

000 001 002 003 004 005 006 007 008 009 010 011 012 013 014 015 016 017 018 019 020 021 022 023 024 025 026 027 028 029 030 031 032 033 034 035 036 037 038 039 040 041 042 043 044 045 046 047 048 049 050 051 052 053 QDOT: AN EFFICIENT QUANTILE-WEIGHTED DISTANCE METRIC FOR GEOMETRIC COMPARISON VIA OPTIMAL TRANSPORT

006
007
008
009
010
011
012
013
014
015
016
017
018
019
020
021
022
023
024
025
026
027
028
029
030
031
032
033
034
035
036
037
038
039
040
041
042
043
044
045
046
047
048
049
050
051
052
053
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; Alatkar & 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 & Guibasm, 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 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:

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

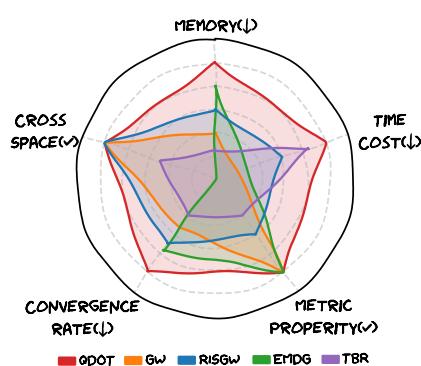


Figure 1: Comparison of various geometric metrics. Lower is better for memory, time cost, and convergence rate. We also evaluate metric properties and cross-space capabilities.

loss; or (3) they introduce non-geometric artifacts that can corrupt the comparison.

108 The primary challenge arises when comparing distributions defined on different metric spaces. To
 109 handle this scenario, the measurable metric space (Mémoli, 2011, mm-space) is defined as the triplet
 110 $\mathcal{X} = (X, d_X, \mu_X)$. For two distinct mm-spaces, $\mathcal{X} = (X, d_X, \mu_X)$ and $\mathcal{Y} = (Y, d_Y, \mu_Y)$, Sturm
 111 (2006) proposed the Gromov-Wasserstein distance, extending the concept from equation 1:

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

114 where the maps $f : X \rightarrow Z$ and $g : Y \rightarrow Z$ are isometric embeddings, satisfying
 115 $d_Z(f(x), f(x')) = d_X(x, x'), \forall x, x' \in \text{supp}(\mu_X)$ 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 "lossless"
 116 projection of the two disparate spaces X and Y into a common latent space Z , within which
 117 their Wasserstein distance can be computed. A significant contribution of the Gromov-Wasserstein
 118 distance is that it provides a complete metric on the space of mm-spaces. This metric is formally
 119 defined as follows.

121 **Definition 1 (Metric on Isometry Classes of mm-spaces)** For two mm-spaces $\mathcal{X} = (X, d_X, \mu_X)$
 122 and $\mathcal{Y} = (Y, d_Y, \mu_Y)$, a map $f : X \rightarrow Y$ is called an isometry if it is a surjection satisfying
 123 $d_Y(f(x), f(x')) = d_X(x, x'), \forall x, x' \in \text{supp}(\mu_X)$. Two mm-spaces are considered isometric if
 124 such an isometry exists between them. A function \mathcal{L} that measures the dissimilarity between two mm-
 125 spaces is a metric on the isometry classes of mm-spaces if it satisfies: (1) **Identity of Indiscernibles**:
 126 $\mathcal{L}(\mathcal{X}, \mathcal{Y}) \geq 0$, and $\mathcal{L}(\mathcal{X}, \mathcal{Y}) = 0$ iff \mathcal{X} and \mathcal{Y} are isometric; (2) **Symmetry**: $\mathcal{L}(\mathcal{X}, \mathcal{Y}) = \mathcal{L}(\mathcal{Y}, \mathcal{X})$;
 127 (3) **Triangle Inequality**: $\mathcal{L}(\mathcal{X}, \mathcal{Y}) \leq \mathcal{L}(\mathcal{X}, \mathcal{Z}) + \mathcal{L}(\mathcal{Y}, \mathcal{Z})$.

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

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

132 where $f : X \rightarrow Z$ and $g : Y \rightarrow Z$ are mapping functions into a common space Z . However,
 133 constructing such maps f and g that yield a valid metric satisfying Definition 1 imposes two key
 134 requirements. First, the maps must be **isometry-invariant**. That is, if mm-spaces \mathcal{X}_1 and \mathcal{X}_2 are
 135 isometric, their representations must be identically distributed, i.e., $f_{\# \mu_{X_1}} = f_{\# \mu_{X_2}}$. Second, to
 136 satisfy the identity of indiscernibles property, the representation must be **information-preserving**;
 137 it must uniquely encode the metric structure of the original space such that non-isometric spaces
 138 map to distinct distributions.

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

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

3 PROPOSED METHOD

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

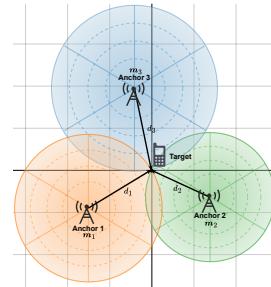


Figure 2: Illustration of trilateration

162 Note that, since it minimizes the aggregate squared distance to the distribution, the barycenter is
 163 isometry-invariant; consequently, the distribution of distances with respect to the Barycenter is also
 164 isometry-invariant. Specifically, if \mathcal{X} and \mathcal{Y} are isometric, then the corresponding distributions of
 165 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 Eu-
 166 clidean space, the Barycenter has an explicit solution corresponding to the mean of the distribution.
 167 Therefore, in the subsequent analysis, for two mm-spaces \mathcal{X} and \mathcal{Y} in Euclidean space, we assume
 168 without loss of generality that they are centered, and we utilize the norm as the distance to the
 169 Barycenter.

170

171 3.1 QUANTILE DISTANCE-WEIGHTED OPTIMAL TRANSPORT

172

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

176

177 **Definition 2 (Quantile-weighted Distance Mean (QDM))** Given a centered mm-space $\mathcal{X} =$
 178 (X, d_X, μ_X) , where $X \subseteq \mathbb{R}^p$ and the metric is the standard Euclidean distance, $d_X(\mathbf{x}_1, \mathbf{x}_2) =$
 179 $\|\mathbf{x}_1 - \mathbf{x}_2\|_2$, we can derive the distribution of its norms, denoted by $\mu_{\|X\|_2} = \|\cdot\|_{2\# \mu_X}$. Let $F_{\|X\|_2}$
 180 denote the cumulative distribution function (CDF) of $\mu_{\|X\|_2}$, and let $F_{\|X\|_2}^{-1}$ be its associated quan-
 181 ticle function. For any quantile level $q \in (0, 1)$, we define the Quantile Distance Weight function
 182 $w : X \times (0, 1) \rightarrow \mathbb{R}$ as

$$183 w^X(\mathbf{x}, q) := e^{-\sigma(\|\mathbf{x}\|_2 - F_{\|X\|_2}^{-1}(q))^2}, \quad (4)$$

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

186

$$187 \mathbf{m}^X(q) := \frac{\mathbb{E}_{\mu_X}[w^X(X, q)X]}{\mathbb{E}_{\mu_X}[w^X(X, q)]}. \quad (5)$$

188

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

194

$$195 \phi^X(\mathbf{x}, q) = d_X(\mathbf{x}, \mathbf{m}^X(q)). \quad (6)$$

196

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

199

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

200

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

204

205 **Definition 3 (Quantile-weighted Distance Optimal Transport (QDOT))** Let $\mathcal{X} = (X, d_X, \mu_X)$
 206 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
 207 and d_Y are the standard Euclidean distances. For a given vector of quantile levels $\mathbf{q} \in (0, 1)^k$,
 208 we compute their corresponding QDMD representations, ϕ^X and ϕ^Y . The QDOT distance is then
 209 defined as the Wasserstein distance between the resulting push-forward measures:

210

$$210 \mathcal{QD}_p(\mathcal{X}, \mathcal{Y}) = \mathcal{W}_p(\phi_{\# \mu_X}^X, \phi_{\# \mu_Y}^Y). \quad (7)$$

211

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

213

214 3.2 THEORETICAL RESULTS

215

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

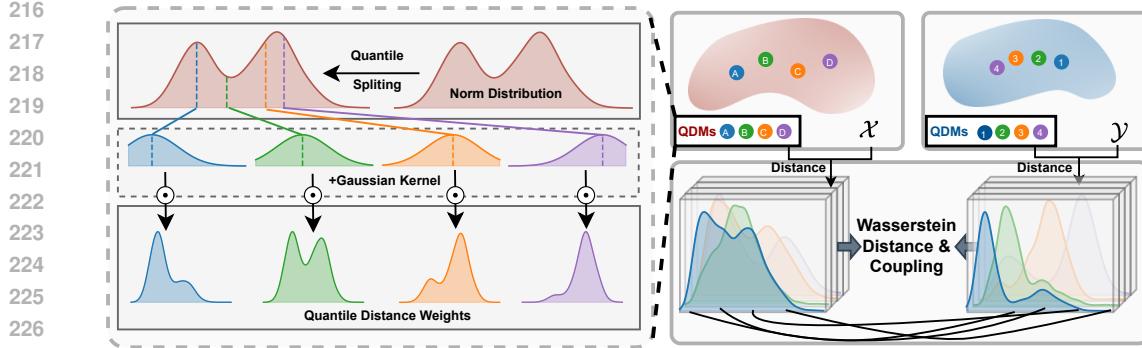


Figure 3: Illustration of the QDOT framework. The method first computes canonical QDM anchors for each distribution, then generates QDMD vector representations based on distances to these anchors. The final QDOT distance is the Wasserstein distance between these two resulting representation distributions.

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

270 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
 271 expectation convergence rate is given by $\mathbb{E}|\mathcal{QD}_p(\mathcal{X}_n, \mathcal{Y}_n) - \mathcal{QD}_p(\mathcal{X}, \mathcal{Y})| = \mathcal{O}(n^{-\frac{1}{\max(d,s)}})$.
 272

273

274

3.3 NUMERICAL IMPLEMENTATION

275 In the numerical setting, we consider discrete probability measures. Let Δ^{n-1} denote the $(n-1)$ -
 276 simplex. Given two sets of samples, $\mathcal{X}_n = \{\mathbf{x}_i\}_{1 \leq i \leq n}$ in \mathbb{R}^d and $\mathcal{Y}_m = \{\mathbf{y}_j\}_{1 \leq j \leq m}$ in \mathbb{R}^s ,
 277 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 $\mathbf{Y} =$
 278 $(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m)^\top \in \mathbb{R}^{m \times s}$. The corresponding probability vectors are denoted as $\mathbf{p}^{\mathbf{X}} \in \Delta^{n-1}$
 279 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 $\mu_{\mathbf{Y}} =$
 280 $\sum_{j=1}^m p_j^{\mathbf{Y}} \delta_{\mathbf{y}_j}$, respectively. For a given quantile level vector $\mathbf{q} \in (0, 1)^k$, the procedure for computing
 281 the QDOT distance is detailed in Algorithm 1.
 282

283

284 **Algorithm 1** QDOT

285 **Require:** $\mathbf{X}, \mathbf{Y}, \mathbf{p}^{\mathbf{X}}, \mathbf{p}^{\mathbf{Y}}, \mathbf{q}$ 286 1: Initialize the data matrices \mathbf{X} and \mathbf{Y} .287 2: Compute the sample norms of \mathbf{X} and \mathbf{Y} : $\phi_0^{\mathbf{X}} \leftarrow (\|\mathbf{x}_i\|_2)_{1 \leq i \leq n}$, $\phi_0^{\mathbf{Y}} \leftarrow (\|\mathbf{y}_j\|_2)_{1 \leq j \leq m}$.288 3: Compute the \mathbf{q} -quantiles of the sample norms: $\triangleright \mathcal{O}(n \log n + m \log m)$ 289 $\mathbf{r}^{\mathbf{X}} \leftarrow (F_{\|\mathbf{X}\|_2}^{-1}(q_1), \dots, F_{\|\mathbf{X}\|_2}^{-1}(q_k)), \quad \mathbf{r}^{\mathbf{Y}} \leftarrow (F_{\|\mathbf{Y}\|_2}^{-1}(q_1), \dots, F_{\|\mathbf{Y}\|_2}^{-1}(q_k))$.290 4: **for** $i \leftarrow 1$ to k **do** $\triangleright \mathcal{O}(knd + kms)$

291 5: Compute the quantile weights:

292
$$\mathbf{w}_{ij}^{\mathbf{X}} \leftarrow \frac{p_j^{\mathbf{X}} \exp\{-\sigma(\mathbf{d}_j^{\mathbf{X}} - \mathbf{r}_i^{\mathbf{X}})^2\}}{\sum_{j'=1}^n p_{j'}^{\mathbf{X}} \exp\{-\sigma(\mathbf{d}_{j'}^{\mathbf{X}} - \mathbf{r}_i^{\mathbf{X}})^2\}}, \quad \mathbf{w}_{ij}^{\mathbf{Y}} \leftarrow \frac{p_j^{\mathbf{Y}} \exp\{-\sigma(\mathbf{d}_j^{\mathbf{Y}} - \mathbf{r}_i^{\mathbf{Y}})^2\}}{\sum_{j'=1}^m p_{j'}^{\mathbf{Y}} \exp\{-\sigma(\mathbf{d}_{j'}^{\mathbf{Y}} - \mathbf{r}_i^{\mathbf{Y}})^2\}}$$

293 6: Compute the q_i -quantile means: $\mathbf{m}_i^{\mathbf{X}} \leftarrow \mathbf{X}^\top \mathbf{w}_i^{\mathbf{X}}$, $\mathbf{m}_i^{\mathbf{Y}} \leftarrow \mathbf{Y}^\top \mathbf{w}_i^{\mathbf{Y}}$

294 7: Compute the distances to quantile means:

295
$$\phi_i^{\mathbf{X}} \leftarrow (\|\mathbf{x}_j - \mathbf{m}_i^{\mathbf{X}}\|_2)_{1 \leq j \leq n}, \quad \phi_i^{\mathbf{Y}} \leftarrow (\|\mathbf{y}_j - \mathbf{m}_i^{\mathbf{Y}}\|_2)_{1 \leq j \leq m}$$

296 8: **end for**297 9: Concatenate representations: $\Phi^{\mathbf{X}} \leftarrow [\phi_0^{\mathbf{X}}, \phi_1^{\mathbf{X}}, \dots, \phi_k^{\mathbf{X}}]$, $\Phi^{\mathbf{Y}} \leftarrow [\phi_0^{\mathbf{Y}}, \phi_1^{\mathbf{Y}}, \dots, \phi_k^{\mathbf{Y}}]$ 298 10: \mathcal{W}_p Computation: $\mathcal{QD}_p, \Pi_{\mathcal{QD}_p} \leftarrow \mathcal{W}_p((\Phi^{\mathbf{X}})_{\#} \mu_{\mathbf{X}}, (\Phi^{\mathbf{Y}})_{\#} \mu_{\mathbf{Y}}) \triangleright \mathcal{O}(n^2 \log n + m^2 \log m)$ 299 11: **return** $\mathcal{QD}_p, \Pi_{\mathcal{QD}_p}$

300

301

302

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

309

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

311

312

3.4 INTEGRAL-QDOT

313

314 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
 315 Optimal Transport have become increasingly popular. We observe that the QDMD representation,
 316 $\phi^X(x, q)$, defined in the previous section, is a scalar value for any given quantile level q . This
 317 structure naturally inspires an alternative approach: instead of comparing the multi-dimensional
 318 representations in \mathbb{R}^{k+1} , we can compare the one-dimensional distributions of the QDMD scalars
 319 for each q and then aggregate the results. This leads to the Integral-QDOT(IQDOT) approach.
 320

321

322 **Definition 4 (Integral-QDOT)** Given two mm-spaces $\mathcal{X} = (X, d_X, \mu_X)$ and $\mathcal{Y} = (Y, d_Y, \mu_Y)$,
 323 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
 324 $q \in (0, 1)$, we can obtain their corresponding scalar QDMD representations, $(\phi^X(\cdot, q))_{\#} \mu_X$ and
 325 $(\phi^Y(\cdot, q))_{\#} \mu_Y$. The Integral-QDOT distance is then defined as the L_p -norm of the 1-D Wasserstein

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

$$326 \quad \mathcal{IQD}_p(\mathcal{X}, \mathcal{Y}) := \inf_{\pi \in \Pi(\mu_X, \mu_Y)} \left(\int_{(0,1)} \int_{X \times Y} |\phi^X(x, q) - \phi^Y(y, q)|^p d\pi(x, y) dq \right)^{1/p}. \quad (8)$$

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

330 **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-
331 spaces embedded in Euclidean spaces. If the set of QDMs satisfies the dimensionality condition
332 $\dim(\{\mathbf{m}^X(q)\}_{q \in (0,1)}) = \dim(X)$ and $\dim(\{\mathbf{m}^Y(q)\}_{q \in (0,1)}) = \dim(Y)$, then \mathcal{IQD}_p defines a
333 metric on the isometry classes of these mm-spaces.*

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

340 4 EXPERIMENTS

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

349 4.1 CROSS SPACE TASKS

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

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

365 Methods	366 <i>Transformed MSE</i> ↓					367 <i>Inlier Ratio (%)</i> ↑					368 <i>Time</i> ($\times 10^2$ s)
	3D	2D _{1st}	2D _{2nd}	2D _{3rd}	Avg.	3D	2D _{1st}	2D _{2nd}	2D _{3rd}	Avg.	
EGW _(λ=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 _(λ=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

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

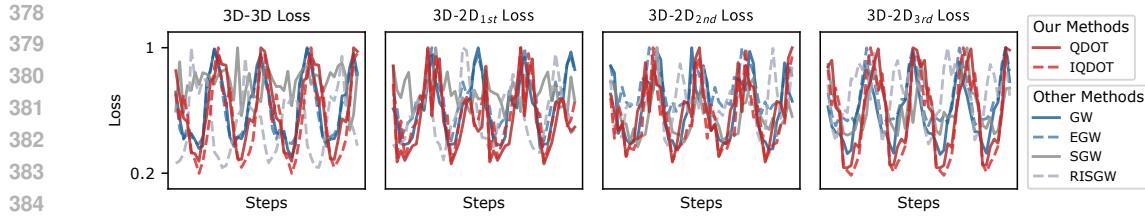


Figure 4: Loss trends between the camel-gallop sequence and the four reference in different spaces. It is evident that QDOT, IQDOT, GW, and EGW successfully capture the four periodic cycles of the camel’s gallop, exhibiting consistent fluctuation amplitudes and trends across all reference spaces. In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

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.

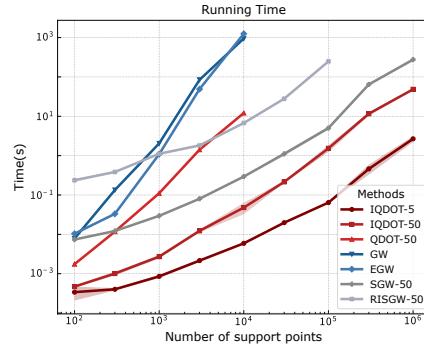


Figure 5: Runtime comparison of various methods. All experiments were conducted in a CPU environment.

432
433
Table 2: Point Cloud Transfer Learning Classification Accuracy (%)

434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 Methods (Mo→Sh)	airplane	car	chair	guitar	lamp	laptop	table	Avg.	Time(h)
SGW-1024	96.69	99.10	60.67	0.00	71.29	98.66	94.25	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	83.82	96.56	51.13	96.45	84.95	83.89	1.59
IQDOT-2048	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)
SGW-1024	89.66	37.03	43.68	0.00	80.55	20.11	99.59	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	87.20	91.10	98.43	81.94	97.04	96.74	93.97	1.59
IQDOT-2048	98.89	91.91	93.22	98.03	77.77	97.63	97.56	95.05	5.98

445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
4.4 MOLECULE GENERATION

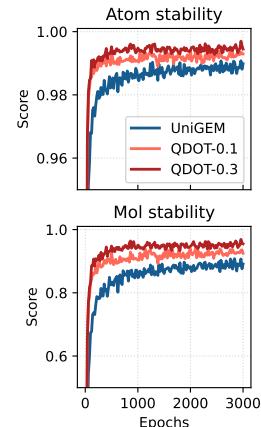
Molecular generation is a central challenge in drug discovery and molecular science. Recent advances show that diffusion models with equivariant neural architectures have become the dominant paradigm (Hoogeboom et al., 2022; Xu et al., 2023; Song et al., 2024; Feng et al., 2025). Since molecular properties 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 (Hoogeboom 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.

460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
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 ± 0.01	86.56 ± 0.27	93.73 ± 0.12	92.04 ± 0.03
→MSE _{0.1}	98.81 ± 0.04	87.63 ± 0.21	94.54 ± 0.15	92.81 ± 0.20
→MSE _{0.3}	98.68 ± 0.07	86.04 ± 0.74	93.83 ± 0.43	92.39 ± 0.55
→ IQDOT _{0.1}	99.15 ± 0.01	90.91 ± 0.21	96.75 ± 0.33	89.06 ± 0.19
→ IQDOT _{0.3}	99.35 ± 0.06	93.29 ± 0.81	97.80 ± 0.22	81.21 ± 0.03
UniGEM	98.90 ± 0.03	89.40 ± 0.02	94.58 ± 0.07	92.75 ± 0.11
→MSE _{0.1}	98.91 ± 0.04	89.00 ± 0.22	95.00 ± 0.16	93.13 ± 0.10
→MSE _{0.3}	99.00 ± 0.07	89.48 ± 0.70	95.16 ± 0.16	93.34 ± 0.18
→ IQDOT _{0.1}	99.24 ± 0.03	92.73 ± 0.09	96.85 ± 0.11	93.42 ± 0.23
→ IQDOT _{0.3}	99.44 ± 0.01	95.23 ± 0.18	97.94 ± 0.06	83.88 ± 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 $\mathcal{L} \leftarrow 0.9\mathcal{L}_{\text{MSE}}$.

477
478
479
480
481
482
483
484
485
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.

486 To assess the fine-tuning capability of the **IQ-**
 487 **DOT** loss, we conducted an additional exper-
 488 iment on GEOM-Drug (Axelrod & Gomez-
 489 Bombarelli, 2022). Starting from a UniGEM
 490 model pre-trained for 13 epochs (Feng et al.,
 491 2025), we introduced **IQDOT** into the loss and
 492 fine-tuned for 3 more epochs with a learning
 493 rate of 10^{-4} . The results in Table 4 show
 494 that while continued training with the original
 495 MSE loss yields no further gains, incorporat-
 496 ing **IQDOT** leads to substantial improvements
 497 in Atom Stability and Molecule Stability, with only a negligible drop in validity.
 498

5 CONCLUSION

501 This work introduces QDOT, a novel geometric metric. By constructing isometry-invariant an-
 502 chors from distance quantiles and deriving a lossless distance representation through trilateration
 503 theory, QDOT establishes a rigorous metric on isometry classes of mm-spaces. Experimental re-
 504 sults demonstrate its effectiveness across cross-space alignment and comparison, computational ef-
 505 ficiency, transfer learning, and molecular generation tasks. Future directions include: (1) extending
 506 QDOT to hyperbolic and spherical spaces; (2) developing more general approaches for constructing
 507 isometry-invariant anchors and representations; and (3) applying QDOT-based alignment techniques
 508 and QDOT as a loss function in complex models.
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539

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	98.46
→ IQDOT _{0.2}	91.87	<u>17.14</u>	97.86
→ IQDOT _{0.3}	91.73	22.05	94.62

540 ETHICS STATEMENT
541542 All datasets utilized in this study are publicly available, and their use raises no ethical concerns.
543544 REPRODUCIBILITY STATEMENT
545546 We are committed to ensuring the reproducibility of our research. The source code for all ex-
547 periments is provided in the supplementary materials, which can be used to reproduce the results
548 presented in this paper. For the molecule generation experiment, we have included the evaluate
549 checkpoint of QDOT-0.3. This checkpoint can also be reproduced using the provided training code.
550 Furthermore, we have explicitly stated all key assumptions for the theorems presented. The sources
551 and preprocessing scripts for all datasets used in this work are also provided.
552553 REFERENCES
554555 Sayali Anil Alatkar and Daifeng Wang. CMOT: cross-modality optimal transport for multimodal
556 inference. *Genome Biology*, 24(1):163, 2023.557 Jason Altschuler, Francis Bach, Alessandro Rudi, and Jonathan Niles-Weed. Massively scalable
558 sinkhorn distances via the nyström method. *Advances in neural information processing systems*,
559 32, 2019.560 Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks.
561 In *International conference on machine learning*, pp. 214–223. PMLR, 2017.563 Arip Asadulaev, Rostislav Korst, Aleksandr Korotin, Vage Egiazarian, Andrey Filchenkov, and
564 Evgeny Burnaev. Rethinking optimal transport in offline reinforcement learning. *Advances in
565 Neural Information Processing Systems*, 37:123592–123607, 2024.566 Simon Axelrod and Rafael Gomez-Bombarelli. Geom, energy-annotated molecular conformations
567 for property prediction and molecular generation. *Scientific Data*, 9(1):185, 2022.568 Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using
569 shape contexts. *IEEE transactions on pattern analysis and machine intelligence*, 24(4):509–522,
570 2002.572 Nicolas Bonneel, Michiel Van De Panne, Sylvain Paris, and Wolfgang Heidrich. Displacement
573 interpolation using lagrangian mass transport. In *Proceedings of the 2011 SIGGRAPH Asia con-
574 ference*, pp. 1–12, 2011.576 Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and Radon Wasserstein
577 barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51:22–45, 2015.578 Chao Chen, Guanbin Li, Ruijia Xu, Tianshui Chen, Meng Wang, and Liang Lin. Clusternet: Deep
579 hierarchical cluster network with rigorously rotation-invariant representation for point cloud anal-
580 ysis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*,
581 pp. 4994–5002, 2019.583 Samir Chowdhury, David Miller, and Tom Needham. Quantized gromov-wasserstein. In *Joint
584 European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 811–
585 827. Springer, 2021.586 Scott Cohen and L Guibas. The earth mover’s distance under transformation sets. In *Proceedings
587 of the Seventh IEEE International Conference on Computer Vision*, volume 2, pp. 1076–1083.
588 IEEE, 1999.589 Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural
590 information processing systems*, 26, 2013.592 Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger
593 bridge with applications to score-based generative modeling. *Advances in neural information
594 processing systems*, 34:17695–17709, 2021.

594 Ishan Deshpande, Yuan-Ting Hu, Ruoyu Sun, Ayis Pyrros, Nasir Siddiqui, Sanmi Koyejo, Zhizhen
 595 Zhao, David Forsyth, and Alexander G Schwing. Max-sliced Wasserstein distance and its use for
 596 GANs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*,
 597 pp. 10648–10656, 2019.

598 Haoqiang Fan, Hao Su, and Leonidas J Guibas. A point set generation network for 3d object recon-
 599 struction from a single image. In *Proceedings of the IEEE conference on computer vision and*
 600 *pattern recognition*, pp. 605–613, 2017.

602 Shikun Feng, Yuyan Ni, Lu yan, Zhi-Ming Ma, Wei-Ying Ma, and Yanyan Lan. UniGEM: A unified
 603 approach to generation and property prediction for molecules. In *The Thirteenth International*
 604 *Conference on Learning Representations*, 2025. URL [https://openreview.net/forum?](https://openreview.net/forum?id=Lb91pXwZMR)
 605 `id=Lb91pXwZMR`.

606 Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z Alaya, Aurélie Boisbunon, Stanis-
 607 las Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, et al. Pot: Python
 608 optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021.

609 Nicolas Fournier and Arnaud Guillin. On the rate of convergence in wasserstein distance of the
 610 empirical measure. *Probability theory and related fields*, 162(3):707–738, 2015.

612 Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. SE (3)-transformers: 3d roto-
 613 translation equivariant attention networks. *Advances in neural information processing systems*,
 614 33:1970–1981, 2020.

616 Michael Gromov. Groups of polynomial growth and expanding maps (with an appendix by Jacques
 617 Tits). *Publications Mathématiques de l'IHÉS*, 53:53–78, 1981.

618 Majdi Hassan, Nikhil Shenoy, Jungyoon Lee, Hannes Stärk, Stephan Thaler, and Dominique Beaini.
 619 Et-flow: Equivariant flow-matching for molecular conformer generation. *Advances in Neural*
 620 *Information Processing Systems*, 37:128798–128824, 2024.

622 Emiel Hoogeboom, Victor Garcia Satorras, Clement Vignac, and Max Welling. Equivariant diffu-
 623 sion for molecule generation in 3d. In *International conference on machine learning*, pp. 8867–
 624 8887. PMLR, 2022.

625 Tanguy Kerdoncuff, Remi Emonet, and Marc Sebban. Sampled gromov wasserstein. *Machine*
 626 *Learning*, 110(8):2151–2186, 2021.

628 Seohyun Kim, Jaeyoo Park, and Bohyung Han. Rotation-invariant local-to-global representation
 629 learning for 3d point cloud. *Advances in Neural Information Processing Systems*, 33:8174–8185,
 630 2020.

631 Pascal Klink, Haoyi Yang, Carlo D'Eramo, Jan Peters, and Joni Pajarinen. Curriculum reinforce-
 632 ment learning via constrained optimal transport. In *International Conference on Machine Learn-
 633 ing*, pp. 11341–11358. PMLR, 2022.

635 Matt Kusner, Yu Sun, Nicholas Kolkin, and Kilian Weinberger. From word embeddings to document
 636 distances. In *International conference on machine learning*, pp. 957–966. PMLR, 2015.

637 Tam Le, Makoto Yamada, Kenji Fukumizu, and Marco Cuturi. Tree-sliced variants of wasserstein
 638 distances. *Advances in neural information processing systems*, 32, 2019.

640 Mengyu Li, Jun Yu, Tao Li, and Cheng Meng. Importance sparsification for sinkhorn algorithm.
 641 *Journal of Machine Learning Research*, 24(247):1–44, 2023a.

642 Mengyu Li, Jun Yu, Hongteng Xu, and Cheng Meng. Efficient approximation of Gromov-
 643 Wasserstein distance using importance sparsification. *Journal of Computational and Graphical*
 644 *Statistics*, 32(4):1512–1523, 2023b.

646 Tao Li, Cheng Meng, Hongteng Xu, and Jun Yu. Hilbert curve projection distance for distribution
 647 comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(7):4993–5007,
 2024.

648 Tao Li, Cheng Meng, Hongteng Xu, and Jun Zhu. Efficient variants of wasserstein distance in
 649 hyperbolic space via space-filling curve projection. *IEEE Transactions on Neural Networks and*
 650 *Learning Systems*, 2025.

651

652 Igor Melnyk, Youssef Mroueh, Brian Belgodere, Mattia Rigotti, Apoorva Nitsure, Mikhail
 653 Yurochkin, Kristjan Greenewald, Jiri Navratil, and Jarret Ross. Distributional preference align-
 654 ment of llms via optimal transport. *Advances in Neural Information Processing Systems*, 37:
 655 104412–104442, 2024.

656

657 Facundo Mémoli. Gromov–Wasserstein distances and the metric approach to object matching.
Foundations of computational mathematics, 11:417–487, 2011.

658

659 Gabriel Peyré, Marco Cuturi, and Justin Solomon. Gromov-Wasserstein averaging of kernel and
 660 distance matrices. In *International conference on machine learning*, pp. 2664–2672. PMLR,
 661 2016.

662

663 Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data
 664 science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

665

666 Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Quantum
 667 chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):1–7, 2014.

668

669 Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover’s distance as a metric for
 670 image retrieval. *International journal of computer vision*, 40(2):99–121, 2000.

671

672 Meyer Scetbon, Gabriel Peyré, and Marco Cuturi. Linear-time Gromov Wasserstein distances using
 673 low rank couplings and costs. In *International Conference on Machine Learning*, pp. 19347–
 674 19365. PMLR, 2022.

675

676 RJ Serfling. Approximation theorems of mathematical statistics. *Wiley series in probability and*
 677 *mathematical statistics Show all parts in this series*, 1980.

678

679 Liangliang Shi, Jack Fan, and Junchi Yan. Ot-clip: Understanding and generalizing clip via optimal
 680 transport. In *Forty-first International Conference on Machine Learning*, 2024.

681

682 Yuxuan Song, Jingjing Gong, Minkai Xu, Ziyao Cao, Yanyan Lan, Stefano Ermon, Hao Zhou,
 683 and Wei-Ying Ma. Equivariant flow matching with hybrid probability transport for 3d molecule
 684 generation. *Advances in Neural Information Processing Systems*, 36:549–568, 2023.

685

686 Yuxuan Song, Jingjing Gong, Hao Zhou, Mingyue Zheng, Jingjing Liu, and Wei-Ying Ma. Unified
 687 generative modeling of 3d molecules with bayesian flow networks. In *The Twelfth International*
 688 *Conference on Learning Representations*, 2024.

689

690 Karl-Theodor Sturm. On the geometry of metric measure spaces. *Acta Mathematica*, 196(1):65–
 691 131, 2006.

692

693 Karl-Theodor Sturm. *The space of spaces: curvature bounds and gradient flows on the space of*
 694 *metric measure spaces*, volume 290. American Mathematical Society, 2023.

695

696 Robert W Sumner and Jovan Popović. Deformation transfer for triangle meshes. *ACM Transactions*
 697 *on graphics (TOG)*, 23(3):399–405, 2004.

698

699 Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. A concise and provably informative multi-scale
 700 signature based on heat diffusion. *Computer Graphics Forum*, 28(5):1383–1392, 2019.

701

702 Federico Thomas and Lluis Ros. Revisiting trilateration for robot localization. *IEEE Transactions*
 703 *on robotics*, 21(1):93–101, 2005.

704

705 Vayer Titouan, Rémi Flamary, Nicolas Courty, Romain Tavenard, and Laetitia Chapel. Sliced
 706 Gromov-Wasserstein. *Advances in Neural Information Processing Systems*, 32, 2019.

707

708 Alexander Tong, Kilian FATRAS, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-
 709 Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative mod-
 710 els with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. ISSN
 711 2835-8856. URL <https://openreview.net/forum?id=CD9Snc73AW>. Expert Certifi-
 712 cation.

702 Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2008.
703

704 Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaou Tang, and Jianxiong
705 Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE*
706 *conference on computer vision and pattern recognition*, pp. 1912–1920, 2015.

707 Hongteng Xu, Dixin Luo, Hongyuan Zha, and Lawrence Carin Duke. Gromov-Wasserstein learning
708 for graph matching and node embedding. In *International conference on machine learning*, pp.
709 6932–6941. PMLR, 2019.

710

711 Minkai Xu, Alexander S Powers, Ron O Dror, Stefano Ermon, and Jure Leskovec. Geometric latent
712 diffusion models for 3d molecule generation. In *International Conference on Machine Learning*,
713 pp. 38592–38610. PMLR, 2023.

714 Yingxue Xu and Hao Chen. Multimodal optimal transport-based co-attention transformer with
715 global structure consistency for survival prediction. In *Proceedings of the IEEE/CVF interna-*
716 *tional conference on computer vision*, pp. 21241–21251, 2023.

717 Jianyu Yang, Hongxing Wang, Junsong Yuan, Youfu Li, and Jianyang Liu. Invariant multi-scale
718 descriptor for shape representation, matching and retrieval. *Computer Vision and Image Under-*
719 *standing*, 145:43–58, 2016.

720

721 Hao Yu, Zheng Qin, Ji Hou, Mahdi Saleh, Dongsheng Li, Benjamin Busam, and Slobodan Ilic.
722 Rotation-invariant transformer for point cloud matching. In *Proceedings of the IEEE/CVF con-*
723 *ference on computer vision and pattern recognition*, pp. 5384–5393, 2023.

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756 A ADDITIONAL BACKGROUND
757758 Here we introduce additional background concepts that are related to our work.
759760 **One-Dimensional Wasserstein Distance.** A key special case of the Wasserstein distance defined
761 in equation 1 occurs in one dimension, where it admits a convenient closed-form solution. For two
762 one-dimensional distributions μ_X and μ_Y , let F_X and F_Y be their respective cumulative distribution
763 functions, and let F_X^{-1} and F_Y^{-1} be the corresponding quantile functions. The p -Wasserstein
764 distance is then given by:
765

766
$$\mathcal{W}_p(\mu_X, \mu_Y) = \left(\int_0^1 |F_X^{-1}(u) - F_Y^{-1}(u)|^p du \right)^{\frac{1}{p}}. \quad (9)$$

767
768

769 For discrete distributions, the empirical quantile functions are readily obtained by sorting, making
770 the computation of the 1D Wasserstein distance explicit and efficient.
771772 **Gromov-Wasserstein Distance(Mémoli, 2011).** While the original formulation of the Gromov-
773 Wasserstein distance in equation 2 is theoretically elegant, its computation is NP-hard. Conse-
774 quently, an alternative and more commonly used formulation was introduced. Given two mm-spaces
775 $\mathcal{X} = (X, d_X, \mu_X)$ and $\mathcal{Y} = (Y, d_Y, \mu_Y)$, for $p \geq 1$, the p -Gromov-Wasserstein distance is defined
776 as:
777

778
$$\mathcal{GW}_p(\mathcal{X}, \mathcal{Y}) := \left(\inf_{\pi \in \Pi(\mu_X, \mu_Y)} \int_{X \times Y} \int_{X \times Y} |d_X(x, x') - d_Y(y, y')|^p d\pi(x, y) d\pi(x', y') \right)^{\frac{1}{p}}. \quad (10)$$

779
780

781 Mémoli (2011) proved that the formulation in equation 10 is bi-Hölder equivalent to the one in equa-
782 tion 2. Moreover, equation 2 constitutes a metric in the sense of Definition 1.
783784 **Computational Costs.** We now briefly review the computational complexity of these OT-based
785 methods for discrete distributions supported on n points. Computing the standard Wasserstein
786 distance in equation 1 using a classic EMD solver has a worst-case complexity of $\mathcal{O}(n^3)$ and
787 an average-case complexity of $\mathcal{O}(n^2)$ (Bonneel et al., 2011). Alternatively, solving the entropy-
788 regularized OT problem with the Sinkhorn algorithm (Cuturi, 2013) has a complexity of up to
789 $\mathcal{O}(n^2 \log n)$. While some targeted algorithms may achieve faster runtimes (Altschuler et al., 2019;
790 Li et al., 2023a), this often comes at the cost of sacrificing metric properties. Computing the
791 Gromov-Wasserstein distance in equation 10 involves solving a non-convex quadratic program,
792 which has a complexity of at least $\mathcal{O}(n^3 \log n)$ (Xu et al., 2019). Faster algorithms, such as those
793 employing entropic regularization (Peyré et al., 2016; Scetbon et al., 2022; Li et al., 2023b), typ-
794 ically represent a trade-off between speed and accuracy. As for the one-dimensional Wasserstein
795 distance in equation 9, the computation is significantly faster, requiring only $\mathcal{O}(n \log n)$ time via
796 sorting. This efficiency has motivated many methods that project higher-dimensional distributions
797 onto one-dimensional lines for comparison (Bonneel et al., 2015; Deshpande et al., 2019; Le et al.,
798 2019; Li et al., 2024; 2025). However, for the Gromov-Wasserstein distance, the development of
799 sliced variants is less mature, with Sliced Gromov-Wasserstein (SGW) (Titouan et al., 2019) being
800 the most prominent example that has achieved empirical success.
801802 B PROOFS
803804 B.1 PROOF OF THEOREM 1
805806 The proof for the symmetry and triangle inequality properties is straightforward, as they are directly
807 inherited from the Wasserstein distance \mathcal{W}_p . We focus on proving the **Identity of Indiscernibles**,
808 which is the most involved part.
809810 To do so, we first recall an equivalent characterization of isometry from (Sturm, 2023, Lemma 1.10):
811812 (i) \mathcal{X}, \mathcal{Y} are isometric.
813

810 (ii) There exists a coupling $\pi \in \Pi(\mu_X, \mu_Y)$ such that $d_X(\mathbf{x}_0, \mathbf{x}_1) = d_Y(\mathbf{y}_0, \mathbf{y}_1)$ holds for
 811 $\pi \otimes \pi$ -almost every pair of points $((\mathbf{x}_0, \mathbf{y}_0), (\mathbf{x}_1, \mathbf{y}_1))$.
 812

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

816 **Proof.** By definition, the condition $\mathcal{QD}_p(\mathcal{X}, \mathcal{Y}) = 0$ implies that there exists an optimal coupling
 817 π between the pushforward measures on the representation space, such that $\phi^X(\mathbf{x}) = \phi^Y(\mathbf{y})$ holds
 818 for all (\mathbf{x}, \mathbf{y}) in the support of the corresponding coupling on $X \times Y$.
 819

820 For a given quantile level vector $\mathbf{q} \in (0, 1)^k$, let us denote the corresponding QDMs as $\mathbf{m}_i^X =$
 821 $\mathbf{m}^X(q_i)$ and $\mathbf{m}_i^Y = \mathbf{m}^Y(q_i)$ for $i = 1, 2, \dots, k$. The equality of the representations implies:

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

825 From $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2$ and the fact that isometric spaces have identical norm distributions (and thus
 826 identical quantile functions), it directly follows that the weights are equal for coupled points:
 827

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

830 For any $(\mathbf{x}, \mathbf{y}) \in \text{supp}(\pi)$, expanding the squared distance equality yields:
 831

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

837 where the last step uses $\|\mathbf{x}\|_2^2 = \|\mathbf{y}\|_2^2$. Integrating this equality against the weight function $w(\cdot, q_j)$
 838 over the respective spaces gives:
 839

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

848 By the definition of QDM in Eq. equation 5, this simplifies to:
 849

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

852 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
 853 back, we obtain the key result that the Gram matrices of the QDM sets are identical: $\mathbf{m}_i^{X^\top} \mathbf{m}_j^X =$
 854 $\mathbf{m}_i^{Y^\top} \mathbf{m}_j^Y$.
 855

856 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:
 857

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

861 Due to the dimensionality condition $\dim(\{\mathbf{m}^X(q_i)\}_{1 \leq i \leq k}) = \dim(X)$, any point $\mathbf{x} \in \text{supp}(\mu_X)$
 862 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
 863 can show that its coupled counterpart \mathbf{y} must be the same linear combination of the corresponding

864 QDMs in Y :

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

877 where the last equality uses $\|\mathbf{x}\|_2^2 = \|\mathbf{y}\|_2^2$. This implies $\mathbf{y} = \sum_{i=1}^k \alpha_i \mathbf{m}_i^Y$.880 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:

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

888 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

891 B.2 PROOF OF THEOREM 2

893 This section provides the proof for Theorem 2. We begin by establishing the necessary notation and 894 preliminary lemmas. Let $r_q = F_{\|\mathbf{X}\|_2}^{-1}(q)$ be the true q -th quantile of the norm distribution, and let 895 $\hat{r}_q = \hat{F}_{\|\mathbf{X}_n\|_2}^{-1}(q)$ be its empirical counterpart estimated from the sample set $\mathbb{X}_n = \{x_1, \dots, x_n\}$. 896 For brevity, we denote the weight functions as $w_i^X(x) = w^X(x, q_i)$ and its empirical version as 897 $\hat{w}_i^{\mathbb{X}_n}(x) = e^{-\sigma(\|\mathbf{x}\|_2 - \hat{r}_q)^2}$.

899 Our proof relies on the well-established convergence rates for the Wasserstein distance, summarized 900 in the following lemma.

901 **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 902 moment $\mathcal{M}_q(\mu)$ is finite, i.e., $\mathcal{M}_q(\mu) = \int_X \|\mathbf{x}\|^q d\mu(x) < \infty$, then the following holds:*

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

910 As discussed in (Fournier & Guillin, 2015), for a sufficiently large moment order q , the term 911 $n^{-(q-p)/q}$ becomes negligible compared to the leading term. For the common case of $p = 2$, 912 this holds for $q \geq 4$. We will assume this condition holds in our subsequent analysis.

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

915 **Lemma 2 (Convergence Rate of Quantiles)** *Let F_X be the CDF of a distribution μ_X , and assume 916 it is continuous on r_q and has a strictly positive density. Then, the mean squared error of the 917 empirical quantile converges as follows:*

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

918 This is a classical result in asymptotic theory, which can be found in (Serfling, 1980, Chap 2.3).
 919 Furthermore, our proof requires the following bounds. The first lemma ensures that the denominator
 920 in the QDM definition is well-behaved.

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

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

$$\begin{aligned} 928 \mathbb{E}_{\mu_X}(w_i^X(X)) &\geq \int_{\|\mathbf{x}\|_2 < M_0} e^{-\sigma(x-r_{q_i})^2} d\mu_X(x) \geq \int_{\|\mathbf{x}\|_2 < M_0} e^{-\sigma \max\{x, r_{q_i}\}^2} d\mu_X(x) \\ 929 &\geq \int_{\|\mathbf{x}\|_2 < M_0} e^{-\sigma \max\{M_0, r_{q_i}\}^2} d\mu_X(x) \geq p_0 e^{-\sigma \max\{M_0, r_{q_i}\}^2}. \end{aligned}$$

933 This implies that the expectation is strictly positive, and thus $\frac{1}{\mathbb{E}_{\mu_X}(w_i^X(X))} < \infty$. \square
 934

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

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

942 **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
 943 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 $\mathbb{E}(\bar{w}(\hat{r}_{q_i})) =
 944 M_1(\hat{w}_i)$. Now, consider the following probability:
 945

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

949 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
 950 Hoeffding's inequality yields:

$$\begin{aligned} 951 \mathbb{P}_1 &= \mathbb{P}(\mathbb{E}(\bar{w}_i(r_{q_i})) - \bar{w}_i(r_{q_i}) > \frac{n}{4} M_1(\hat{w}_i)) \\ 952 &= \mathbb{P}\left(\sum_{j=1}^n \mathbb{E}(w_i^X(\mathbf{x}_j)) - \sum_{j=1}^n w_i^X(\mathbf{x}_j) > \frac{n}{4} M_1(\hat{w}_i)\right) \\ 953 &\leq \exp(-\frac{n}{8} M_1^2(\hat{w}_i)). \end{aligned}$$

957 For the second term, \mathbb{P}_2 , we first bound the difference by applying the Mean Value Theorem:
 958

$$\begin{aligned} 959 |\bar{w}_i(r_{q_i}) - \bar{w}_i(\hat{r}_{q_i})| &\leq \frac{1}{n} \sum_{j=1}^n |e^{-\sigma(\|\mathbf{x}_j\|_2 - r_{q_i})} - e^{-\sigma(\|\mathbf{x}_j\|_2 - \hat{r}_{q_i})}| \\ 960 &\leq \frac{1}{n} \sum_{j=1}^n |2\sigma(\|\mathbf{x}_j\|_2 - r_{j0})e^{-\sigma(\|\mathbf{x}_j\|_2 - r_{j0})^2}| |r_{q_i} - \hat{r}_{q_i}| \\ 961 &\leq \sqrt{2\sigma} e^{-1/2} |r_{q_i} - \hat{r}_{q_i}| \quad (\text{since } \sup_{t \geq 0} 2\sigma t e^{-\sigma t^2} \leq \sqrt{2\sigma} e^{-1/2}). \end{aligned} \tag{11}$$

967 Consequently, we have:

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

971 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
 972 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

1026 **Proof of Theorem 2.** First, we use a property derived from the triangle inequality:

$$\begin{aligned} 1027 \quad |d(\mathbf{x}, \mathbf{x}') - d(\mathbf{y}, \mathbf{y}')| &\leq |d(\mathbf{x}, \mathbf{x}') - d(\mathbf{x}, \mathbf{y}')| + |d(\mathbf{x}, \mathbf{y}') - d(\mathbf{y}, \mathbf{y}')| \\ 1028 \quad &\leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{x}', \mathbf{y}'). \end{aligned}$$

1029 This allows us to bound the QDOT distance.

$$\begin{aligned} 1031 \quad \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(x, y) \\ 1032 \quad &\leq \inf_{\pi \in \Pi(\mu_X, \mu_Y)} \int \sum_{i=0}^k |\phi_i^X - \phi_i^Y|^p d\pi(x, y) \\ 1033 \quad &\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(x, y) \\ 1034 \quad &\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(x, y) \quad (12) \\ 1035 \quad &= (k+1)2^{p-1} \inf_{\pi \in \Pi(\mu_X, \mu_Y)} \int d(\mathbf{x}, \mathbf{y})^p d\pi(x, y) + 2^{p-1} \sum_{i=0}^k d(\mathbf{m}_i^X, \mathbf{m}_i^Y)^p \\ 1036 \quad &= C_0 \underbrace{\mathcal{W}_p(\mu_X, \mu_Y)^p}_{\text{Wasserstein Discrepancy}} + C_1 \underbrace{\sum_{i=1}^k d(\mathbf{m}_i^X, \mathbf{m}_i^Y)^p}_{\text{QDM Discrepancy}} \\ 1037 \quad &= C_0 \mathcal{W}_p(\mu_X, \mu_Y)^p + C_1 \sum_{i=1}^k d(\mathbf{m}_i^X, \mathbf{m}_i^Y)^p \\ 1038 \quad &= C_0 \mathcal{W}_p(\mu_X, \mu_Y)^p + C_1 \sum_{i=1}^k d(\mathbf{m}_i^X, \mathbf{m}_i^Y)^p \\ 1039 \quad &= C_0 \mathcal{W}_p(\mu_X, \mu_Y)^p + C_1 \sum_{i=1}^k d(\mathbf{m}_i^X, \mathbf{m}_i^Y)^p \\ 1040 \quad &= C_0 \mathcal{W}_p(\mu_X, \mu_Y)^p + C_1 \sum_{i=1}^k d(\mathbf{m}_i^X, \mathbf{m}_i^Y)^p \\ 1041 \quad &= C_0 \mathcal{W}_p(\mu_X, \mu_Y)^p + C_1 \sum_{i=1}^k d(\mathbf{m}_i^X, \mathbf{m}_i^Y)^p \\ 1042 \quad &= C_0 \mathcal{W}_p(\mu_X, \mu_Y)^p + C_1 \sum_{i=1}^k d(\mathbf{m}_i^X, \mathbf{m}_i^Y)^p \\ 1043 \quad &= C_0 \mathcal{W}_p(\mu_X, \mu_Y)^p + C_1 \sum_{i=1}^k d(\mathbf{m}_i^X, \mathbf{m}_i^Y)^p \\ 1044 \quad &= C_0 \mathcal{W}_p(\mu_X, \mu_Y)^p + C_1 \sum_{i=1}^k d(\mathbf{m}_i^X, \mathbf{m}_i^Y)^p \\ 1045 \quad &= C_0 \mathcal{W}_p(\mu_X, \mu_Y)^p + C_1 \sum_{i=1}^k d(\mathbf{m}_i^X, \mathbf{m}_i^Y)^p \\ 1046 \quad &= C_0 \mathcal{W}_p(\mu_X, \mu_Y)^p + C_1 \sum_{i=1}^k d(\mathbf{m}_i^X, \mathbf{m}_i^Y)^p \\ 1047 \quad &= C_0 \mathcal{W}_p(\mu_X, \mu_Y)^p + C_1 \sum_{i=1}^k d(\mathbf{m}_i^X, \mathbf{m}_i^Y)^p \\ 1048 \quad &= C_0 \mathcal{W}_p(\mu_X, \mu_Y)^p + C_1 \sum_{i=1}^k d(\mathbf{m}_i^X, \mathbf{m}_i^Y)^p \end{aligned}$$

1049 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
1050 (Fournier & Guillin, 2015, Theorem 1), we have the bound for the Wasserstein Discrepancy between
1051 the true measure and its empirical version:

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

1052 Next, we analyze the QDM Discrepancy term, $d(\mathbf{m}_i^X, \hat{\mathbf{m}}_i^{\mathbb{X}_n})$.

$$\begin{aligned} 1053 \quad d(\mathbf{m}_i^X, \hat{\mathbf{m}}_i^{\mathbb{X}_n}) &= \left\| \frac{\frac{1}{n} \sum_{j=1}^n \hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_j) \mathbf{x}_j}{\frac{1}{n} \sum_{j=1}^n \hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_j)} - \frac{\mathbb{E}_{\mu_X}(w_i^X(X)X)}{\mathbb{E}_{\mu_X}(w_i^X(X))} \right\|_2 \\ 1054 \quad &= \left\| \frac{\frac{1}{n} \sum_{j=1}^n \hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_j) \mathbf{x}_j \mathbb{E}_{\mu_X}(w_i^X(X)) - \mathbb{E}_{\mu_X}(w_i^X(X)X) \frac{1}{n} \sum_{j=1}^n \hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_j)}{\frac{1}{n} \sum_{j=1}^n \hat{w}_i^{\mathbb{X}_n}(\mathbf{x}_j) \mathbb{E}_{\mu_X}(w_i^X(X))} \right\|_2 \\ 1055 \quad &\leq C_2 \underbrace{\left| \mathbb{E}_{\mu_X}(w_i^X(X)) - \frac{1}{n} \sum_{j=1}^n w_i^X(x_j) \right|}_{\text{Term (I)}} \\ 1056 \quad &\quad + C_2 \underbrace{\left\| \hat{w}_i^{\mathbb{X}_n} \right\|_2 \left\| \frac{1}{n} \sum_{j=1}^n \hat{w}_i^{\mathbb{X}_n}(x_j) x_j - \mathbb{E}_{\mu_X}(w_i^X(X)X) \right\|_2}_{\text{Term (II)}}. \end{aligned} \quad (13)$$

1057 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
1058 now focus on the two main terms. Term (I) can be decomposed as:

$$\begin{aligned} 1059 \quad \text{Term (I)} &\leq \underbrace{\left| \mathbb{E}_{\mu_X}(w_i^X(X)) - \frac{1}{n} \sum_{j=1}^n w_i^X(x_j) \right|}_{\text{Term (I.1)}} + \underbrace{\left| \frac{1}{n} \sum_{j=1}^n w_i^X(x_j) - \frac{1}{n} \sum_{j=1}^n \hat{w}_i^{\mathbb{X}_n}(x_j) \right|}_{\text{Term (I.2)}}. \end{aligned} \quad (14)$$

1060 For Term (I.1), standard results for the mean of i.i.d. variables give:

$$\begin{aligned} 1061 \quad \mathbb{E}(\text{Term (I.1)}) &\leq \sqrt{\mathbb{E} \left(\mathbb{E}_{\mu_X}(w_i^X(X)) - \frac{1}{n} \sum_{j=1}^n w_i^X(x_j) \right)^2} = \sqrt{\frac{1}{n} \text{Var}(w_i^X(x_1))} = O(n^{-1/2}). \end{aligned}$$

1080 Using the bound from equation 11, we have:
 1081

$$1082 \text{Term (I.2)} \leq \sqrt{2\sigma} e^{-1/2} |r_{q_i} - \hat{r}_{q_i}|.$$

1083 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
 1084 similar decomposition:
 1085

$$1086 \text{Term (II)} \leq \|\hat{m}_i^{\mathbb{X}_n}\|_2 \underbrace{\left\| \frac{1}{n} \sum_{j=1}^n \hat{w}_i^{\mathbb{X}_n}(x_j) x_j - \frac{1}{n} \sum_{j=1}^n w_i^X(x_j) x_j \right\|_2}_{\text{Term (II.1)}} \\ 1087 \\ 1088 \\ 1089 \\ 1090 \\ 1091 \\ 1092 \\ 1093 \\ 1094 \\ 1095 \\ 1096 \\ 1097 \\ 1098 \\ 1099 \\ 1100 \\ 1101 \\ 1102 \\ 1103 \\ 1104 \\ 1105 \\ 1106 \\ 1107 \\ 1108 \\ 1109 \\ 1110 \\ 1111 \\ 1112 \\ 1113 \\ 1114 \\ 1115 \\ 1116 \\ 1117 \\ 1118 \\ 1119 \\ 1120 \\ 1121 \\ 1122 \\ 1123 \\ 1124 \\ 1125 \\ 1126 \\ 1127 \\ 1128 \\ 1129 \\ 1130 \\ 1131 \\ 1132 \\ 1133$$

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

For Term (II.1), we have:

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

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:

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

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 $\Pi(\mu_X, \mu_Y)$.

1134 A proof of this lemma can be found in (Villani et al., 2008, Chap 1).
 1135

1136 **Proof of Triangle Inequality.** Consider three mm-spaces $\mathcal{X} = (X, d_X, \mu_X)$, $\mathcal{Y} = (Y, d_Y, \mu_Y)$,
 1137 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
 1138 the respective pairwise optimal couplings for the IQDOT distance. According to Lemma 7, we can
 1139 find a joint coupling $\pi_{X,Y,Z} \in \Pi(\mu_X, \mu_Y, \mu_Z)$ such that

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

1140 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
 1141 IQDOT distance as follows:
 1142

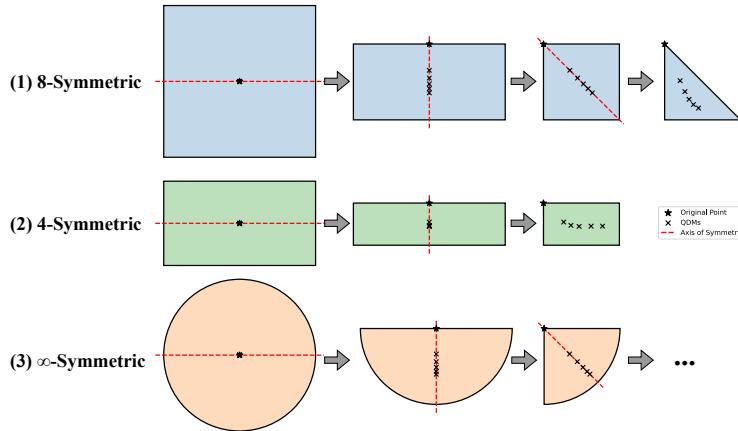
$$\begin{aligned} \mathcal{IQD}_p(\mathcal{X}, \mathcal{Y}) &= \mathbb{E}_{(0,1) \times \pi_{X,Y}^*} (|\phi^X(x, q) - \phi^Y(y, q)|^p)^{1/p} \\ &\leq \mathbb{E}_{(0,1) \times \pi_{X,Y}} (|\phi^X(x, q) - \phi^Y(y, q)|^p)^{1/p} \\ &\leq \mathbb{E}_{(0,1) \times \pi_{X,Y,Z}} ((|\phi^X(x, q) - \phi^Z(z,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

1143 where the last inequality follows from Minkowski’s inequality. This completes the proof. \square
 1144

B.4 RELATIONSHIP BETWEEN QDOT AND IQDOT

1145 It is evident that both QDOT and IQDOT utilize the same QDMD representations in their numerical
 1146 implementations. Since IQDOT is formulated as a pairwise 1-D Wasserstein distance, the inequality
 1147 $\mathcal{IQD} \leq \mathcal{QD}$ consistently holds in practice. Consequently, the convergence $\mathcal{QD}(\mu_n, \mu) \rightarrow 0$ implies
 1148 $\mathcal{IQD}(\mu_n, \mu) \rightarrow 0$. Furthermore, as both constitute distinct well-defined metrics, under the requisite
 1149 conditions, they share the equivalence property regarding in distinguishability: $\mathcal{QD}(\mu, \nu) = 0 \Leftrightarrow$
 1150 $\mathcal{IQD}(\mu, \nu) = 0$.
 1151

B.5 DISCUSSIONS ON THE SYMMETRIC SHAPES.



1152
 1153 Figure 7: Illustrations of simple symmetric shapes. Note that for non-radially symmetric shapes, utilizing a
 1154 finite number of QDMs suffices to satisfy the dimension conditions in Theorem 1 (i.e., ensuring that they are not
 1155 collinear). For radially symmetric shapes, infinite subdivision implies a degeneration into a radial distribution.
 1156

C IMPLEMENTATION DETAILS

C.1 DETAILS OF ALGORITHM 1

1157 The following provides further details on some of the operations and parameter choices in Algo-
 1158 rithm 1.
 1159

1188
 1189 **Data Initialization.** For two unknown distributions situated in Euclidean spaces, preprocessing
 1190 the data is necessary. The objective is to obtain the distribution of distances relative to the barycen-
 1191 ter, which is isometry-invariant. However, this initialization step is not necessary for intra-space
 1192 comparisons or in cases where a correspondence between the origins is already known. The in-
 1193 tuiti on of scalization aligns with the numerical implementation of Gromov-Wasserstein (Flamary
 1194 et al., 2021), serving as a common technique for metric space alignment.

1195
 1196 **Quantile Level Vector q .** The quantile levels in the vector q are theoretically chosen from the
 1197 open interval $(0, 1)$. In practice, we select these levels by taking equispaced points within a truncated
 1198 interval $[\delta, 1 - \delta]$. The default value for δ is 0.1, which avoids numerical instability at the extreme
 1199 tails of the distribution. For larger datasets where the empirical quantiles are more stable, a smaller
 1200 δ may be used to satisfy the theoretical conditions more closely.

1201
 1202 **Choice of the Number of Quantiles k .** A large number of quantiles, k , is often not required
 1203 to achieve strong performance. For low-dimensional data (e.g., 3D point clouds), the trilateration
 1204 condition is readily satisfied, and a small value such as $k = 5$ can be sufficient. For high-dimensional
 1205 data, which often exhibits low-rank structure, a value of k smaller than the ambient dimension can
 1206 also be effective. While a larger k can lead to improved numerical stability, we find that $k = 50$
 1207 provides a robust default choice across most applications.

1208
 1209 **Bandwidth Parameter σ .** The choice of the bandwidth parameter σ in the Gaussian kernel is
 1210 related to the number of support points, n . A very small σ will cause the weights to concentrate on a
 1211 single point, while a very large σ will lead to nearly uniform weights across all samples. Empirically,
 1212 we recommend choosing σ such that the maximum weight is on the order of the average weight,
 1213 i.e., $\max_i\{w_i\} \sim n^{-1/2}$. In practice, a default value of $\sigma = 10$ provides robust and effective
 1214 performance across most of our experiments.

1215 We analyze the robustness of our method to these parameters through a toy example detailed in
 1216 Appendix D.1.

1217 C.2 ALGORITHM OF IQDOT

1219 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 corre-
 1220 sponding probability vectors $\mathbf{p}^X \in \Delta^{n-1}$ and $\mathbf{p}^Y \in \Delta^{m-1}$, we now detail the numerical implemen-
 1221 tation of the IQDOT, as introduced in Definition 4. The procedure is summarized in Algorithm 2.

1224 Algorithm 2 Integal-QDOT

1225 **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: $\mathbf{r}^X \leftarrow (F_{\|\mathbf{X}\|}^{-1}(q_1), \dots, F_{\|\mathbf{X}\|}^{-1}(q_k))$, $\mathbf{r}^Y \leftarrow (F_{\|\mathbf{Y}\|}^{-1}(q_1), \dots, F_{\|\mathbf{Y}\|}^{-1}(q_k))$. $\triangleright \mathcal{O}(n \log n + m \log m)$
- 4: **for** $i \leftarrow 1$ to k **do** $\triangleright \mathcal{O}(knd + kms)$
 - 5: Compute the quantile weights:

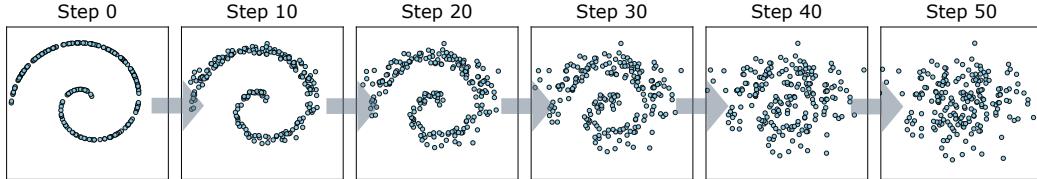
$$\mathbf{w}_{ij}^X \leftarrow \frac{p_j^X \exp\{-\sigma(\mathbf{d}_j^X - \mathbf{r}_i^X)^2\}}{\sum_{j'=1}^n p_{j'}^X \exp\{-\sigma(\mathbf{d}_{j'}^X - \mathbf{r}_i^X)^2\}}, \quad \mathbf{w}_{ij}^Y \leftarrow \frac{p_j^Y \exp\{-\sigma(\mathbf{d}_j^Y - \mathbf{r}_i^Y)^2\}}{\sum_{j'=1}^m p_{j'}^Y \exp\{-\sigma(\mathbf{d}_{j'}^Y - \mathbf{r}_i^Y)^2\}}$$
 - 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:

$$\phi_i^X \leftarrow (\|\mathbf{x}_j - \mathbf{m}_i^X\|)_{1 \leq j \leq n}, \quad \phi_i^Y \leftarrow (\|\mathbf{y}_j - \mathbf{m}_i^Y\|)_{1 \leq j \leq m}$$
 - 8: Calculate the 1D- \mathcal{W}_p loss: $\mathcal{L}_i \leftarrow \mathcal{W}_p((\phi_i^X)_{\#} \mu_{\mathbf{X}}, (\phi_i^Y)_{\#} \mu_{\mathbf{Y}})$ $\triangleright \mathcal{O}(n \log n + m \log m)$
- 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

1242 **Computational Cost.** Similar to the analysis for QDOT, the pre-computation of QDMs and QD-
 1243 MDs is highly efficient. The final one-dimensional Wasserstein distances can also be computed
 1244 in $\mathcal{O}(n \log n)$ time by North-West corner rule(Peyré et al., 2019). Consequently, the overall time
 1245 complexity of the IQDOT algorithm is $\mathcal{O}(n \log n)$, assuming k and the data dimensions are small
 1246 constants.

D ADDITIONAL EXPERIMENT RESULTS

D.1 TOY EXAMPLE AND PARAMETER ANALYSIS



1259 Figure 8: Illustration of the Wasserstein flow from a spiral point cloud to a random Gaussian noise distribution
 1260 over 50 steps.
 1261

1262 To provide an intuitive visualization of QDOT’s behavior and to analyze its sensitivity to key parameters,
 1263 we present a toy example. As illustrated in Figure 8, the experiment consists of a Wasserstein
 1264 flow sequence that interpolates between a 2D spiral distribution and a random Gaussian noise distribution.
 1265 Each point cloud in the 50-step sequence comprises $n = 200$ points. We conduct a series of
 1266 tests to: (1) compare the dissimilarity trends produced by different geometric metrics as the spiral
 1267 deforms into noise; (2) analyze the effect of varying the bandwidth parameter σ ; (3) analyze the
 1268 effect of varying the quantile interval parameter δ ; and (4) assess the impact of different initialization
 1269 and scaling strategies.

1270 The results, presented in Figure 9, reveal several key insights. First, when compared to baselines
 1271 such as GW, EGW, SGW, and RISGW, both QDOT and IQDOT produce significantly smoother
 1272 and more stable dissimilarity curves. Second, the analysis of the bandwidth parameter shows that
 1273 for $\sigma = 1$ and $\sigma = 10$, the resulting curves are stable and smooth. A larger value of $\sigma = 100$
 1274 introduces some volatility, which can be attributed to the interaction between a high bandwidth and
 1275 the relatively small sample size. Third, the dissimilarity curve is largely insensitive to the choice
 1276 of the quantile interval parameter δ in this experiment, indicating good robustness. Finally, while
 1277 different initialization and scaling strategies affect the absolute values of the dissimilarity, they all
 1278 produce smooth, monotonically increasing curves that preserve the overall trend, demonstrating the
 1279 robustness of the underlying geometric representation.

D.2 CROSS SPACE TASKS

1282 **Evaluation Metrics.** For the numerical experiments, the inputs are a 3D point cloud sequence,
 1283 represented by sample matrices $\mathbf{X} \in \mathbb{R}^{n \times 3}$, and a reference shape. The reference is provided both
 1284 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
 1285 have corresponding probability vectors, $\mathbf{p}^{\mathbf{X}}$ and $\mathbf{p}^{\mathbf{Y}}$, we assume access to a ground-truth alignment
 1286 between the sequence frame \mathbf{X} and the reference shape \mathbf{Y} , denoted by a map \mathcal{T} , i.e., $\mathbf{X} = \mathcal{T}(\mathbf{Y})$.
 1287 Algorithm 1 computes a coupling matrix $\Pi \in \mathbb{R}^{n \times m}$ between the input distributions, which satisfies
 1288 the marginal constraints $\Pi \mathbf{1}_m = \mathbf{p}^{\mathbf{Y}}$ and $\Pi^{\top} \mathbf{1}_n = \mathbf{p}^{\mathbf{X}}$. To evaluate the quality of the resulting
 1289 match, we define the following metrics.

1290 The **Transformed Mean Squared Error (TMSE)** evaluates the alignment cost in the original 3D
 1291 ambient space. Even when the coupling Π is computed between a 3D shape \mathbf{X} and a 2D projection
 1292 $\hat{\mathbf{Y}}$, the error is measured by using the original 3D coordinates of the reference shape, \mathbf{Y} :

$$1294 \text{TMSE} = \sum_{i=1}^n \sum_{j=1}^m \Pi_{ij} d_X(\mathbf{x}_i, \mathcal{T}(\mathbf{y}_j));$$

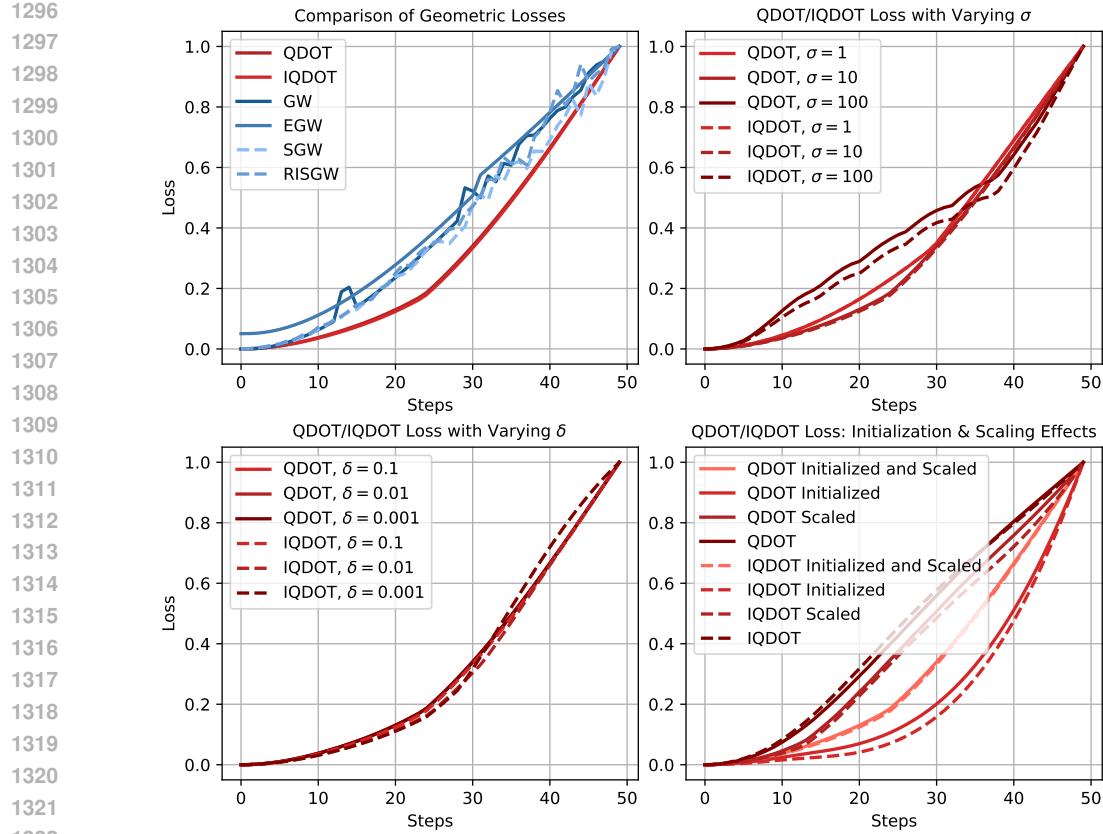


Figure 9: Results from the four experimental settings. The x-axis represents the step in the Wasserstein flow sequence, and the y-axis represents the computed dissimilarity (loss).

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., $\tau = F_{d_X}^{-1}(0.2)$.

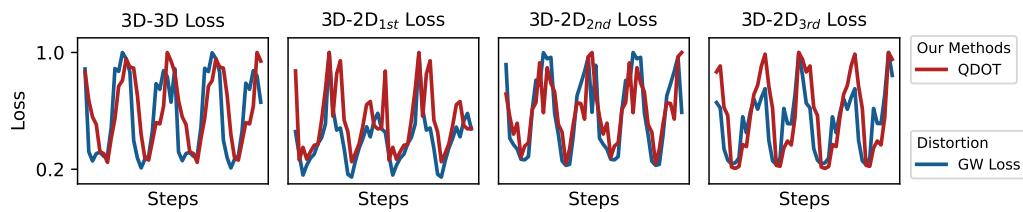
Additional comparisons. To demonstrate the efficiency of QDOT, we compared it against several fast GW/EGW variants, such as PGW (Kerdoncuff et al., 2021), QGW (Chowdhury et al., 2021), and LRGW (Scetbon et al., 2022), as shown below. The penalty parameter for all methods was set to 0.01.

Table 5: Additional comparisons on the camel-gallop dataset

Methods	Transformed MSE \downarrow					Inlier Ratio (%) \uparrow					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.

1350
 1351 **Distortion.** We substitute the coupling obtained via Algorithm 1 into the Gromov-Wasserstein loss
 1352 function in 10. By plotting the loss curves over the optimization steps, we compare the behavior of
 1353 the QDOT coupling with respect to both the QDOT loss and the GW loss. We employed equidistant
 1354 sampling to select approximately $n \approx 1000$ points. The experimental results are presented below.
 1355



1363 Figure 10: Distortion results on the camel-gallop dataset. The x-axis represents the time steps of the galloping
 1364 camel sequence, and the y-axis indicates the loss corresponding to the coupling at each time step.
 1365

1366 As shown in Figure 10, the QDOT loss and the GW loss maintain a high degree of consistency under
 1367 the same coupling. This validates the effectiveness of QDOT as a metric.
 1368

1369 **Ablation study.** To analyze the sensitivity of the results to the parameter settings, we consider
 1370 $\sigma \in \{200, 1000, 5000\}$ and $k \in \{20, 100, 500\}$. The experimental results are presented below.
 1371

Table 6: QDOT results with different parameter selections on the camel-gallop dataset.
 1372

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_{mean±std}	0.22	0.24	0.23	0.25	0.24	70.95	63.68	68.22	63.04	66.47
	(± 0.00)	(± 0.01)	(± 0.00)	(± 0.01)	(± 0.01)	(± 0.28)	(± 0.45)	(± 0.27)	(± 0.19)	(± 0.25)

1384 As shown in Table 6, the QDOT alignment results are not particularly sensitive to the selection of
 1385 σ and k . Generally, $k = 100$ yields sufficiently good performance, with larger k providing minimal
 1386 additional gains.
 1387

1388 D.3 TRANSFER LEARNING

1389 **Comparison with other sliced methods.** Setting $n = 2048$, we compared the results against
 1390 recently efficient sliced methods, including Sliced Wasserstein(Bonneel et al., 2015), Max-Sliced
 1391 Wasserstein(Deshpande et al., 2019), and Hilbert Curve Projection(Li et al., 2024), as shown below.
 1392

1393 As shown in Table 7, it is evident that IQDOT exhibits superior comprehensive performance. Notably,
 1394 in the ‘guitar’ category, QDOT achieves near-perfect accuracy, whereas other methods fail
 1395 almost entirely. This is attributed to the differing orientations of the guitars in the two datasets. As
 1396 illustrated in Figure 11, guitars in ModelNet are horizontally aligned, while those in ShapeNet are
 1397 vertically aligned, leading to the failure of other sliced-based methods. This highlights the critical
 1398 importance of isometry invariance in transfer learning.
 1399

1400 **Ablation study.** On the ablation study, we consider $n = 1024$, $\sigma \in \{50, 200, 1000\}$, and $k \in$
 1401 $\{10, 50, 200\}$. The results are shown below.
 1402

1403 As indicated in Table 8, apart from some numerical instability when $k = 10$, the differences between
 1404 results are not substantial. This demonstrates that our method is relatively insensitive to parameter
 1405 variations.
 1406

1404

1405

1406

1407

1408

1409

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

1420

1421

1422

1423

1424

1425

1426

1427

1428

1429

1430

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 (± 0.33)	82.54 (± 2.30)	83.33 (± 1.02)	96.65 (± 0.55)	50.32 (± 0.72)	96.68 (± 0.46)	84.40 (± 0.81)	83.48 (± 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 (± 0.34)	85.03 (± 2.31)	90.35 (± 1.04)	98.34 (± 0.63)	79.63 (± 1.50)	96.32 (± 0.87)	95.82 (± 0.75)	93.14 (± 0.67)

1454

1455

1456

1457

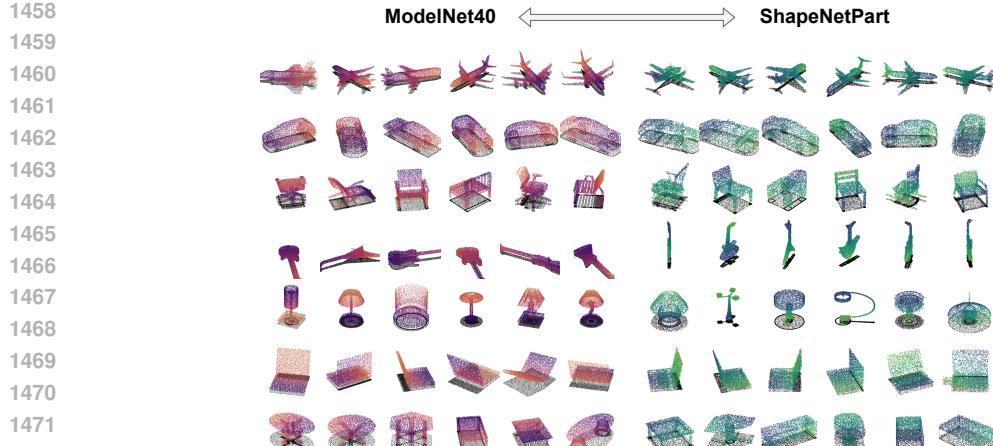


Figure 11: Visualization of the ModelNet40 and ShapeNetPart datasets.

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:

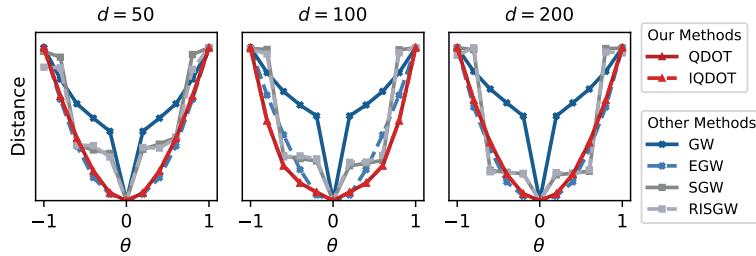


Figure 12: High dimension results.

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
\rightarrow 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
\rightarrow 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
\rightarrow 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
\rightarrow 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
\rightarrow 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
\rightarrow 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
\rightarrow 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
\rightarrow 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
\rightarrow 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
\rightarrow 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
\rightarrow 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
\rightarrow 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
\rightarrow 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
\rightarrow 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
\rightarrow 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
\rightarrow 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

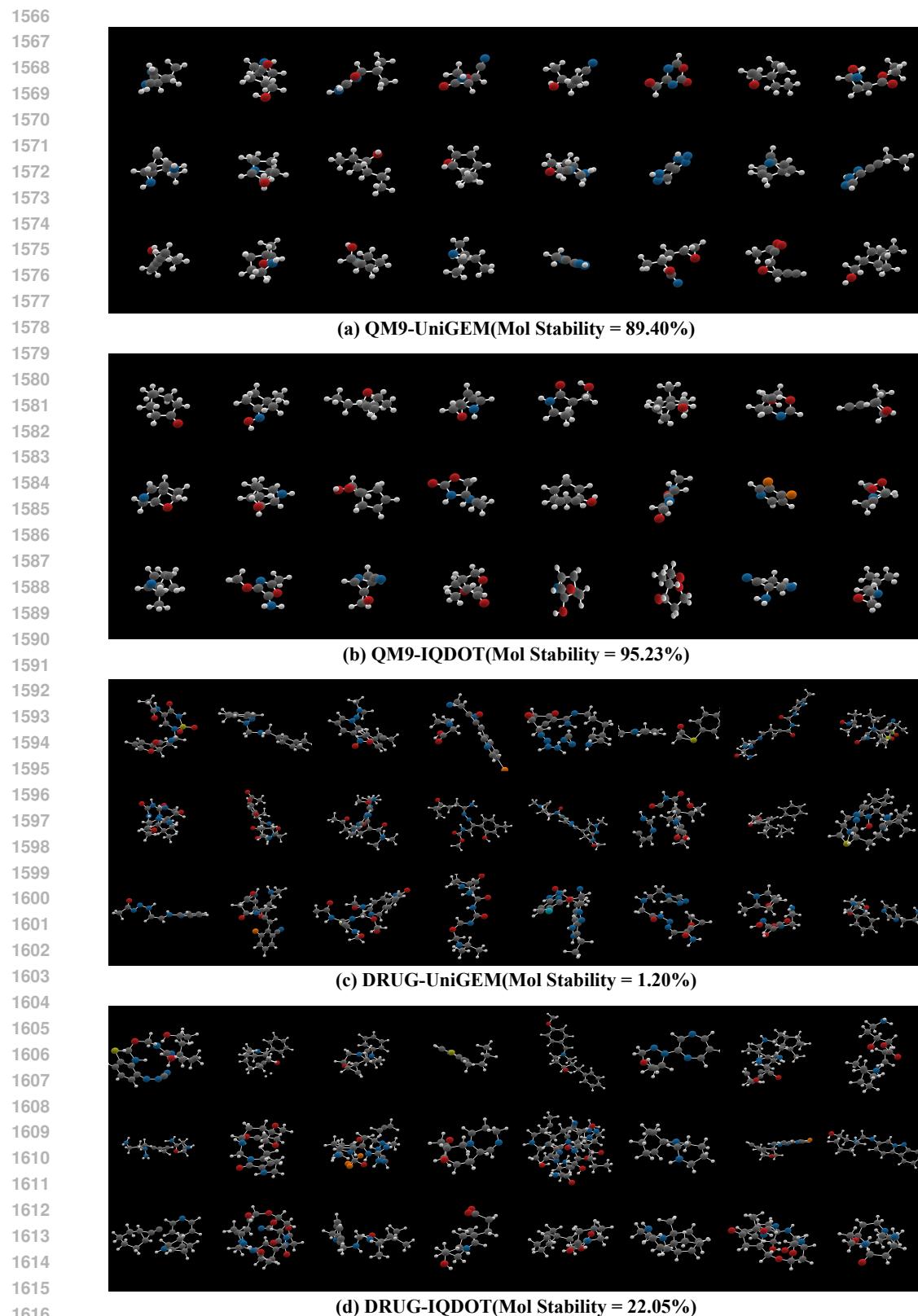
¹ The notation \rightarrow IQDOT_{0.2} indicates that the objective function $\mathcal{L} = 0.2\mathcal{L}_{\text{IQDOT}} + 0.8\mathcal{L}_{\text{EDM}}$.

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

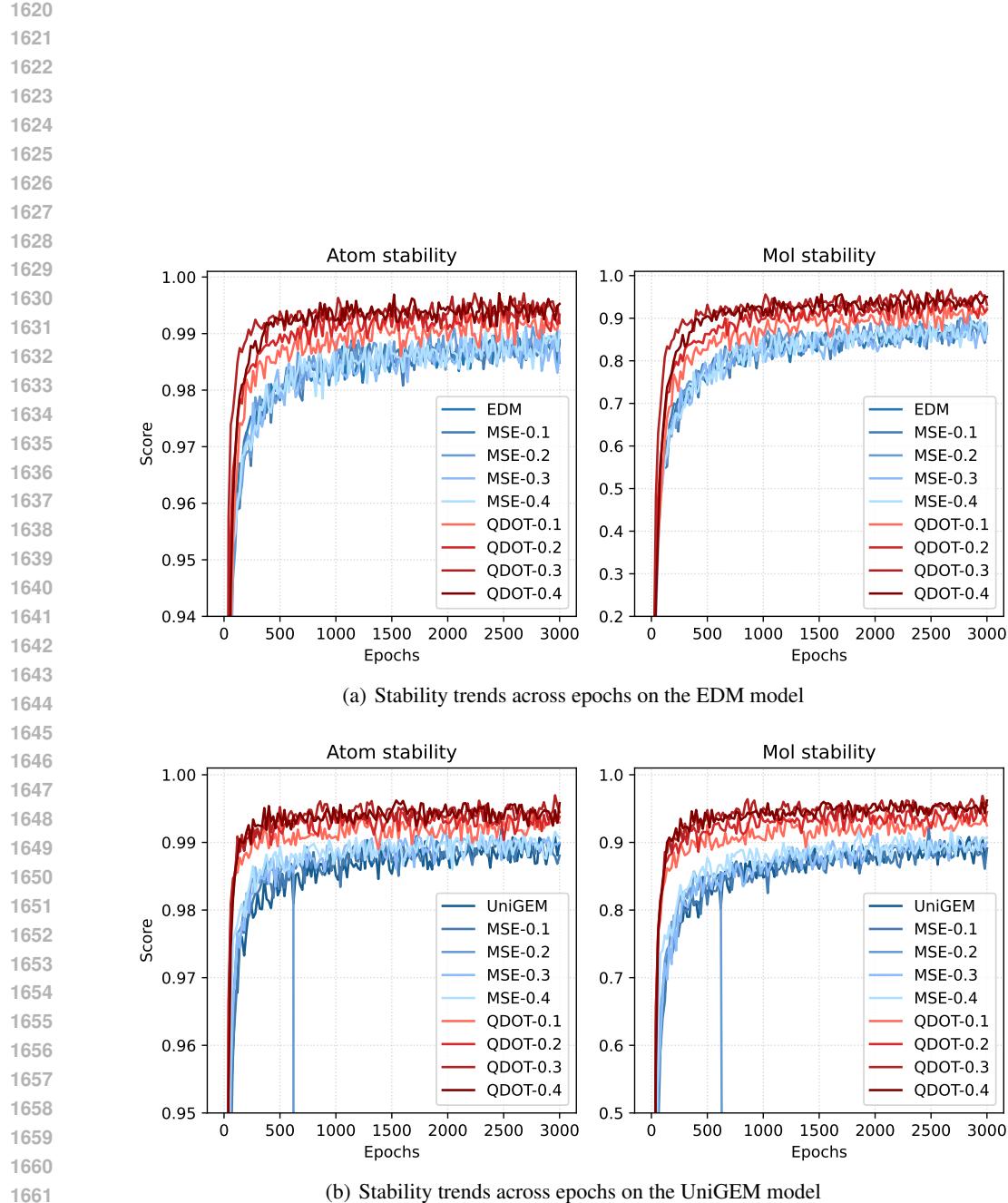
As illustrated in Figure 14, models trained with the IQDOT loss achieve stable convergence on both Atom and Molecule Stability metrics in just 1000 epochs, for both the EDM and UniGEM architectures. In contrast, the baseline models and those fine-tuned by simply adjusting the MSE loss weight require 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.



1618 Figure 13: Visualization of molecular generation results comparing IQDOT and UniGEM on the QM9 and
 1619 GEOM-Drugs datasets.



1663 Figure 14: Convergence trends of Atom Stability and Molecule Stability for various methods on the EDM and
1664 UniGEM models. The horizontal axis represents the number of training epochs, and the vertical axis represents
1665 the stability percentage. The comparison includes the baseline models and variants with different weights for
1666 the MSE and IQDOT loss terms (MSE-0.1 to MSE-0.4 and IQDOT-0.1 to IQDOT-0.4).

1667
1668
1669
1670
1671
1672
1673