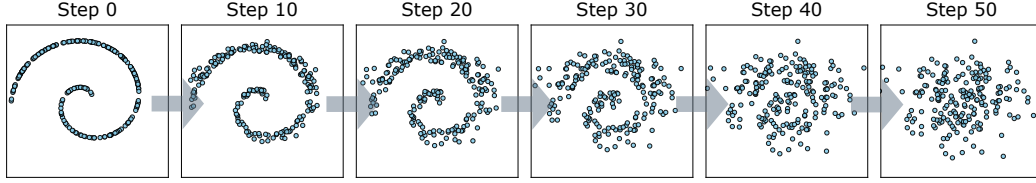


Figure 8: Illustration of the Wasserstein flow from a spiral point cloud to a random Gaussian noise distribution over 50 steps.

To provide an intuitive visualization of QDOT’s behavior and to analyze its sensitivity to key parameters, we present a toy example. As illustrated in Figure 8, the experiment consists of a Wasserstein flow sequence that interpolates between a 2D spiral distribution and a random Gaussian noise distribution. Each point cloud in the 50-step sequence comprises $n = 200$ points. We conduct a series of tests to: (1) compare the dissimilarity trends produced by different geometric metrics as the spiral deforms into noise; (2) analyze the effect of varying the bandwidth parameter σ ; (3) analyze the effect of varying the quantile interval parameter δ ; and (4) assess the impact of different initialization and scaling strategies.

The results, presented in Figure 9, reveal several key insights. First, when compared to baselines such as GW, EGW, SGW, and RISGW, both QDOT and IQDOT produce significantly smoother and more stable dissimilarity curves. Second, the analysis of the bandwidth parameter shows that for $\sigma = 1$ and $\sigma = 10$, the resulting curves are stable and smooth. A larger value of $\sigma = 100$ introduces some volatility, which can be attributed to the interaction between a high bandwidth and the relatively small sample size. Third, the dissimilarity curve is largely insensitive to the choice of the quantile interval parameter δ in this experiment, indicating good robustness. Finally, while different initialization and scaling strategies affect the absolute values of the dissimilarity, they all produce smooth, monotonically increasing curves that preserve the overall trend, demonstrating the robustness of the underlying geometric representation.

D.2 CROSS SPACE TASKS

Evaluation Metrics. For the numerical experiments, the inputs are a 3D point cloud sequence, represented by sample matrices $\mathbf{X} \in \mathbb{R}^{n \times 3}$, and a reference shape. The reference is provided both in its original 3D form, $\mathbf{Y} \in \mathbb{R}^{m \times 3}$, and as a 2D projection, $\hat{\mathbf{Y}} \in \mathbb{R}^{m \times 2}$. Both distributions have corresponding probability vectors, $\mathbf{p}^{\mathbf{X}}$ and $\mathbf{p}^{\mathbf{Y}}$, we assume access to a ground-truth alignment between the sequence frame \mathbf{X} and the reference shape \mathbf{Y} , denoted by a map \mathcal{T} , i.e., $\mathbf{X} = \mathcal{T}(\mathbf{Y})$. Algorithm 1 computes a coupling matrix $\Pi \in \mathbb{R}^{n \times m}$ between the input distributions, which satisfies the marginal constraints $\Pi \mathbf{1}_m = \mathbf{p}^{\mathbf{X}}$ and $\Pi^\top \mathbf{1}_n = \mathbf{p}^{\mathbf{Y}}$. To evaluate the quality of the resulting match, we define the following metrics.

The **Transformed Mean Squared Error (TMSE)** evaluates the alignment cost in the original 3D ambient space. Even when the coupling Π is computed between a 3D shape \mathbf{X} and a 2D projection $\hat{\mathbf{Y}}$, the error is measured by using the original 3D coordinates of the reference shape, \mathbf{Y} :

$$\text{TMSE} = \sum_{i=1}^n \sum_{j=1}^m \Pi_{ij} d_X(\mathbf{x}_i, \mathcal{T}(\mathbf{y}_j));$$

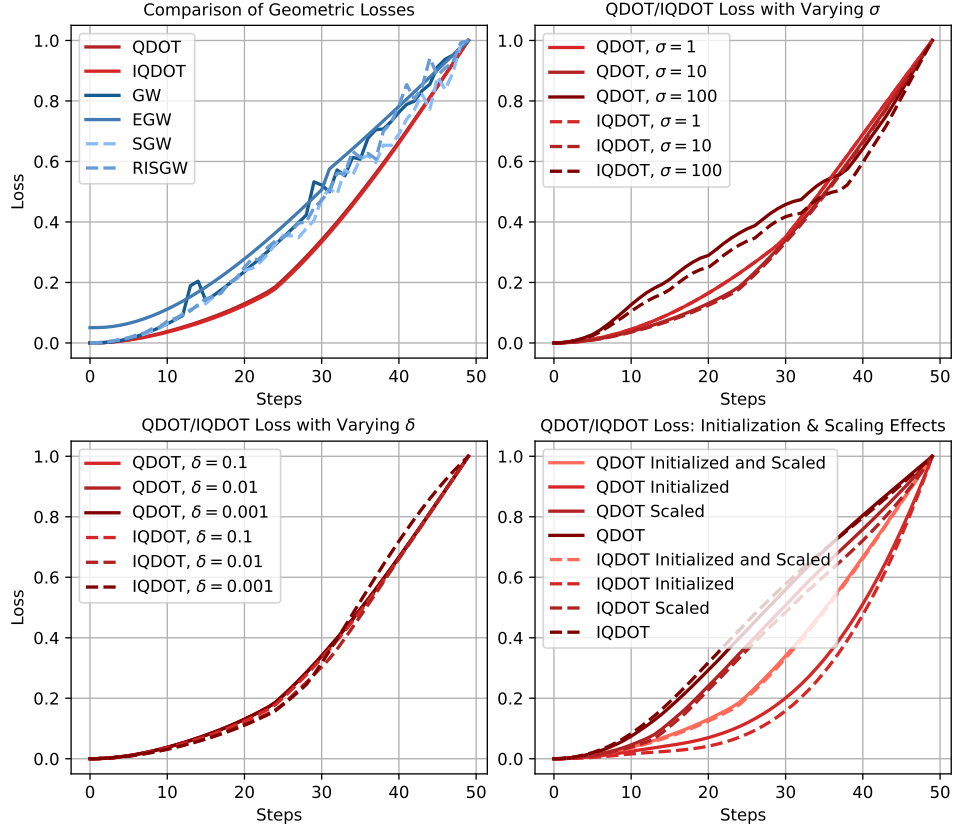


Figure 9: Results from the four experimental settings. The x-axis represents the step in the Wasserstein flow sequence, and the y-axis represents the computed dissimilarity (loss).

The **Inlier Ratio (IR)** then measures the total probability mass of the coupling Π that is placed on geometrically correct matches. A match between x_i and y_j is considered correct if x_i is close to the ground-truth location corresponding to y_j :

$$\text{IR} = \sum_{i=1}^n \sum_{j=1}^m \Pi_{ij} \mathbf{1}(d_X(x_i, \mathcal{T}(y_j)) < \tau) \times 100\%;$$

where the threshold τ is set to the 0.2-quantile of the intra-point distance distribution within \mathbf{X} , i.e., $\tau = F_{d_X}^{-1}(0.2)$.

Additional comparisons. To demonstrate the efficiency of QDOT, we compared it against several fast GW/EGW variants, such as PGW (Kerdoncuff et al., 2021), QGW (Chowdhury et al., 2021), and LRGW (Scetbon et al., 2022), as shown below. The penalty parameter for all methods was set to 0.01.

Table 5: Additional comparisons on the camel-gallop dataset

Methods	Transformed MSE ↓					Inlier Ratio (%) ↑					Time ($\times 10^2$ s)
	3D	2D _{1st}	2D _{2nd}	2D _{3rd}	Avg.	3D	2D _{1st}	2D _{2nd}	2D _{3rd}	Avg.	
PGW	0.35	0.34	0.32	0.33	0.33	53.85	55.38	55.71	51.35	54.07	0.42
QGW	0.56	0.62	0.53	0.64	0.59	23.42	18.30	27.98	17.74	21.85	0.19
LRGW	0.38	0.35	0.42	0.45	0.40	46.00	49.53	39.40	36.35	42.82	1.53
QDOT	0.22	0.24	0.23	0.25	0.24	71.15	63.75	68.48	63.20	66.65	0.74

As shown in Table 5, it is evident that the QDOT method achieves the best performance under the same order of computational cost, highlighting its efficiency.

Distortion. We substitute the coupling obtained via Algorithm 1 into the Gromov-Wasserstein loss function in 10. By plotting the loss curves over the optimization steps, we compare the behavior of the QDOT coupling with respect to both the QDOT loss and the GW loss. We employed equidistant sampling to select approximately $n \approx 1000$ points. The experimental results are presented below.

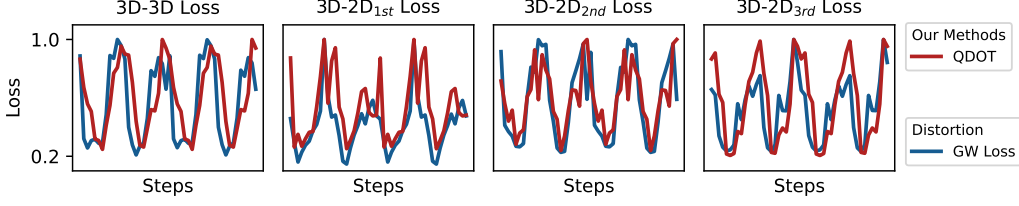


Figure 10: Distortion results on the camel-gallop dataset. The x-axis represents the time steps of the galloping camel sequence, and the y-axis indicates the loss corresponding to the coupling at each time step.

As shown in Figure 10, the QDOT loss and the GW loss maintain a high degree of consistency under the same coupling. This validates the effectiveness of QDOT as a metric.

Ablation study. To analyze the sensitivity of the results to the parameter settings, we consider $\sigma \in \{200, 1000, 5000\}$ and $k \in \{20, 100, 500\}$. The experimental results are presented below.

Table 6: QDOT results with different parameter selections on the camel-gallop dataset.

Parameters	Transformed MSE \downarrow					Inlier Ratio (%) \uparrow				
	3D	2D _{1st}	2D _{2nd}	2D _{3rd}	Avg.	3D	2D _{1st}	2D _{2nd}	2D _{3rd}	Avg.
$\sigma = 200, k = 20$	0.22	0.25	0.23	0.26	0.24	70.49	62.79	67.70	62.82	65.95
$\sigma = 200, k = 100$	0.22	0.24	0.23	0.25	0.24	70.82	63.34	68.05	63.09	66.33
$\sigma = 200, k = 500$	0.22	0.24	0.23	0.25	0.24	70.79	63.36	68.05	63.09	66.32
$\sigma = 1000, k = 20$	0.22	0.24	0.23	0.25	0.24	70.68	63.68	68.11	62.82	66.32
$\sigma = 1000, k = 100$	0.22	0.24	0.23	0.25	0.24	71.15	63.75	68.49	63.20	66.65
$\sigma = 1000, k = 500$	0.22	0.24	0.23	0.25	0.24	71.12	63.73	68.46	63.18	66.62
$\sigma = 5000, k = 20$	0.22	0.24	0.23	0.26	0.24	70.83	64.46	68.07	62.71	66.52
$\sigma = 5000, k = 100$	0.22	0.24	0.23	0.25	0.23	71.35	64.04	68.52	63.22	66.78
$\sigma = 5000, k = 500$	0.22	0.24	0.23	0.25	0.23	71.33	63.97	68.50	63.23	66.76
Summary $\bar{y}_{\text{mean} \pm \text{std}}$	0.22 (± 0.00)	0.24 (± 0.01)	0.23 (± 0.00)	0.25 (± 0.01)	0.24 (± 0.01)	70.95 (± 0.28)	63.68 (± 0.45)	68.22 (± 0.27)	63.04 (± 0.19)	66.47 (± 0.25)

As shown in Table 6, the QDOT alignment results are not particularly sensitive to the selection of σ and k . Generally, $k = 100$ yields sufficiently good performance, with larger k providing minimal additional gains.

D.3 TRANSFER LEARNING

Comparison with other sliced methods. Setting $n = 2048$, we compared the results against recently efficient sliced methods, including Sliced Wasserstein (Bonneel et al., 2015), Max-Sliced Wasserstein (Deshpande et al., 2019), and Hilbert Curve Projection (Li et al., 2024), as shown below.

As shown in Table 7, it is evident that IQDOT exhibits superior comprehensive performance. Notably, in the ‘guitar’ category, QDOT achieves near-perfect accuracy, whereas other methods fail almost entirely. This is attributed to the differing orientations of the guitars in the two datasets. As illustrated in Figure 11, guitars in ModelNet are horizontally aligned, while those in ShapeNet are vertically aligned, leading to the failure of other sliced-based methods. This highlights the critical importance of isometry invariance in transfer learning.

Ablation study. On the ablation study, we consider $n = 1024$, $\sigma \in \{50, 200, 1000\}$, and $k \in \{10, 50, 200\}$. The results are shown below.

As indicated in Table 8, apart from some numerical instability when $k = 10$, the differences between results are not substantial. This demonstrates that our method is relatively insensitive to parameter variations.

Table 7: Additional Transfer Learning Classification Accuracy Results (%)

Methods (Mo→Sh)	airplane	car	chair	guitar	lamp	laptop	table	Avg.	Time(h)
SW	96.69	95.77	94.81	0.00	68.71	97.78	89.60	86.03	17.36
MSW	96.25	95.55	95.53	0.00	67.29	98.23	89.91	86.09	17.05
HCP	93.38	95.55	95.85	0.25	74.73	98.23	91.29	86.90	22.58
IQDOT	95.35	87.86	85.01	96.44	52.55	98.00	87.15	85.42	5.98
Methods (Sh→Mo)	airplane	car	chair	guitar	lamp	laptop	table	Avg.	Time(h)
SW	100.00	95.96	96.76	0.00	93.06	95.86	97.76	89.36	17.36
MSW	99.86	95.62	97.67	0.00	93.75	96.45	96.54	89.45	17.05
HCP	98.48	96.63	97.47	0.00	92.36	97.63	98.78	89.52	22.58
IQDOT	98.89	91.91	93.22	98.03	77.77	97.63	97.56	95.05	5.98

Table 8: Point Cloud Transfer Learning Results with different QDOT parameter selections (%)

Params. (Mo→Sh)	airplane	car	chair	guitar	lamp	laptop	table	Avg.
$\sigma = 50, k = 10$	94.80	78.06	81.93	96.82	50.42	96.90	83.53	82.53
$\sigma = 50, k = 50$	95.46	82.74	82.70	96.44	50.81	96.67	84.50	83.46
$\sigma = 50, k = 200$	95.32	82.96	82.78	96.44	50.36	96.90	84.50	83.42
$\sigma = 200, k = 10$	94.94	80.07	83.10	97.46	50.10	96.45	84.03	83.12
$\sigma = 200, k = 50$	94.86	84.18	83.82	96.56	51.13	96.45	84.95	83.89
$\sigma = 200, k = 200$	95.65	84.41	84.11	96.32	50.23	97.12	85.30	84.15
$\sigma = 1000, k = 10$	94.98	80.73	82.01	97.71	48.48	95.57	82.68	82.26
$\sigma = 1000, k = 50$	95.35	84.08	84.78	95.93	50.81	97.12	84.90	84.14
$\sigma = 1000, k = 200$	95.72	85.63	84.73	96.19	50.55	96.90	85.18	84.37
Summary_{mean±std}	95.23	82.54	83.33	96.65	50.32	96.68	84.40	83.48
	(±0.33)	(±2.30)	(±1.02)	(±0.55)	(±0.72)	(±0.46)	(±0.81)	(±0.70)
Params. (Sh→Mo)	airplane	car	chair	guitar	lamp	laptop	table	Avg.
$\sigma = 50, k = 10$	98.62	83.84	89.79	97.65	79.17	95.27	94.92	92.58
$\sigma = 50, k = 50$	98.90	86.53	90.19	98.43	78.47	95.27	95.33	93.13
$\sigma = 50, k = 200$	98.76	88.89	90.09	98.82	79.86	95.86	95.73	93.49
$\sigma = 200, k = 10$	98.62	83.16	90.60	97.25	79.86	95.27	96.14	92.97
$\sigma = 200, k = 50$	98.89	87.20	91.10	98.43	81.94	97.04	96.74	93.97
$\sigma = 200, k = 200$	98.76	86.87	92.11	98.82	79.86	97.04	96.54	94.14
$\sigma = 1000, k = 10$	98.48	81.14	88.27	97.65	76.39	96.45	95.12	91.76
$\sigma = 1000, k = 50$	98.21	83.84	89.69	99.22	81.25	97.63	96.95	93.13
$\sigma = 1000, k = 200$	97.80	83.84	91.30	98.82	79.86	97.04	94.92	93.10
Summary_{mean±std}	98.56	85.03	90.35	98.34	79.63	96.32	95.82	93.14
	(±0.34)	(±2.31)	(±1.04)	(±0.63)	(±1.50)	(±0.87)	(±0.75)	(±0.67)

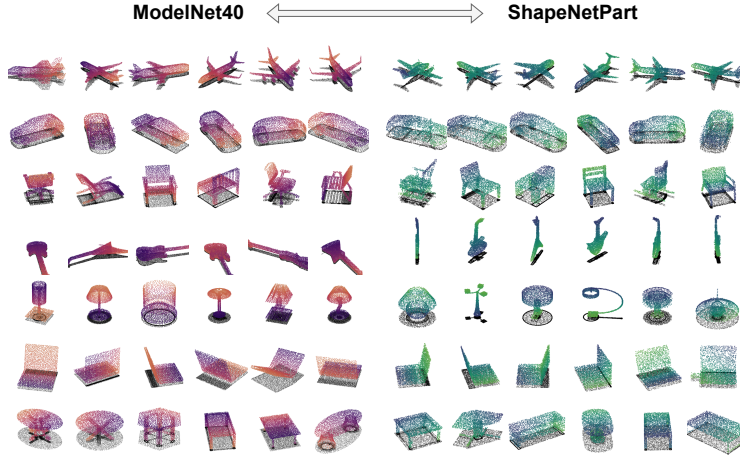


Figure 11: Visualization of the ModelNet40 and ShapeNetPart datasets.

D.4 HIGH DIMENSIONAL CAPABILITIES.

Following the experimental setup in Li et al. (2024), we sample source and target data from multivariate normal distributions $\mu_X \sim N(\mathbf{0}_d, \Sigma_X)$ and $\mu_Y \sim N(\mathbf{0}_d, \Sigma_Y)$, respectively, where the covariance matrices are defined as $\Sigma_X = \text{diag}(3\mathbf{I}_2, \mathbf{I}_{d-2})$ and $\Sigma_Y = \text{diag}(3\mathbf{I}_2 + 3\theta\mathbf{B}_2, \mathbf{I}_{d-2})$ and $\mathbf{I}_2, \mathbf{B}_2$ are identity and backward identity matrices. We evaluate the distance between these distributions as a function of the varying parameter θ . We fix the sample size at $n = 100$ and consider high-dimensional settings with $d \in \{50, 100, 200\}$. We compare our method against GW, EGW, SGW, and RISGW. The results are presented below:

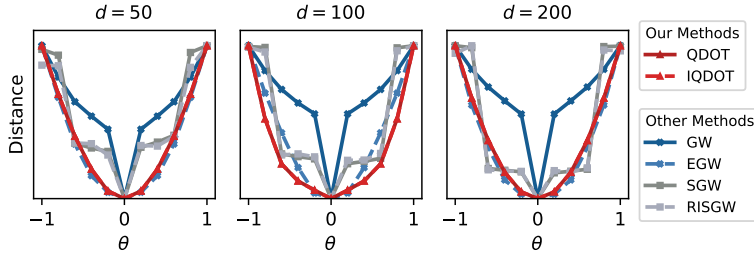


Figure 12: High dimension results.

As illustrated in Figure 12, QDOT/IQDOT and EGW exhibit remarkably stable and smooth trends across varying dimensions and θ values. In contrast, other methods display significant volatility. These results underscore the robustness of QDOT in high-dimensional settings.

D.5 MOLECULE GENERATION

Visualization results. The visualization results of molecular generation are presented in Figure 13. It can be observed that the IQDOT-regularized model tends to preserve stronger structural information and exhibits greater stability.

For the molecular generation experiment, we present the complete results in Table 9.

As shown in Table 9, simply reducing the weight of the MSE loss does not yield significant improvements for the baseline models. We also observe that the benefits of the IQDOT loss do not necessarily increase monotonically with its proportion in the final objective. Overall, these results suggest that the standard MSE loss primarily captures information about absolute atomic positions, while the IQDOT loss focuses on the intrinsic molecular structure. The strongest performance is

Table 9: Molecule Generation Results on QM9 Dataset.

Methods	Atom sta. (%)	Mol sta. (%)	Validity (%)	Uniqueness (%)	V * U (%)
Data	99.0	95.2	97.7	100.0	97.7
EDM	98.70 \pm 0.01	86.56 \pm 0.27	93.73 \pm 0.12	98.19 \pm 0.14	92.04 \pm 0.03
→MSE _{0.1}	98.81 \pm 0.04	87.63 \pm 0.21	94.54 \pm 0.15	98.17 \pm 0.08	92.81 \pm 0.20
→MSE _{0.2}	98.78 \pm 0.01	87.60 \pm 0.30	94.18 \pm 0.15	98.25 \pm 0.10	92.54 \pm 0.19
→MSE _{0.3}	98.68 \pm 0.07	86.04 \pm 0.74	93.83 \pm 0.43	98.47\pm0.22	92.39 \pm 0.55
→MSE _{0.4}	98.80 \pm 0.02	87.44 \pm 0.35	94.59 \pm 0.25	98.38 \pm 0.07	93.05\pm0.31
→IQDOT _{0.1}	99.15 \pm 0.01	90.91 \pm 0.21	96.75 \pm 0.33	92.06 \pm 0.15	89.06 \pm 0.19
→IQDOT _{0.2}	99.30 \pm 0.03	92.73 \pm 0.21	97.63 \pm 0.11	84.48 \pm 0.59	82.48 \pm 0.49
→IQDOT _{0.3}	99.35 \pm 0.06	93.29 \pm 0.81	97.80 \pm 0.22	83.05 \pm 2.27	81.21 \pm 2.03
→IQDOT _{0.4}	99.48\pm0.01	94.37\pm0.18	98.16\pm0.13	79.42 \pm 0.27	77.96 \pm 0.17
UniGEM	98.90 \pm 0.03	89.40 \pm 0.02	94.58 \pm 0.07	98.07 \pm 0.05	92.75 \pm 0.11
→MSE _{0.1}	98.91 \pm 0.04	89.00 \pm 0.22	95.00 \pm 0.16	98.04 \pm 0.07	93.13 \pm 0.10
→MSE _{0.2}	98.65 \pm 0.06	85.27 \pm 0.35	93.93 \pm 0.50	98.05 \pm 0.13	92.15 \pm 0.42
→MSE _{0.3}	99.00 \pm 0.07	89.48 \pm 0.70	95.16 \pm 0.16	98.09\pm0.23	93.34 \pm 0.18
→MSE _{0.4}	99.03 \pm 0.05	90.09 \pm 0.53	95.17 \pm 0.30	98.06 \pm 0.13	93.32 \pm 0.18
→IQDOT _{0.1}	99.24 \pm 0.03	92.73 \pm 0.09	96.85 \pm 0.11	96.45 \pm 0.21	93.42\pm0.23
→IQDOT _{0.2}	99.35 \pm 0.07	94.06 \pm 0.43	97.42 \pm 0.28	92.59 \pm 0.13	90.20 \pm 0.36
→IQDOT _{0.3}	99.44\pm0.01	95.23\pm0.18	97.94\pm0.06	85.64 \pm 0.28	83.88 \pm 0.30
→IQDOT _{0.4}	99.38 \pm 0.02	94.81 \pm 0.25	97.66 \pm 0.24	78.67 \pm 0.38	76.83 \pm 0.39

¹ The notation →IQDOT_{0.2} indicates that the objective function $\mathcal{L} = 0.2\mathcal{L}_{\text{IQDOT}} + 0.8\mathcal{L}_{\text{EDM}}$.

achieved when these two complementary objectives are fused, indicating that they provide a more comprehensive learning signal.

As illustrated in Figure 14, models trained with the IQDOT loss achieve stable convergence on both Atom and Molecule Stability metrics in just 1000 epochs, for both the EDM and UniGEM architectures. In contrast, the baseline models and those fine-tuned by simply adjusting the MSE loss weight require significantly more training: approximately 2500 epochs on EDM and over 2000 epochs on UniGEM to reach a similar level of stability. This ablation confirms that simply reducing the MSE loss weight does not accelerate convergence, highlighting the efficiency gains are a direct result of the structural guidance provided by the IQDOT loss.

THE USE OF LARGE LANGUAGE MODELS

We acknowledge the use of Large Language Models in the preparation of this manuscript. Their role was strictly limited to functioning as a tool for language polishing and improving readability. We affirm that all definitions and theorems presented in this paper are the original intellectual contributions of the authors. The core code is intricately linked to the algorithm we propose, and the experiments were designed specifically to investigate the properties of our method. We take full responsibility for the integrity of this work, which is free from any form of academic misconduct.

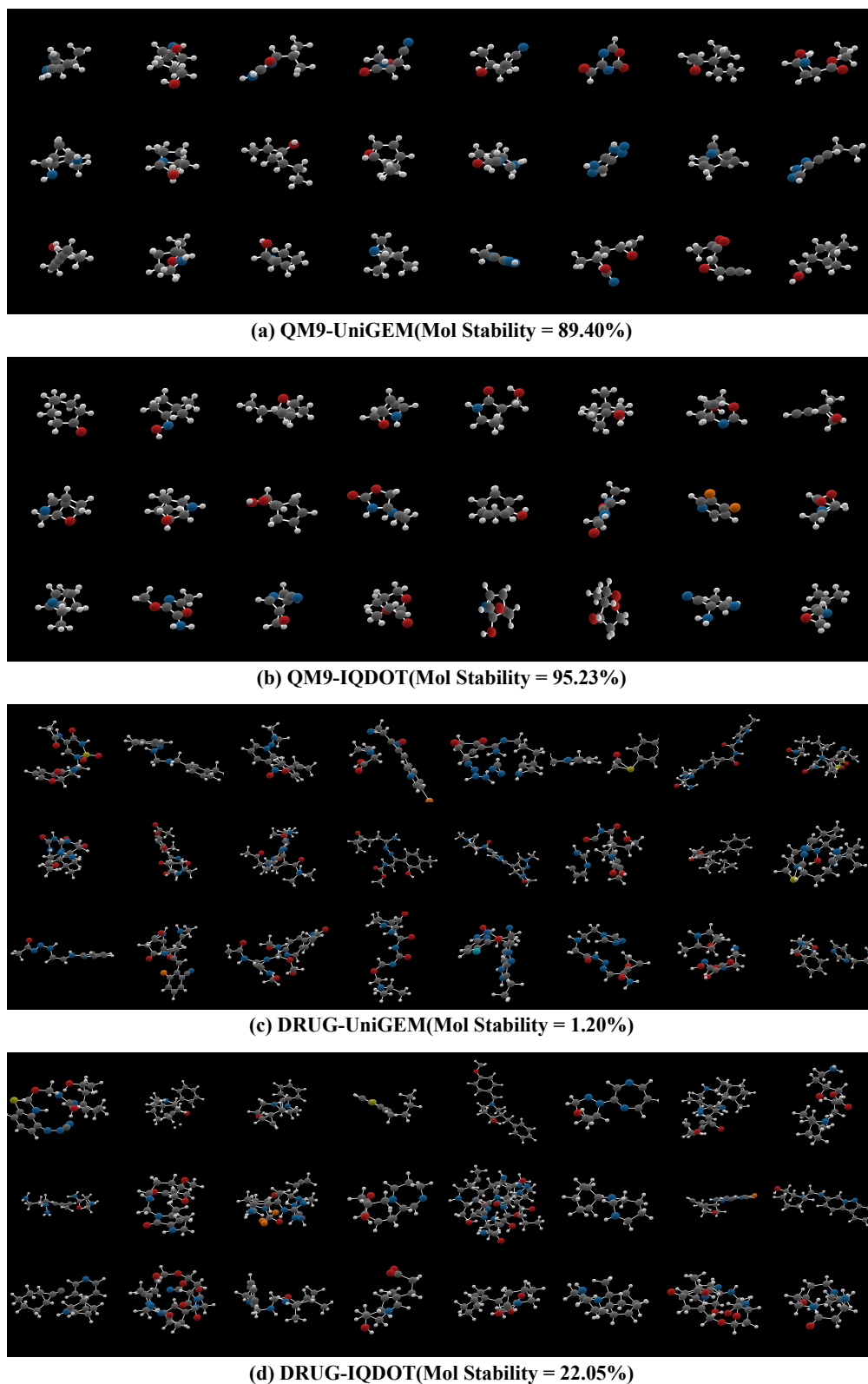
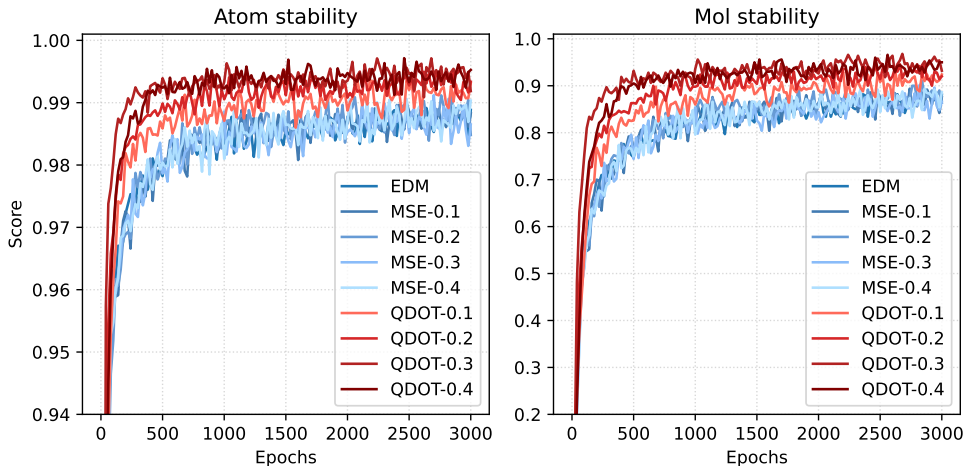
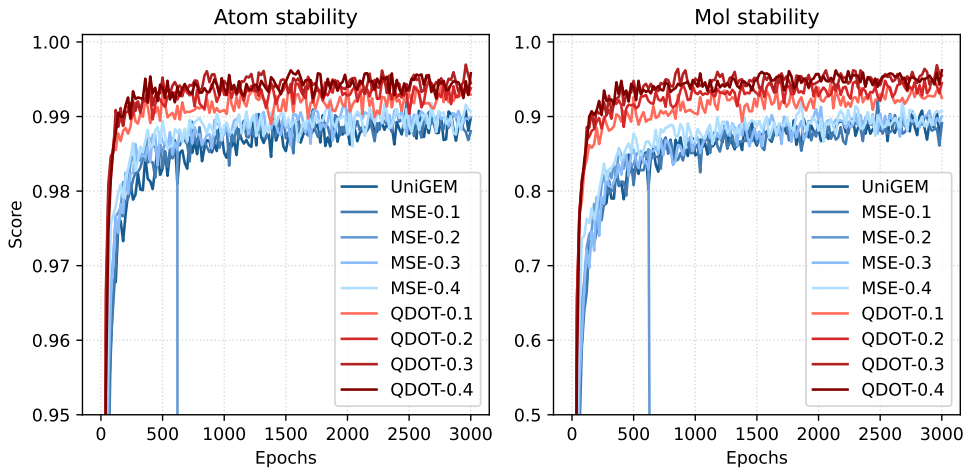


Figure 13: Visualization of molecular generation results comparing IQDOT and UniGEM on the QM9 and GEOM-Drugs datasets.



(a) Stability trends across epochs on the EDM model



(b) Stability trends across epochs on the UniGEM model

Figure 14: Convergence trends of Atom Stability and Molecule Stability for various methods on the EDM and UniGEM models. The horizontal axis represents the number of training epochs, and the vertical axis represents the stability percentage. The comparison includes the baseline models and variants with different weights for the MSE and IQDOT loss terms (MSE-0.1 to MSE-0.4 and IQDOT-0.1 to IQDOT-0.4).