# Two-sample goodness-of-fit tests on the flat torus based on Wasserstein distance and their relevance to structural biology[*]

**Javier González-Delgado**[1,2], **Alberto González-Sanz**[1,3],
**Juan Cortés**[2] **and Pierre Neuvial**[1]

[1]*Institut de Mathématiques de Toulouse, UMR 5219, Université de Toulouse, CNRS*
*e-mail:* javier.gonzalez-delgado@math.univ-toulouse.fr
alberto.gonzalez_sanz@math.univ-toulouse.fr; pierre.neuvial@math.univ-toulouse.fr

[2]*LAAS-CNRS, Université de Toulouse, CNRS*
*e-mail:* juan.cortes@laas.fr

[3]*ImUva, Universidad de Valladolid*

**Abstract:** This work is motivated by the study of local protein structure, which is defined by two variable dihedral angles that take values from probability distributions on the flat torus. Our goal is to provide the space $\mathcal{P}(\mathbb{R}^2/\mathbb{Z}^2)$ with a metric that quantifies local structural modifications due to changes in the protein sequence, and to define associated two-sample goodness-of-fit testing approaches. Due to its adaptability to the geometry of the underlying space, we focus on the Wasserstein distance as a metric between distributions.

We extend existing results of the theory of Optimal Transport to the $d$-dimensional flat torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$, in particular a Central Limit Theorem for the fluctuations of the empirical optimal transport cost. Moreover, we propose different approaches for two-sample goodness-of-fit testing for the one and two-dimensional case, based on the Wasserstein distance. We prove their validity and consistency. We provide an implementation of these tests in R. Their performance is assessed by numerical experiments on synthetic data and illustrated by an application to protein structure data.

**Keywords and phrases:** Optimal transport, flat torus, Wasserstein distance, central limit theorem, goodness-of-fit test, structural biology, intrinsically disordered proteins.

Received April 2022.

## Contents

## 1. Introduction

When it comes to measure the distance between two probability distributions, the well known Wasserstein distance, derived from the theory of Optimal Transport (OT), provides both strong theoretical guarantees –it metrizes weak convergence [66]– and attractive empirical performance [50]. Most of the applications of such theory are related to the very active field of machine learning, notably in the framework of generative networks [2], robustness [59] or fairness [21], among others.

From a statistical point of view, one of the main caveats of the theory of OT comes from the *curse of dimensionality*: the rate of convergence of the empirical Wasserstein distance decreases as $n^{-1/d}$ with the dimension [26]. Another important issue is the asymptotic behavior of the fluctuations of the empirical optimal transport cost. For probability measures supported in $\mathbb{R}^d$, it has been proved, using Efron–Stein's inequality that, for the cost $L^2$, the difference $\sqrt{n}(\mathcal{W}_2^2(P_n,Q) - \mathbb{E}\mathcal{W}_2^2(P_n,Q))$ is asymptotically Gaussian [22]. Recently, the proofs have been extended to some general costs in $\mathbb{R}^d$, including the cost $L^p$, for $p > 1$ [19]. Concerning statistical goodness-of-fit tests based on Wasserstein distance, the one-sample case has already been addressed in [30] and, when the probability distributions are defined over $\mathbb{R}$, two-sample tests can be derived from [17, 47].

In this paper, we focus on the $d$-dimensional flat torus $\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$ where, even from the purely theoretical point of view, OT has not been completely addressed, besides the work in [14], [42] or, more recently, in [39]. However, this space appears naturally when the probability measures are periodic (e.g. for distributions of angles). The main objective of this work is (1) to extend recent existing OT results to the space of probability measures on the flat torus $\mathcal{P}(\mathbb{T}^d)$, especially a Central Limit Theorem (CLT) for the fluctuations of the empirical optimal transport cost, and (2) to address in particular the two-dimensional case, by constructing two-sample goodness-of-fit tests based on the Wasserstein distance.

Our motivation for extending the theory of OT to $\mathbb{T}^2$ comes from the investigation of proteins. Understanding the relationships between protein sequence, structure and function is the main goal of Structural Biology. In addition to its scientific importance, a better understanding of these relationships is essential for applications in diverse areas, such as biomedicine and biotechnology. The conformational state of a protein can be defined by a vector of angles, corresponding to rotations around the chemical bonds between the atoms that constitute its "backbone". This vector contains two values per amino-acid, $\phi$ and $\psi$, which follow a certain distribution, and which are usually represented using the so-called Ramachandran plots [53] (see also Figure 3). The analysis of these distributions has several important applications, such as the validation or refinement of protein structures determined from biophysical techniques [46, 38], the prediction of some biophysical measurements to complement experiments [60], and the development of potential energy models or scoring methods for protein structure modeling, prediction and design [4, 55, 63].

In this context, the definition of a suitable distance between distributions on $\mathbb{T}^2$ is essential. This would allow to quantify the expected magnitude of structural effects associated with local changes in the sequence, and therefore to develop improved versions of the aforementioned modeling and prediction techniques. Nevertheless, this has not been done satisfactorily in previous works. For example, significant differences between two laws are stated after visual comparison of two empirical distributions in [55] and [60], and the Hellinger distance is used to compare distributions on a non-periodic $[-\pi, \pi] \times [-\pi, \pi]$ in [63]. Powerful statistical tests remain to be defined and implemented in order to state such differences, being based on a metric that takes geometry into consideration. As many other commonly-used metrics, Hellinger distance ignores the underlying geometry of the space. Here, we propose to use the Wasserstein distance, whose advantageous geometrical and mathematical properties are described in [50], [65] and [66], to define goodness-of-fit testing techniques for two measures on $\mathbb{T}^2$, allowing a more accurate study of the distribution of protein local conformations.

The paper is organized as follows:

- Section 2 starts by introducing the general framework of measures on the flat torus in general dimension, followed by the precise formulation of the optimal transport problem. Section 2.1 is devoted to the study of the shape

of the solutions, recalling that they are the gradients of periodic convex functions and showing the uniqueness of the potential in Corollary 2.2. Section 2.2 proves through Theorem 2.5 that the optimal transport potentials converge, up to an additive constant, when the measures converge weakly. This result implies that the method of [22] based on Efron–Stein's inequality can be applied to derive a Central Limit Theorem, see Theorem 2.6 in Subsection 2.3. Finally, we show how the previously defined CLT does not allow the definition of an asymptotic test.

- Section 3 shows how Wasserstein distance can be used to define two-sample goodness-of-fit tests in the two-dimensional flat torus. We propose two testing approaches. The first one, introduced in Subsection 3.1, consists in testing the equality of two measures projected into a finite number of closed geodesics on $\mathbb{T}^2$. The second, presented in Subsection 3.2, is a conservative procedure based on upper-bounding the exact $p$-values. This is possible thanks to a concentration inequality given in Theorem 3.7, together with faster convergence rates for the expectation.
- Section 4 reports numerical experiments illustrating the relevance of these theoretical results, first with synthetic data and then with real data from protein structures, showing that our methods behave well in both cases.

To facilitate reading, the proofs are relegated to the Appendix, but in some cases the intuition behind the proof is provided in the main text for clarity.

## 2. Optimal transport in $\mathbb{R}^d/\mathbb{Z}^d$

Let $\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$ be defined as the quotient space derived from the equivalence relation $\mathbf{x}\mathcal{R}\mathbf{y}$ if $\mathbf{x} - \mathbf{y} \in \mathbb{Z}^d$. For each $\mathbf{x} \in \mathbb{R}^d$ we denote as $\bar{\mathbf{x}} \in \mathbb{T}^d$ its equivalence class and reserve the notation $\tau$ for the canonical projection map $\mathbf{x} \mapsto \tau(\mathbf{x}) = \bar{\mathbf{x}}$. The topology of the quotient space is defined as the finest one that makes $\tau$ continuous. With this topology, the space $\mathbb{T}^d$ is a Polish space with the distance derived from the Euclidean norm $\|\cdot\|$,

$$d(\bar{\mathbf{x}}, \bar{\mathbf{y}}) := \inf_{\mathbf{p} \in \mathbb{Z}^d} \|\mathbf{x} - \mathbf{y} + \mathbf{p}\|. \tag{2.1}$$

Note that the last claim is true since the projection map $\tau$ is in fact a metric identification, $(\mathbb{R}^d, \|\cdot\|)$ is a Banach space and $\mathbb{Z}^d$ is a closed subset, then it is complete, metrizable through $d$ and separable.

Set $p > 1$. For two probability measures $P, Q \in \mathcal{P}(\mathbb{T}^d)$, a probability measure $\pi \in \mathcal{P}(\mathbb{T}^d \times \mathbb{T}^d)$ is said to be an *optimal transport plan for the cost $d^p$* between $P$ and $Q$ if it solves

$$\mathcal{T}_p(P, Q) := \inf_{\gamma \in \Pi(P,Q)} \int_{\mathbb{T}^d \times \mathbb{T}^d} d^p(\bar{\mathbf{x}}, \bar{\mathbf{y}}) d\gamma(\bar{\mathbf{x}}, \bar{\mathbf{y}}), \tag{2.2}$$

where $\Pi(P, Q)$ is the set of probability measures $\gamma \in \mathcal{P}(\mathbb{T}^d \times \mathbb{T}^d)$ such that $\gamma(A \times \mathbb{R}^d) = P(A)$ and $\gamma(\mathbb{T}^d \times B) = Q(B)$ for all Borel measurable subsets $A, B$ of $\mathbb{T}^d$.

The Kantorovich problem (2.2) can be formulated in a dual form, as follows

$$\mathcal{T}_p(P,Q) = \sup_{(f,g)\in\Phi_p(P,Q)} \int_{\mathbb{T}^d} f(\bar{\mathbf{x}})dP(\bar{\mathbf{x}}) + \int_{\mathbb{T}^d} g(\bar{\mathbf{y}})dQ(\bar{\mathbf{y}}), \qquad (2.3)$$

where

$$\Phi_p(P,Q) = \{(f,g) \in L_1(P) \times L_1(Q): \ f(\bar{\mathbf{x}}) + g(\bar{\mathbf{y}}) \le d^p(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \quad \forall\, \bar{\mathbf{x}}, \bar{\mathbf{y}} \in \mathbb{T}^d\}.$$

The element $\psi \in L_1(P)$ is said to be an *optimal transport potential from $P$ to $Q$ for the cost $d^p$* if there exists $\varphi \in L_1(Q)$ such that the pair $(\psi, \varphi)$ solves (2.3). Recall from [66] that the solutions of (2.3) are pairs $(f,g)$ of $d^p$-conjugate $d^p$-concave functions. This means that

$$f(\mathbf{x}) = \inf_{\bar{\mathbf{y}}\in\mathbb{T}^d}\{d(\bar{\mathbf{x}}, \bar{\mathbf{y}})^p - g(\bar{\mathbf{y}})\} \text{ and } \ g(\bar{\mathbf{y}}) = f^{d^p}(\bar{\mathbf{y}}) = \inf_{\bar{\mathbf{x}}\in\mathbb{T}^d}\{d(\bar{\mathbf{x}}, \bar{\mathbf{y}})^p - f(\bar{\mathbf{x}})\}. \tag{2.4}$$

Furthermore, since $\mathbb{T}^d$ is a Polish space, then Theorem 4.1 in [66] implies that there exists a solution $\pi^*$ of (2.2). Additionally, Theorem 5.10 in [66] establishes that $\mathrm{supp}(\pi^*)$ is $d^p$-*cyclically monotone*. This means that for any finite sequence $\{(\mathbf{x}_k, \mathbf{y}_k)\}_{k=1}^n \subset \mathrm{supp}(\pi^*)$ and any bijection $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$, the following inequality holds:

$$\sum_{k=1}^n d^p(\bar{\mathbf{x}}_k, \bar{\mathbf{y}}_k) \le \sum_{k=1}^n d^p(\bar{\mathbf{x}}_k, \bar{\mathbf{y}}_{\sigma(k)}).$$

Note that, if $Q$ is a probability measure in $\mathbb{T}^d$, its support is defined as the closed set $\mathrm{supp}(Q) \subset \mathbb{T}^d$ composed by $\bar{\mathbf{x}} \in \mathbb{T}^d$ such that for any neighborhood $\mathcal{U}_{\bar{\mathbf{x}}}$ of $\bar{\mathbf{x}}$ it holds that $Q(\mathcal{U}_{\bar{\mathbf{x}}}) > 0$. The interior of the support is denoted by $\mathcal{X}_Q$.

With the same obvious notation we can define a $\|\cdot\|^p$-*cyclically monotone* set. Note that for $p = 2$, $\|\cdot\|^2$-cyclical monotonicity is equivalent to the concept of cyclical monotonicity in convex analysis, described in [56]. Recall that a set $A \subset \mathbb{R}^d \times \mathbb{R}^d$ is *cyclically monotone* if for every finite sequence $\{(\mathbf{x}_k, \mathbf{y}_k)\}_{k=1}^n \subset A$ and every bijection $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ it holds that

$$\sum_{k=1}^n \langle \mathbf{x}_k, \mathbf{y}_k \rangle \ge \sum_{k=1}^n \langle \mathbf{x}_k, \mathbf{y}_{\sigma(k)} \rangle.$$

Consequently, the concept of $d^p$ (resp. $\|\cdot\|^p$) -cyclical monotonicity is the natural generalization, to other spaces and costs, of cyclical monotonicity.

In some cases, that we will study later on, there exists some measurable map $T$ such that the optimal transport plan $\pi$ satisfies $\pi = (I \times T)\#P$, where the symbol $T\#P$ denotes the *push forward measure* of $P$ through $T$, which is defined by $T\#P(A) := P(T^{-1}(A))$, for all measurable $A \subset \mathbb{T}^d$, and $I$ denotes the identity map. Therefore, the problem becomes equivalent to the following *Monge* formulation:

$$\mathcal{T}_p(P,Q) = \inf_{T\#P=Q} \int_{\mathbb{T}^d} d^p(\bar{\mathbf{x}}, T(\bar{\mathbf{x}}))dP(\bar{\mathbf{x}}). \tag{2.5}$$

## 2.1. Existence of $\|\cdot\|^p$-cyclically monotone mappings

A cyclically monotone map is the natural generalization of a non decreasing function in the real line (as being the gradient of a convex function, see [56]). Cyclical monotonicity provides a powerful tool for statistical studies, see [30, 18, 13] among others. The existence of cyclically monotone maps between probability measures in $\mathbb{R}^d$ has been investigated, in parallel, by [15] and [12], with the restrictive assumption of finite second order moment, relaxed in [43]. For periodic measures, the celebrated result of [14] showed the existence. The concept of cyclically monotone map also appears naturally when solving an optimal transport problem with quadratic cost in $\mathbb{R}^d$. Therefore, for any potential cost $\|\cdot\|^p$, the natural generalization is the one of $\|\cdot\|^p$-cyclically monotone. In fact, [28] proved the existence of a $\|\cdot\|^p$-cyclically monotone mapping between probability measures with finite moment of order $p > 1$. To the authors' knowledge, no previous work has dealt with the existence of $\|\cdot\|^p$-cyclically monotone mappings between periodic probability measures. Consequently, the main result of this section is Theorem 2.1, which shows the existence and uniqueness of a $\|\cdot\|^p$-cyclically monotone preserving map $\mathbf{S}_p$ between periodic measures, for $p > 1$, and relates it with the solution of (2.5). Then, Theorem 2.2 guarantees, under certain assumptions of regularity on the support of $P$, that the solution of (2.3) is unique up to an additive constant.

Note that, in practice, a probability $P \in \mathcal{P}(\mathbb{T}^d)$ defines a periodic measure $\mu_P \in \mathcal{M}(\mathbb{R}^d)$ w.r.t. any $\mathbf{p} \in \mathbb{Z}^d$. In other words, $T_{\mathbf{p}}\#\mu_P = \mu_P$, for all $\mathbf{p} \in \mathbb{Z}^d$, where $T_{\mathbf{p}} : \mathbb{R}^d \to \mathbb{R}^d$ is the shift operator $\mathbf{x} \mapsto \mathbf{x} + \mathbf{p}$. A measure $\mu_P$ is *periodic* if it is the natural extension of some probability measure $P \in \mathcal{P}(\mathbb{T}^d)$. As anticipated, the goal of this section is to show the existence of $\|\cdot\|^p$-cyclically monotone mappings between two periodic measures $\mu_P, \mu_Q \in \mathcal{M}(\mathbb{R}^d)$ absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^d$, denoted as $\mu_P, \mu_Q \ll \ell_d$. As commented before, [14] established the existence of a $\|\cdot\|^2$-cyclically monotone map (which a.s. is the gradient of a convex function $\varphi$) such that $\nabla\varphi\#\mu_P = \mu_Q$. Theorem 1.25 in [57] entails that there is a unique solution of the Monge problem in the torus, described by the relation $T = \mathbf{x} - \nabla f(\mathbf{x})$, where the sum is to be intended modulo $\mathbb{Z}^d$ and $f$ is an optimal transport potential for the quadratic cost. Note that this is a quite similar relation (between potentials and transport) to the one in the quadratic transport problem in $\mathbb{R}^d$.

The proof of Theorem 2.1 starts by realizing that since $\mathbb{T}^d$ is a Polish space, then Theorem 4.1 in [66] implies that there exists a solution $\pi^*$ of (2.2). Furthermore, Theorem 5.10 in [66] establishes that $\mathrm{supp}(\pi^*)$ is $d^p$-cyclically monotone, which implies that the set

$$\Gamma = \{(\mathbf{x}+p, \mathbf{y}+p) : (\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in \mathrm{supp}(\pi^*), \ \mathbf{x} \in [0,1]^d, \ d(\bar{\mathbf{x}}, \bar{\mathbf{y}}) = \|\mathbf{x}-\mathbf{y}\| \text{ and } p \in \mathbb{Z}^d\} \tag{2.6}$$

is cyclically monotone. Corollary 3.5 in [28] implies that this cyclically monotone set is contained in the graph of a $\|\cdot\|^p$-differential

$$\partial^{\|\cdot\|^p}\varphi_p(\mathbf{x}) = \{\bar{\mathbf{y}} : \ \varphi_p(\mathbf{z}) \le \varphi_p(\mathbf{x}) + \|\mathbf{z} - \mathbf{y}\|^p - \|\mathbf{x} - \mathbf{y}\|^p, \text{ for all } \mathbf{z} \in \mathbb{R}^d\}$$

of a $\| \cdot \|^p$-concave function $\varphi_p$ (defined as in (2.4) but replacing $d^p$ with $\| \cdot \|^p$). In conclusion, the a.s. uniqueness of this $\| \cdot \|^p$-differential ends the proof.

**Theorem 2.1.** *Let $P, Q \in \mathcal{P}(\mathbb{T}^d)$ be probability measures such that $\mu_P \ll \ell_d$. Then, there exists a unique solution $\mathbf{T}_p$ of (2.5). Moreover, there exists a $\mu_P$-a.e. defined $\| \cdot \|^p$-cyclically monotone map $\mathbf{S}_p$ such that*

- *the relation $\mathbf{T}_p \circ \tau = \tau \circ (\mathbf{S}_p)$ holds $\mu_P$-almost surely,*
- *and $\mathbf{S}_p \# \mu_P = \mu_Q$.*

The following result gives the uniqueness, up to additive constants, of the optimal transport potential, where the assumptions are given with respect to its associated periodic measures. In particular, we need to have *negligible boundary* of $\mu_P$ which means that the boundary of its support has Lebesgue measure 0, $\ell_d(\partial \operatorname{supp}(\mu_P)) = 0$. The proof investigates the intrinsic relation between the optimal transport potentials and the previously described $\mathbf{T}_p$, which allows the use of general results for the uniqueness of $\| \cdot \|^p$-concave functions (see [19]) which have the same gradient a.s. in a connected domain of $\mathbb{R}^d$.

**Theorem 2.2.** *Let $P, Q \in \mathcal{P}(\mathbb{T}^d)$ be probability measures with connected support such that their associated periodic measures satisfy $\mu_P, \mu_Q \ll \ell_d$ with negligible boundary. Then, there exists a unique, up to an additive constant, $d^p$-concave function $f_p$ solution of (2.3).*

The assumption of connected support can be relaxed, via [62, Theorem 2], to the setting where both measures have disconnected support. If the supports of $\mu_P$ and $\mu_Q$ decompose into closures of connected open components

$$\operatorname{supp}(\mu_P) = \bigcup_{i \in \mathcal{I}} \mathcal{X}_{i, \mu_P}, \quad \operatorname{supp}(\mu_Q) = \bigcup_{j \in \mathcal{J}} \mathcal{X}_{j, \mu_Q}, \tag{2.7}$$

where $\mathcal{I}$ is finite index set and $\mathcal{J}$ is a countable index set, then, assuming for all non-empty proper $\mathcal{I}' \subset \mathcal{I}$ and $\mathcal{J}' \subset J$ that

$$\sum_{i \in \mathcal{I}'} \mu_P(\mathcal{X}_{i, \mu_P}) \neq \sum_{j \in \mathcal{J}'} \mu_Q(\mathcal{X}_{j, \mu_Q}), \tag{2.8}$$

it follows by [62, Lemma 5] that no degenerate transport plan exists. Hence, invoking Theorem 2 in [62] in conjunction with Theorem 2.2, yields an extension of the uniqueness result to measures with disconnected support.

**Corollary 2.3.** *Let $P, Q \in \mathcal{P}(\mathbb{T}^d)$ be probability measures such that their associated periodic measures satisfy $\mu_P, \mu_Q \ll \ell_d$ with negligible boundary where (2.7) and (2.8) hold. Then, there exists a unique, up to an additive constant, $d^p$-concave function $f_p$ solution of (2.3).*

The importance of Corollary 2.3 mainly lies in that it enables the study of the asymptotic behavior of the potential, allowing us to apply Arzelá-Ascoli like reasoning, as explained in the following section.

## 2.2. Asymptotic behaviour

This section deals with the asymptotic properties of the transport map and potentials. We consider two sequences of probability measures $\{\alpha_n\}_{n\in\mathbb{N}}, \{\beta_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(\mathbb{T}^d)$ converging weakly to $P$ and $Q$ respectively,

$$\alpha_n \xrightarrow{w} P \ \text{ and } \ \beta_n \xrightarrow{w} Q.$$

Since $\mathbb{T}^d$ is compact, here the weak convergence is in the sense that for every continuous function $h \in \mathcal{C}(\mathbb{T}^d)$, $\int h(\bar{\mathbf{x}})d\alpha_n(\bar{\mathbf{x}}) \to \int h(\bar{\mathbf{x}})dP(\bar{\mathbf{x}})$. Once again, thanks to that compactness the existence of moments of any order is always fulfilled for $P \in \mathcal{P}(\mathbb{T}^d)$. As a consequence, Theorem 7.12 in [65] implies that $\alpha_n \xrightarrow{w} P$ if and only if the *p-Wasserstein distance* $\mathcal{W}_p(\alpha_n, P) := (\mathcal{T}_p(\alpha_n, P))^{\frac{1}{p}}$ tends to 0. An analogous reasoning implies the convergence $\mathcal{T}_p(\alpha_n, \beta_n) \to \mathcal{T}_p(P, Q)$ for the two-sample case.

The idea of this section is to take advantage of the fact that any $d^p$-concave function $f$ is continuous whereby it is finite. Moreover, it has bounded continuity modulus, so we can apply Arzelá-Ascoli's Theorem by fixing the constants.

**Lemma 2.4.** *Every $d^p$-concave function $f$ is Lipschitz (in its definition domain* $\mathrm{dom}(f)$) *with constant $L = 2\,p\,d^{\frac{p-1}{2}}$, with respect to the metric* (2.1).

The proof of the next Theorem first proceeds by choosing the sequence $\{a_n\}_{n\in\mathbb{N}}$ to guarantee the uniform boundedness of the sequence $\{(f_n, g_n)\}_{n\in\mathbb{N}}$ of solutions of (2.3). This, together with Lemma 2.4 and Arzelá-Ascoli's Theorem, implies that $\{(f_n, g_n)\}_{n\in\mathbb{N}}$ is relatively compact. The uniqueness of solutions of (2.3), described in Theorem 2.2, allows us to conclude.

**Theorem 2.5.** *Let $P, Q \in \mathcal{P}(\mathbb{T}^d)$ be probability measures with connected supports whose associated periodic measures satisfy $\mu_P, \mu_Q \ll \ell_d$ with negligible boundary. Let $\{\alpha_n\}_{n\in\mathbb{N}}$ and $\{\beta_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(\mathbb{T}^d)$ be two sequences of probability measures converging weakly to $P$ and $Q$ respectively. Denote by $(f_n, g_n)$ (resp. $(f, g)$) the solution of the dual problem between $\alpha_n$ and $\beta_n$ (resp. $P$ and $Q$). Then there exists a sequence of real numbers $\{a_n\}_{n\in\mathbb{N}}$ such that $f_n + a_n \to f$ uniformly on the compact sets of $\mathcal{X}_P$.*

## 2.3. Asymptotic normality

This section is devoted a proof of a Central Limit Theorem (CLT) for the fluctuations of the empirical optimal transport cost. Recall that the previous section proves that, under certain regularity assumptions, there exists a unique optimal transport potential from $P$ to $Q$. Let $f_p$ be such a potential. We will use Efron-Stein's inequality to derive that

$$\sqrt{n}\,(\mathcal{T}_p(P_n, Q) - \mathbb{E}\mathcal{T}_p(P_n, Q)) \xrightarrow{w} N(0, \sigma_p^2(P, Q)),$$

with

$$\sigma_p^2(P, Q) = \mathrm{Var}(f_p(X)). \tag{2.9}$$

Then, we will see that the same holds in the two sample case. The idea is not new: it has already been used with the same goal in [22] for the quadratic cost in $\mathbb{R}^d$, and in its extension to general costs in [19]. Moreover, when using regularized optimal transport, [45] showed that the same technique can be applied. A similar result, but using the idea in [20] of differentiating the supremum in the functional sense by applying the general result of [16], yields also a CLT on the torus for $p \geq 2$, see [34].

**Theorem 2.6.** *Let $P, Q \in \mathcal{P}(\mathbb{T}^d)$ be probability measures with connected supports such that their associated periodic measures satisfy $\mu_P, \mu_Q \ll \ell_d$ with negligible boundary. Then, for any $p > 1$, we have*

$$\sqrt{n}\left(\mathcal{T}_p(P_n, Q) - \mathbb{E}\mathcal{T}_p(P_n, Q)\right) \xrightarrow{w} N(0, \sigma_p^2(P, Q)),$$

*and, if $m = m(n)$ satisfies that $m \longrightarrow +\infty$ and $\frac{n}{n+m} \to \lambda \in (0,1)$ as $n \to \infty$,*

$$\sqrt{\tfrac{nm}{n+m}}\left(\mathcal{T}_p(P_n, Q_m) - \mathbb{E}\mathcal{T}_p(P_n, Q_m)\right) \xrightarrow{w} N\left(0, (1-\lambda)\sigma_p^2(P, Q) + \lambda\sigma_p^2(Q, P)\right),$$

*where $\sigma_p^2(P, Q)$ and $\sigma_p^2(Q, P)$ are defined in (2.9) and satisfy*

$$\sqrt{\tfrac{nm}{n+m}}\mathrm{Var}(\mathcal{T}_p(P_n, Q_m)) \longrightarrow (1-\lambda)\sigma_p^2(P, Q) + \lambda\sigma_p^2(Q, P). \tag{2.10}$$

It is clear that the limit of Theorem 2.6 degenerates to 0 when $P = Q$. Suppose now that $P \neq Q$ are satisfying the assumption of Theorem 2.6. The limit, in the one-sample case, is degenerate if and only if $\mathrm{Var}(f_p(X)) = 0$. Since the optimal transport potentials are unique up to additive constants, see Theorem 2.2, we can suppose that $E(f_p(X)) = 0$. Thus, the degeneracy is equivalent to $E(f_p(X)^2) = 0$, hence $f_p = 0$ $P$-a.s. and the same holds for $f_p^{d^p}$. This implies, in particular, that $\mathcal{W}_p(P, Q) = 0$ which occurs only if $P = Q$.

Our initial motivation to prove Theorem 2.6 was to find an asymptotic distribution of $\mathcal{T}_p(P_n, Q_m)$ allowing the definition of a two-sample goodness-of-fit test. Even for measures supported on the real line, the only asymptotic results account for the case $P \neq Q$, providing the asymptotic behaviour of the Wasserstein statistic under the alternative hypothesis. The idea of switching $H_0$ and $H_1$ and testing for similarities has been studied in several previous works, all considering measures supported on $\mathbb{R}$. Gaussian deviations from the true distance $\mathcal{T}_2(P, Q)$ are proved in [21], which allows testing of $\mathcal{T}_2(P, Q) \geq \Delta_0$, for a given threshold $\Delta_0$. In the same way, the earlier work [27] introduced such an asymptotic test for assessing similarities based on the trimmed Wasserstein distance, allowing sample dependency.

Unfortunately, the same strategy can not be applied in our case, as the derived CLT for measures supported on $\mathbb{T}^2$ (Theorem 2.6) only states Gaussian deviations from the mean. Indeed, if we use (2.10), we could consider the statistic

$$\frac{\mathcal{T}_p(P_n, Q_m) - \mathbb{E}\mathcal{T}_p(P_n, Q_m)}{\sqrt{\mathrm{Var}(\mathcal{T}_p(P_n, Q_m))}} \xrightarrow[P \neq Q]{w} N(0, 1), \tag{2.11}$$

where, in practice, the variance and expectation could be estimated by bootstrapping the given samples (as long as bootstrap consistency is ensured). The recent works of [33] and [39] show that, in small dimension –$d = 2, 3$ and at most 4–, the value $\mathbb{E}\mathcal{T}_p(P_n, Q_m)$ can be substituted by the population $\mathcal{T}_p(P, Q)$. That gives rise to

$$\frac{\mathcal{T}_p(P_n, Q_m) - \mathcal{T}_p(P, Q)}{\sqrt{\mathrm{Var}(\mathcal{T}_p(P_n, Q_m))}} \xrightarrow[P \neq Q]{w} N(0, 1), \tag{2.12}$$

see [33, Example 5.7] for general $p$ or [39, Corollary 8] for $p = 2$. However, for dimension $d > 4$ and $p = 2$, this substitution is no longer valid [40, Proof of Proposition 21].

When $P \neq Q$, the statistics in (2.11), (2.12) converge in law to a standard Gaussian distribution. This is illustrated in Figure 8. However, one would expect the statistic to be stochastically larger under $P \neq Q$ than under $P = Q$, allowing the distinction of the null and the alternative hypotheses. Nevertheless, due to the aforementioned degeneracy of Theorem 2.6 when $P = Q$, this condition fails to be satisfied and no asymptotic test can be implemented from this result. Further discussion about this issue can be found in Section 5. Therefore, the rest of this paper is devoted to alternative approaches to define suitable two-sample goodness-of-fit tests for measures supported on $\mathbb{T}^2$.

## 3. Two-sample goodness-of-fit tests

Let us first formulate the problem. Denote by $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_m)$ two independent and identically distributed random samples of laws $P, Q \in \mathcal{P}(\mathbb{T}^2)$ respectively, and by $P_n, Q_m$ their corresponding empirical probability measures. We aim to test

$$H_0 : P = Q \qquad \text{against} \qquad H_1 : P \neq Q \tag{3.1}$$

via the the definition of a statistic $T_{nm} = T(P_n, Q_m)$, representing an estimate of discrepancy between $P_n$ and $Q_m$, together with the critical region

$$R = \{(x_1, \ldots, x_n; y_1, \ldots, y_m) : T_{nm} \geq c_{nm}(\alpha)\}, \tag{3.2}$$

where $x_i$ (resp. $y_j$) denotes a realization of $X_i$ (resp. $Y_j$) for $i = 1, \ldots, n$ (resp. $j = 1, \ldots, m$). The critical value $c_{nm}(\alpha)$ in (3.2) is given for a fixed significance level $\alpha$ by

$$c_{nm}(\alpha) = \inf\{t > 0 : F_{nm}(t) \geq 1 - \alpha\}, \tag{3.3}$$

where $F_{nm}$ is the distribution function of the statistic $T_{nm}$ under $H_0$. We are therefore considering the test

$$\pi_{nm} = \left\{ \begin{array}{ll} 1 & \text{if} \qquad T_{nm} \geq c_{nm}(\alpha) \\ 0 & \text{otherwise} \end{array} \right. \tag{3.4}$$

Equivalently, a $p$-value for this test is $p_{nm} = 1 - F_{nm}(T_{nm})$. Ideally, we would like $T_{nm}$ to be $\mathcal{T}_p(P_n, Q_m)$. However, knowing the distribution of the latter statistic

under $H_0$ remains an open problem. The one-sample case in $\mathbb{R}^d$ has recently been addressed in [30], but approaches for two-sample testing in arbitrary dimension, and for measures on more general spaces, have not already been proposed to the best of our knowledge. The lack of solutions may be explained by the intrinsic difficulty of characterizing the distribution of $\mathcal{T}_p(P_n, Q_m)$ when $P = Q$ especially when the dimension is larger than one. In the next subsections, we propose two alternative approaches to define (3.4), both based on the 2-Wasserstein distance, that allow two-sample goodness-of-fit testing for measures on $\mathbb{T}^2$.

## *3.1. Geodesic projections into* $\mathbb{R}/\mathbb{Z}$

Our first approach for testing the equality of two measures $P$, $Q$ on $\mathbb{R}^2/\mathbb{Z}^2$ is to test the equality of their geodesic projections. This bypasses the dimension problem and allows the implementation of testing techniques based on Wasserstein distance for one-dimensional spaces. Geodesics on $\mathbb{T}^2$ are the images by the canonical projection $\tau$ of straight lines on $\mathbb{R}^2$ [9]. Lines with irrational slope map to geodesics which are dense on $\mathbb{T}^2$, and only lines with rational slope map to *closed* geodesics on the torus, which are closed spirals isomorphic to $\mathbb{R}/\mathbb{Z}$ (see [9, Figure VII.10] for an illustration).

The strategy is to project $P_n$ and $Q_m$ into $N_g$ closed geodesics, and to test the equality of each pair of projected measures, which will be supported on $\mathbb{R}/\mathbb{Z}$. These geodesics can be chosen a priori by the practitioner, or sampled from the set of all closed geodesics on $\mathbb{T}^2$. We propose a sampling method in Appendix A.1. This method prioritizes simpler geodesics (that is, with a smaller number of revolutions over the torus) in order to ease computational implementation. The algorithm we used to project samples on $\mathbb{T}^2$ to a given geodesic is described in Appendix A.2. To avoid repetition of the same test, and to ensure independence between the computed $p$-values, we require all the $N_g$ geodesics to be different.

In this section, we propose a two-sample Wasserstein test to assess the equality of two measures supported on the circle, and state how to combine the resulting $N_g$-tuple of $p$-values into a global $p$-value for the bi-dimensional problem. From now on, to simplify notation, we will denote by $\mathcal{T}_2$ any squared Wasserstein distance, the ground space being inferred from the corresponding measures.

### *3.1.1. Two-sample goodness-of-fit test on* $\mathbb{R}/\mathbb{Z}$

Optimal Transport on the circle has been recently studied in detail in [32], where the limit laws of the one and two-sample empirical Wasserstein distance for measures on $\mathbb{R}/\mathbb{Z}$ are derived. However, the considered statistics are not distribution-free, so that only one-sample goodness-of-fit tests can be derived from these results. Still, the authors of [32] also propose a $b$-out-of-$n$ bootstrap approach, for $b = o(n)$, to define a two-sample goodness-of-fit test. Unfortunately, type I error fails to be controlled since the bootstrapped $p$-value under

the null hypothesis is (substantially) stochastically smaller than a uniform random variable. This can be observed by simple numerical experiments based on the implementation proposed by [32], for example by comparing two equally-sized samples from a Uniform distribution. We believe that this is due to a lack of consistency of the two-sample bootstrap for the proposed statistic. In order to bypass this issue, we now propose a convenient alternative approach based on a distribution-free two-sample statistic.

Let $P^c, Q^c \in \mathcal{P}(\mathbb{R}/\mathbb{Z})$ and $P_n^c, Q_m^c$ be their corresponding empirical probability measures. We aim to test

$$H_0 : P^c = Q^c \qquad \text{against} \qquad H_1 : P^c \neq Q^c.$$

If $\mathbb{R}/\mathbb{Z}$ is parameterized by the set $[0, 1)$ with the geodesic distance

$$d_{\mathbb{R}/\mathbb{Z}}(x, y) = \min\{|x - y|, 1 - |x - y|\},$$

the cumulative distribution functions of $P^c, Q^c$, denoted as $F$, $G$ respectively, can be defined as in [32] as

$$F(t) = P^c([0, t]), \quad G(t) = Q^c([0, t]) \qquad \forall\, t \in [0, 1). \tag{3.5}$$

Then, we can write

$$\mathcal{T}_2(P^c, Q^c) = \inf_{\alpha \in \mathbb{R}} \int_0^1 \left( F^{-1}(t) - (G - \alpha)^{-1}(t) \right)^2 \, dt, \tag{3.6}$$

where the pseudo-inverse is defined as $H^{-1}(s) = \inf\{t \,:\, H(t) > s\}$, for any distribution function $H$. The formulation (3.6) was first proved in [52] for discrete measures, and extended to arbitrary measures in [23]. It shows how the Optimal Transport problem on the circle reduces to the same problem on $[0, 1) \subset \mathbb{R}$ if both measures are relocated on the real line choosing as origin the minimizing element $\alpha$. This is well illustrated in [32]. We first remark that if one of the two measures is the uniform law on $\mathbb{R}/\mathbb{Z}$, the infimum on (3.6) has an explicit formulation.

**Lemma 3.1.** *Let $P^c \in \mathcal{P}(\mathbb{R}/\mathbb{Z})$, and $F$ be its cumulative distribution function. Let $U$ be the uniform distribution on $\mathbb{R}/\mathbb{Z}$. Then,*

$$\mathcal{T}_2(P^c, U) = \int_0^1 \left( F^{-1}(t) - t - \alpha_0(F) \right)^2 \, dt,$$

*where the optimal origin is given by*

$$\alpha_0(F) = \int_0^1 (F^{-1}(t) - t) \, dt.$$

If we replace $P^c$ and $F$ by their empirical counterparts, $P_n^c$ and $F_n$, Lemma 3.1 allows the definition of the statistic $\mathcal{T}_2(P_n^c, U)$, which is distribution-free when $P^c = U$.

**Lemma 3.2.** *Let $P^c \in \mathcal{P}(\mathbb{R}/\mathbb{Z})$, $P_n^c$ be its empirical probability measure, and $U$ be the uniform distribution on $\mathbb{R}/\mathbb{Z}$. Then, if $P^c = U$,*

$$n\,\mathcal{T}_2(P_n^c, U) \xrightarrow[n]{w} \int_0^1 \mathbb{B}(t)^2\, dt - \left(\int_0^1 \mathbb{B}(t)\, dt\right)^2,$$

*where $\mathbb{B}$ is a standard Brownian bridge, and the weak convergence is understood as convergence of probability measures on the space of right-continuous functions with left limits.*

Lemma 3.2 can be used to define of a one-sample goodness-of-fit uniformity test, based on the squared Wasserstein distance on the circle. This would complement the work in [32], where such a test was introduced for the 1-Wasserstein distance. As our aim here is to define a two-sample test, we adapt the idea of [54] to compare two measures on the circle, by considering the 2-Wasserstein distance between $G_m^{-1}(F_n)$ and the uniform distribution. We can therefore consider the statistic

$$T_{nm}^c = \frac{nm}{n+m}\mathcal{T}_2(G_m \# P_n^c, U) = \qquad (3.7)$$

$$\frac{nm}{n+m}\int_0^1 \left(G_m(F_n^{-1}(t)) - t - \alpha_0(F_n^{-1}(G_m))\right)^2 dt,$$

which is also distribution-free when $P^c = Q^c$. The following result is the counterpart of Lemma 3.2.

**Proposition 3.3.** *Let $P^c, Q^c \in \mathcal{P}(\mathbb{R}/\mathbb{Z})$, having continuous and strictly increasing cumulative distribution functions. Let $P_n^c, Q_m^c$ be their corresponding empirical probability measures, and $F_n, G_m$ be their empirical cumulative distribution functions. If $\frac{n}{m} \to \lambda$ when $n, m \to \infty$ for some $\lambda \in [0, \infty)$ then, under $P^c = Q^c$, it holds that*

$$T_{nm}^c = \frac{nm}{n+m}\mathcal{T}_2(G_m \# P_n^c, U) \xrightarrow[n,m]{w} \int_0^1 \mathbb{B}(t)^2\, dt - \left(\int_0^1 \mathbb{B}(t)\, dt\right)^2.$$

Consequently, with the notation of the beginning of Section 3, we propose the test

$$\pi_{nm}^c = \begin{cases} 1 & \text{if} & T_{nm}^c \geq c_{nm}^c(\alpha) \\ 0 & \text{otherwise} \end{cases} \qquad (3.8)$$

where the critical value $c_{nm}^c(\alpha)$ is given by

$$c_{nm}^c(\alpha) = \inf\left\{t > 0 \,:\, F_{nm}^c(t) \geq 1 - \alpha\right\},$$

with $F_{nm}^c$ denoting the distribution function of $T_{nm}^c$ under $H_0$. Equivalently, a $p$-value for this test is $p_{nm}^c = 1 - F_{nm}^c(T_{nm}^c)$. Following Proposition 3.3, the critical value or, equivalently, the $p$-value for a given sample, can be approximated with arbitrary precision using a Monte Carlo algorithm. The following result guarantees the consistency of (3.8).

**Proposition 3.4** (Consistency). *Let $P^c, Q^c \in \mathcal{P}(\mathbb{R}/\mathbb{Z})$ having continuous and strictly increasing cumulative distribution functions. If $P^c \neq Q^c$, it holds*

$$\lim_{n,m \to \infty} \mathbb{P}\left(\pi_{nm}^c = 1\right) = 1 \quad \text{for any } \alpha > 0.$$

*3.1.2. Combining a $N_g$-tuple of tests on $\mathbb{R}/\mathbb{Z}$*

Consider the problem of testing the equality of $N_g$ pairs of projections of $P_n$ and $Q_m$ into $N_g$ different closed geodesics. Instead of a single statistic, we now have a sample $(T_{nm,1}^c, \ldots, T_{nm,N_g}^c)$ of statistics which, under the null hypothesis, are identically distributed as $T_{nm}^c$ (by Proposition 3.3). Equivalently, one can think of a sample of $p$-values $(p_1, \ldots, p_{N_g})$ which, following (3.8), are given by

$$p_i = 1 - F_{nm}^c(T_{nm,i}^c) \quad i = 1, \ldots, N_g. \tag{3.9}$$

These individual $p$-values can be aggregated as follows:

$$p^{N_g} = N_g \min_{i=1}^{N_g} p_i. \tag{3.10}$$

This aggregation is akin to the Bonferroni correction for Family Wise Error Rate (FWER) control in multiple testing [6]. As such, $p^{N_g}$ defined in (3.10) is a valid $p$-value for the two-dimensional test, regardless of the possible dependencies between the $N_g$ individual $p$-values. This implies that the two-dimensional test

$$\pi_{nm,N_g}^g = \left\{ \begin{array}{cc} 1 & \text{if} \quad p^{N_g} \leq \alpha \\ 0 & \text{otherwise} \end{array} \right. \qquad (N_g\text{-geod})$$

controls the type I error for any $\alpha > 0$ (see Appendix B.2 for a proof). Regarding consistency under fixed alternatives, by construction, ($N_g$-geod) will fail to detect differences between two measures on $\mathbb{T}^2$ whose projected distributions are identical for all the $N_g$ geodesics considered. Therefore, $\pi_{nm,N_g}^g$ will not be consistent under such alternatives, which are arguably very unlikely in practice if $N_g$ is large enough. Otherwise, consistency is guaranteed.

**Proposition 3.5** (Consistency). *Let $P, Q \in \mathcal{P}(\mathbb{T}^2)$ such that $\mu_P, \mu_Q \ll \ell_2$ and $P_i^c$ (resp. $Q_i^c$), $i = 1, \ldots, N_g$, be the circular projected distributions of $P$ (resp. $Q$) to $N_g$ closed geodesics of $\mathbb{T}^2$. If $P_i^c \neq Q_i^c$ for at least one $i \in \{1, \ldots, N_g\}$, it holds*

$$\lim_{n,m \to \infty} \mathbb{P}\left(\pi_{nm,N_g}^g = 1\right) \quad \text{for any } \alpha > 0.$$

**Remark 3.6.** *The assumption in Proposition 3.3 that the projected measure $P^c \in \mathcal{P}(\mathbb{R}/\mathbb{Z})$ has continuous and strictly increasing cumulative distribution function is satisfied if the underlying measure $P \in \mathcal{P}(\mathbb{T}^2)$ satisfies $\mu_P \ll \ell_2$. See Appendix B.2 for a proof.*

The time complexity of ($N_g$-geod) is $\mathcal{O}(n + m)$. Indeed, $n + m$ operations are needed to compute $G_m(F_n^{-1}(t))$ and $F_n^{-1}(G_m(t))$ for a given $t$. Therefore, computing the test statistic (3.7) can be done in $\mathcal{O}(n + m)$ operations, where the complexity constant depends on the number of subdivisions of $[0, 1]$ set by the numerical integration method chosen to compute (3.7). Moreover, the time complexity of the algorithm described in Appendix A.1 to sample closed geodesics is also $\mathcal{O}(n + m)$ in practice, as a consequence of the distribution from which the geodesics are drawn. This is empirically illustrated in Figure 9.

### *3.2. p-value upper bounding*

If we set $\mathcal{T}_2(P_n, Q_m)$ as the statistic $T_{nm}$ for the test (3.4), the p-value for a given sample would be given by

$$\mathbb{P}_{H_0}(\mathcal{T}_2(P_n, Q_m) \geq t_{nm}), \tag{3.11}$$

where $t_{nm}$ denotes the statistic realization. The goal of this section is to find an upper bound for (3.11), which will itself be a valid $p$-value for (3.4) if it controls type I error (that is, if it remains with probability $1 - \alpha$ over a fixed significance level $\alpha$ under $H_0$). We will also require the power of the corresponding test to tend to 1 under fixed alternatives. We start by upper bounding the deviations of the statistic from the mean. Using McDiarmid's inequality [44], we obtain the following result, which extends to the two-sample case the inequality in [67, Proposition 20], for the quadratic cost.

**Theorem 3.7.** *Let $P, Q \in \mathcal{P}(\mathbb{T}^2)$ and $P_n, Q_m$ be two empirical probability measures of laws $P$, $Q$ respectively. Then, for all $t \in \mathbb{R}$, we have*

$$\mathbb{P}\left(\mathcal{T}_2(P_n, Q_m) - \mathbb{E}\mathcal{T}_2(P_n, Q_m) > t\right) \leq \exp\left(-\frac{nm}{n + m}8t^2\right). \tag{3.12}$$

After that, we study the convergence speed of the expectation under the null hypothesis. Using directly the results exposed in [26], only bounds of order

$$\mathbb{E}\mathcal{T}_2(P_n, Q_m) = O\left(n^{-\frac{1}{2}} + m^{-\frac{1}{2}}\right) \tag{3.13}$$

can be expected. However, the recent work in [1] shows that the convergence of the mean (3.13) becomes faster under some regularity assumptions. On the one hand, we require the density of the induced periodic measure $\mu_P$ to be Hölder continuous[1] and absolutely continuous w.r.t. the Lebesgue measure $\ell_2$ in $\mathbb{R}^2$. On the other hand, we require the set $\text{supp}(P)$ to be connected and to have $\mathcal{C}^1$ boundary, in the sense that it can be locally parameterized by a $\mathcal{C}^1$ curve.

**Assumption 1.** *(1) $P \in \mathcal{P}(\mathbb{T}^2)$ is supported in a connected set with $\mathcal{C}^1$ boundary, with $\mu_P \ll \ell_2$. (2) Its probability density $p$ is Hölder continuous and bounded from below in its support ($p(x) \geq \lambda > 0$ for all $x \in \text{supp}(\mu_P)$).*

---

[1]A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be locally Hölder continuous in a compact set $X$ for some $\alpha > 0$ if, for every $x \in X$, there exists some $\epsilon > 0$ such that $|f(x) - f(y)| \leq C\|x - y\|^\alpha$ if $y \in X$ and $\|y - x\| < \epsilon$.

If Assumption 1 is satisfied, then from Lemma B.1 and Theorem 6.3. in [1] we can derive the following asymptotic bound for the two-sample null expectation.

**Lemma 3.8.** *Let* $P = Q \in \mathcal{P}(\mathbb{T}^2)$ *satisfy Assumption 1 and* $m = m(n)$ *be a sequence such that* $m \xrightarrow[n \to \infty]{} \infty$ *and* $\frac{n}{m} \to \lambda \in (0, 1)$. *Then, we have*

$$\limsup_{n \to \infty} \frac{n}{\log(n)} \, \mathbb{E}\mathcal{T}_2(P_n, Q_m) \leq \frac{1}{4\pi} \left( 1 + \frac{1}{\lambda} \right). \tag{3.14}$$

Note that Assumption 1 is not especially restrictive. It is satisfied by any continuously differentiable density, whose connected support can be locally given by the graph of a continuously differentiable function. Examples include bivariate von Mises distributions or uniform distributions in connected smooth sets.

The idea to define the test is to combine Theorem 3.7 with Lemma 3.8 and upper bound (3.11) for sufficiently large sample sizes. If we take the limit for the expectation in (3.12) under the null, we have the following result.

**Proposition 3.9.** *Let* $P, Q \in \mathcal{P}(\mathbb{T}^2)$ *and* $P_n, Q_m$ *be two empirical probability measures of laws* $P, Q$ *respectively. For all* $\varepsilon > 0$, *there exists* $N_\varepsilon \in \mathbb{N}$ *such that for all* $n, m \geq N_\varepsilon$, *we have*

$$\mathbb{P}_{H_0} \left( \mathcal{T}_2(P_n, Q_m) > t \right) \leq \exp \left( -\frac{nm}{n+m} 8(t - \varepsilon)^2 \right) =: \xi_{nm,\varepsilon}(t) \qquad \forall \, t > 0. \tag{3.15}$$

For a fixed $\varepsilon > 0$, the bound (3.15) can be used to define a test (3.4) for any $\alpha > 0$ as follows:

$$\pi_{nm,\varepsilon}^{ub} = \left\{ \begin{array}{ll} 1 & \text{if} \quad \xi_{nm,\varepsilon} \left( \mathcal{T}_2(P_n, Q_m) \right) \leq \alpha \\ 0 & \text{otherwise} \end{array} \right. \tag{UB}$$

By Proposition 3.9, the test (UB) will control type I error for all $n, m \geq N_\epsilon$. In practice, the threshold $N_\varepsilon$ depends on the unspecified constant hidden in (3.14), which is dragged from the results in [1]. The following result shows that, nevertheless, asymptotic consistency at level $\alpha$ of (UB) is guaranteed.

**Proposition 3.10** (Asymptotic consistency at level $\alpha$). *Let* $P, Q \in \mathcal{P}(\mathbb{Z}^2)$. *The test* (UB) *is asymptotically of level* $\alpha$. *If* $P = Q$, *we have, for any* $\varepsilon > 0$,

$$\lim_{n,m \to \infty} \mathbb{P} \left( \pi_{nm,\varepsilon}^{ub} = 1 \right) \leq \alpha \quad \text{for any } \alpha > 0. \tag{3.16}$$

*Under fixed alternatives, the test is consistent if* $\mathcal{T}_2(P, Q) > \varepsilon$:

$$\lim_{n,m \to \infty} \mathbb{P} \left( \pi_{nm,\varepsilon}^{ub} = 1 \right) = 1 \quad \text{for any } \alpha > 0.$$

The last result ensures asymptotic consistency at level $\alpha$ if the two compared measures are further than $\varepsilon$ in the squared 2-Wasserstein distance. This can be used to calibrate the sensibility of (UB) if the practitioner possesses some prior information about the differences that the test should accept. This would

ensure smaller $N_\varepsilon$ without implying a power decrease. For the simulation and case studies presented here, we will set $\varepsilon$ to the machine precision $\varepsilon_m = 2.2 \cdot 10^{-16}$ (for a standard double-precision floating-point format). The corresponding $N_\varepsilon$ should be affordable thanks to Lemma 3.8, responsible of the satisfactory power of (UB). Due to the improved convergence speed of the expectation, we will have sharp bounds (3.15) for reasonable sample sizes, allowing the detection of differences for our practical purposes. This is illustrated in Section 4.2.

The computational complexity of (UB) is given by the numerical algorithm solving the Optimal Transport problem. Here, we used the Fast Network Simplex for Optimal Transport [8], which has $\mathcal{O}((n+m)^2)$ time complexity and $\mathcal{O}((n+m)^2)$ memory cost, due to the cost matrix computation.

## 4. Numerical experiments

This section is devoted to assess the performance of the two-sample goodness-of-fit tests ($N_g$-geod) and (UB), and to show how they can be implemented to evaluate differences on protein structure data. In Section 4.1 and 4.2, we evaluate the relative efficiency of both tests, comparing their performance with other methods not based on Optimal Transport. Section 4.3 illustrates one possible application to protein structure investigations, by stating statistical evidence of nearest neighbors effects on local protein conformations.

### *4.1. Small-sample performance*

To make an informative analysis of the performance of tests ($N_g$-geod) and (UB), we studied how their power function behaves for alternatives converging to the null hypothesis. We also assessed whether the proposed approach to define a Wasserstein test on the circle contributes to a better power. In particular, we compared the power function of ($N_g$-geod) with variations of the same test. On the one hand, to evaluate whether the choice of an optimal origin to relocate the measures on $[0,1)$ is advantageous, we considered the same statistic (3.7) but with $\alpha_0$ being random and uniformly chosen in $[0,1]$. It is easy to check that the modified statistic is distribution-free under the null, by proceeding analogously to Proposition 3.3. On the other hand, to study whether the use of Wasserstein distance for the one-dimensional statistic contributes to a better power, we relocated the measures in $[0,1)$ (again after choosing a random origin on $\mathbb{R}/\mathbb{Z}$) and compared them with the well-known Anderson-Darling two-sample statistic. To study the effect of the number $N_g$ of geodesics, we performed the test ($N_g$-geod) for $N_g \in \{2,3,4,5\}$. We also compared all the previous approaches with the two-dimensional extension of the Kolmogorov-Smirnov two-sample test proposed by Fasano and Franceschini [24], defined for measures supported on $\mathbb{R}^2$. This allows the assessment of whether taking into account the geometry of the underlying space contributes to a better performance.

For the small-sample case, we compared samples of size $n = m = 50$ drawn from a bivariate von Mises (bvM) distribution [41] of means $\mu = \nu = 0.5$, and
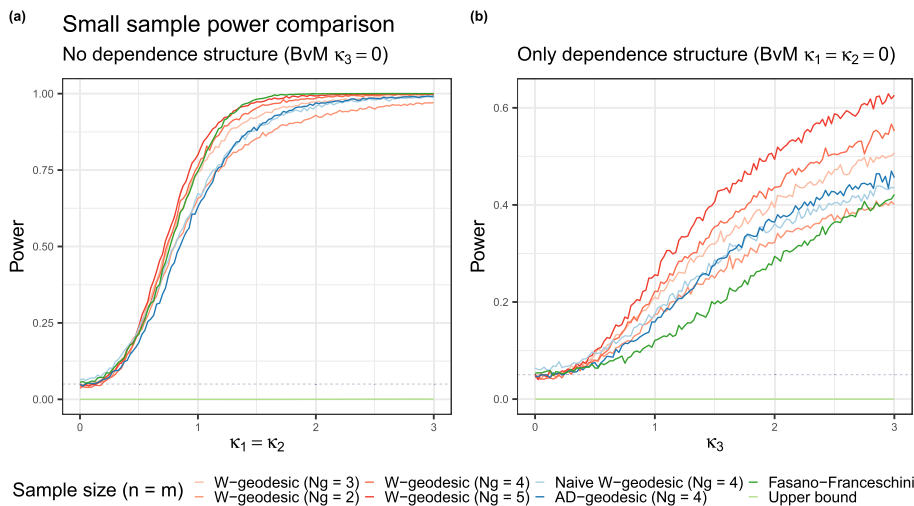
FIG 1. *Empirical power of two-sample goodness-of-fit tests for measures supported on* $\mathbb{T}^2$, *under bivariate von Mises (BvM) alternatives with no dependence structure and different marginal laws (a) and with equal marginal laws and dependence structure (b). The simulated samples had sizes* $n = m = 50$. *The empirical power corresponds to the proportion of rejections at level* $\alpha = 0.05$ *(dashed line) among* 5000 *repetitions of the test for fixed concentration parameters.*

concentration parameters $\kappa_1, \kappa_2, \kappa_3$ with equally-sized samples drawn from a uniform distribution on $\mathbb{T}^2$. The density of the bvM cosine model is given by

$$f(\varphi, \psi) = c(\kappa_1, \kappa_2, \kappa_3)(\exp(\kappa_1 \cos(\varphi - \mu) + \kappa_2 \cos(\phi - \nu) - \\ \kappa_3 \cos(\varphi - \mu - \psi + \nu)),$$

where the explicit form of the normalization constant $c(\kappa_1, \kappa_2, \kappa_3)$ is stated in [41]. The null hypothesis corresponds to the case $\kappa_1 = \kappa_2 = \kappa_3 = 0$. For the converging alternatives, we distinguished two scenarios:

(a) No dependence structure: $\kappa_3 = 0$ and $\kappa_1 = \kappa_2 \in [0, 3]$ as varying parameter.

(b) Only dependence structure: $\kappa_1 = \kappa_2 = 0$ and $\kappa_3 \in [0, 3]$ as varying parameter. Here, the marginal laws are uniform distributions on $[0, 1]$ [41].

The rejection probability was estimated as the proportion of rejections at level $\alpha = 0.05$ among 5000 repetitions of each test for a fixed value of the corresponding varying parameter. Results for both scenarios are shown in Figure 1, where 'W-geodesic' stands for the test ($N_g$-geod), 'Naive W-geodesic' for its random origin variation, 'AD-geodesic' for the comparison with the Anderson-Darling two-sample statistic, and 'Upper bound' for the test (UB).

The first conclusion that we can state after Figure 1 is that the test (UB) has zero power for small sample sizes. This was expected by Proposition 3.9, as large values of $n, m$ are required to ensure sharp bounds. However, some interesting

conclusions can be extracted regarding the other tests. First, the test ($N_g$-geod) has power $\alpha$ under $H_0$. Indeed, further simulations confirmed that the approach described in Section 3.1.2 ensures the uniformity of the combined $p$-value's null distribution. Together with the illustrated consistency of test ($N_g$-geod), we can observe the considerable gain in power when comparing measures with the Wasserstein statistic (3.7) by choosing an optimal origin on the circle. The choice of a random origin ('Naive W-geodesic' curve) or the use of techniques that do not rely on Optimal Transport ('AD-geodesic' or Fasano-Franceschini curve) notably reduce the test power, specially when differences are presented on the dependence structure (Figure 1b). Finally, the choice of the number $N_g$ of geodesics seems to have an effect on power. As one could have expected, increasing the number of geodesic projections improves the test's ability to detect slighter differences. Consequently, the practitioner is entitled to indefinitely increase $N_g$, paying back on computation time (or implementation complexity, if geodesics are randomly chosen, see Appendix A.1).

### 4.2. Asymptotic performance

This section is devoted to assess the suitability of the upper bound testing technique (UB) when large sample sizes are available. Here, we studied the relative efficiency of tests (UB) and ($N_g$-geod) for the same converging alternatives as in Section 4.1, with $n = m \in \{1000, 1500, 2000\}$. Results are shown in Figure 2, where the parameter of interest ($\kappa_1 = \kappa_2$ or $\kappa_3$) took values in $\{0.1, 0.2, \ldots, 4\}$, and the empirical power was estimated as the proportion of rejections at level $\alpha = 0.05$ among 1000 repetitions of each test.

Figure 2 shows that the test (UB) is powerful when sample sizes are large enough. As its corresponding $p$-value has been defined as an upper bound of the actual $p$-value (3.11), it will be quite a conservative test and, therefore, relatively less efficient than ($N_g$-geod). This is illustrated in both panels. In any case, the test (UB) can be useful in practice. Besides the detection of big differences, the practitioner may be interested in the acceptance of small and controlled discrepancies between samples, which may be due, for instance, to experimental inaccuracies. In scenarios where a less conservative method as ($N_g$-geod) may detect such differences, one might prefer to rely on a test method that allows slight dissimilarities and stands out only the more relevant ones. Consequently, even if the test (UB) is clearly less efficient than our first candidate ($N_g$-geod), we believe it can be of interest in some practical scenarios, such as several situations appearing in Structural Biology problems. This is further discussed in Section 5.

### 4.3. Application to protein structure analysis

A method to accurately compare local structural preferences in conformational ensemble models of proteins is useful to investigate sequence-structure-function relationships, allowing for instance to understand the effect of mutations. The
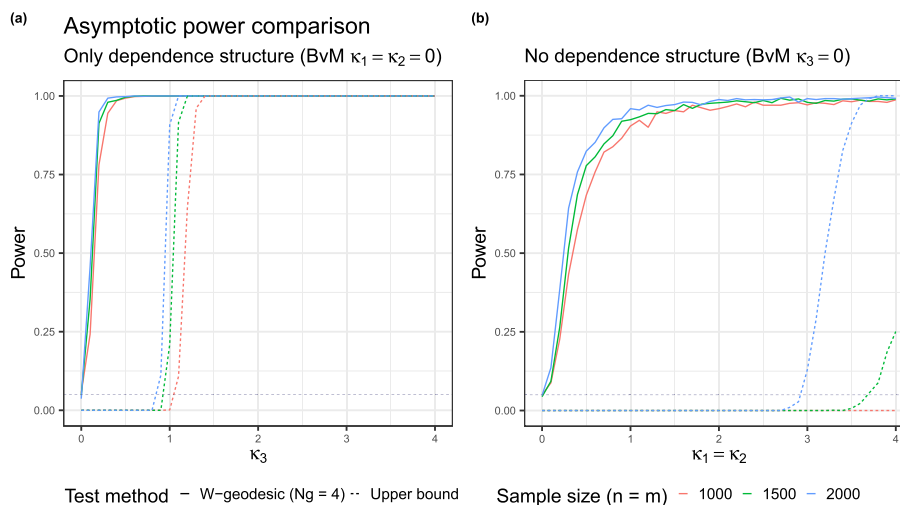
FIG 2. *Empirical power of two-sample goodness-of-fit tests for measures supported on* $\mathbb{T}^2$, *under bivariate von Mises (BvM) alternatives with no dependence structure and different marginal laws (a) and with equal marginal laws and dependence structure (b). The empirical power corresponds to the proportion of rejections at level* $\alpha = 0.05$ *(dashed line) among* $1000$ *repetitions of each test for fixed concentration parameters.*

local structure of a protein is determined by two dihedral angles usually denoted by $\phi$ and $\psi$, which describe the conformational state of each amino acid residue along the sequence [11, 37]. For most amino acid types (for all excepting proline and glycine), the distribution of $\phi$ and $\psi$ angles is supported on the same subset of $\mathbb{T}^2$, which, even if there exist some physically forbidden regions due to strong repulsive forces between non-bonded atoms at short distance, is connected and has a smooth boundary. We can also assume that density is continuously differentiable and strictly positive in its support, so that Assumption 1 is satisfied.

The aim of this section is to make use of the tests ($N_g$-geod) and (UB) to show that the distribution of $(\phi, \psi)$ does not depend only on the amino acid type, but also on the sequence context, and particularly on the closest neighbors. This corresponds to rejecting Flory's isolated-pair hypothesis [25]. Even if the importance of the closest neighbors effect is widely accepted in the Structural Biology community [35, 29, 4, 55, 63], only purely descriptive methods have been employed to state so, and no goodness-of-fit techniques have been used to the best of our knowledge. For a given amino acid $C$, we denote by $P_C$ the distribution of $(\phi, \psi)$ supported on $\mathbb{T}^2$. If we take into account the identities $L, R$ of $C$'s left and right neighbors, the distribution of $(\phi, \psi)$ is now given by $P_{LCR}$. The objective is to test

$$H_0 : P_C = P_{LCR} \qquad \text{against} \qquad H_1 : P_C \neq P_{LCR} \qquad (4.1)$$

to assess whether nearest neighbors significantly affect dihedral angles distributions. An example of two samples drawn from $P_C$ and $P_{LCR}$ is depicted in
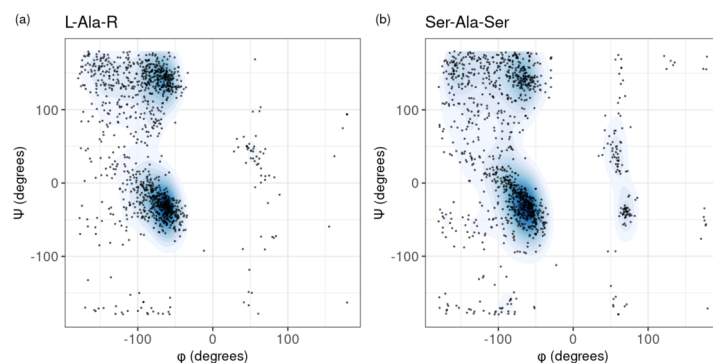
FIG 3. *(a) Sample and kernel density estimate of alanine* $(\phi, \psi)$ *distribution when the identity of its left and right neighbor is not taken into account. (b) Sample and kernel density estimate of* $(\phi, \psi)$ *distribution corresponding to tripeptide Ser-Ala-Ser (a fragment of the three consecutive amino-acids serine, alanine, serine).*

Figure 3. For the analysis presented here, we used a structural database of three-residue fragments (also called tripeptides) extracted from experimentally-determined high-resolution protein structures [48]. The large available sample sizes allow us to illustrate the asymptotic behaviour of (UB). We selected the 71 tripeptides *L-C-R* for which the database contained more than 3000 points. For each one, we compared the corresponding sample of $(\phi, \psi)$ values with an equally-sized sample drawn from $P_C$ (sampled from the sub-database containing $(\phi, \psi)$ values from tripeptides having $C$ as central amino-acid). The data were rescaled to $[0, 1] \times [0, 1]$ before applying the tests. As the *p*-values for the test ($N_g$-geod) are computed by Monte Carlo simulation, they are lower-bounded by $1/N_{MC}$, where $N_{MC}$ is the number of Monte Carlo replicas [51]. This point is important here, as due to the large number of performed tests, we had to correct *p*-values for multiplicity [31]. The results are depicted in Figure 4, where we show the empirical cumulative distribution function of both tests' corrected *p*-values, for three increasing ranges of sample sizes.

From Figure 4, we can state that the geodesic projection test ($N_g$-geod) strongly rejects the null hypothesis at level $\alpha = 0.05$ for the three considered sample size ranges, being all *p*-values truncated to the Monte Carlo precision. Repeating the same analysis for $N_g = 3, 4$ did not change the shape of the ($N_g$-geod) *p*-values curves, which was expected as higher values of $N_g$ yield a power increase. A clear asymptotic behaviour is observed for the upper bound technique (UB), as power at level $\alpha$ tends to one when sample sizes increase. Note that, for the largest range of sample sizes, (UB) is relatively more efficient than ($N_g$-geod), due to the Monte Carlo truncation. Both procedures lead to rejection of the null hypothesis, and therefore to the statement that nearest neighbors effect on $(\phi, \psi)$ distributions is statistically significant. This analysis suggests that both ($N_g$-geod) and (UB) are suitable for assessing differences on local protein structures, as the available sample sizes (which may be up to $\sim 10^5$
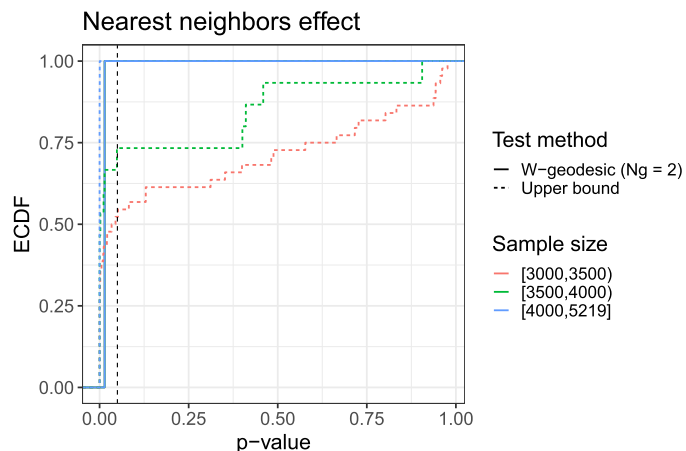
FIG 4. *Empirical cumulative distribution function of p-values corresponding to test hypotheses* (4.1) *with* ($N_g$-geod) *('W-marginal') and* (UB) *(Upper bound) testing methods, for* 71 *different combinations of L,C,R. To illustrate the asymptotic behaviour, p-values were classified in three ranges of sample sizes. For each test method, p-values were corrected for multiplicity using Holm-Bonferroni correction [31]. Marginal test p-values were computed with a Monte Carlo simulation of* $N_{MC} = 5000$ *replicas. The black dashed line indicates an arbitrary significance level of* $\alpha = 0.05$.

in some practical scenarios) are large enough to state significant conclusions.

## 5. Discussion

The main goal of this work was to define suitable two-sample goodness-of-fit tests for measures on $\mathbb{T}^2$. This naturally led us to enrich the existing theoretical results [14, 39, 42] on Optimal Transport for periodic measures. In particular, we studied the shape of the solutions to the Monge problem (2.5), which allowed the extension of a Central Limit Theorem to $\mathbb{T}^d$, for any $p > 1$. Our original inspiration when first investigating these theoretical results was to use the Central Limit Theorem 2.6 to define a two-sample asymptotic test. However, the derived limit distribution degenerates when $P = Q$ and prevents such an application. Nevertheless, the Wasserstein distance on $\mathbb{T}^2$ for the quadratic cost was used to define two efficient testing techniques, which address our initial goals.

The first approach bypasses the dimension problem by projecting the measures to closed geodesics on $\mathbb{T}^2$ and subsequently test their equality. This required the investigation of how to project samples on closed geodesics and, moreover, how to conveniently sample closed geodesics. The answers we propose here, notably in Sections A.1 and A.2, together with their supplied practical implementations, may be of interest in further practical situations. Furthermore, they suggest one possible extension of the Sliced Wasserstein distance [7] to the two-dimensional flat torus. As closed geodesics on $\mathbb{T}^2$ are isomorphic to $\mathbb{R}/\mathbb{Z}$, the equality of the projected measures is assessed through a two-sample Wasser-

stein test on the circle which, to the best of our knowledge, is the only efficient procedure proposed up to now.

The second proposed approach consists in upper-bounding the exact *p*-values (3.11). This is possible thanks to the derived concentration inequalities (3.12) for the two-sample empirical Wasserstein distance with the quadratic cost, and to the improved convergence speed of its expectation, as shown in Lemma 3.8. As with any upper-bounding technique, the corresponding test is conservative and only efficient for large sample sizes, which reduces its range of application. However, this test could be relevant in some practical scenarios. For example, Molecular Dynamics simulations (which simulate the temporal evolution of the structure of a protein using force-fields based on physical models), produce samples on $\mathbb{T}^2$ that may present small and meaningless differences when re-running simulations multiple times with slightly different initial conditions. In such a situation, we expect that the first technique ($N_g$-geod) will reject the equality of their corresponding distributions, while the conservative test (UB) will accept differences between independent replicas of the same simulation. Consequently, (UB) will only detect more important discrepancies, which are the only ones of interest for practical purposes.

Regarding the practical implementation of both tests, some differences appear with respect to computing time. The main advantage of ($N_g$-geod) is the explicit formulation of Wasserstein distance on one-dimensional spaces, which avoids the use of any Optimal Transport solver. As a result, its time complexity is linear in the sample size. However, the statistic null-distribution must be simulated with the desired precision, which may slow down the procedure. Note that, in any case, this distribution can be simulated once and be tabulated for any further implementation. The time complexity of (UB) exclusively lies on the Optimal Transport solver chosen to compute Wasserstein distance. For very large sample sizes, this might lead to a substantially slower process.

The issue of two-sample goodness-of-fit testing studied in Section 3 remains largely open. Our contribution in this respect is to propose easily implementable goodness-of-fit testing approaches that are built on top of state-of-the-art tools in Optimal Transport. Finding the exact or asymptotic distribution of the Wasserstein statistic in general dimension remains one of the main unsolved problems of the theory of Optimal Transport, preventing the construction of more efficient two-sample goodness-of-fit tests. An asymptotic approach for measures supported on a finite set has been presented in [61] and, in the one-dimensional case, [3] have obtained a CLT under the null $P = Q$ for deviations of $W_p(P_n, Q_n)$ from the true distance $W_p(P, Q)$ (instead of $\mathbb{E}(W_p(P_n, Q_n))$). The results of [3] are already quite challenging mathematically, and extensions to higher dimensions are clearly beyond the scope of the present work. Altogether, we believe that the goodness-of-fit tests defined in this paper constitute a relevant building block for the study of the sequence-structure-function relationship in proteins, and in particular for Intrinsically Disordered Proteins (IDPs), allowing their structural investigation with mathematical guarantees. Furthermore, the interest of the techniques here presented may go beyond the Structural Biology community, as they allow solving the goodness-of-fit testing

problem for two distributions lying in general periodic spaces, which appears in various application domains.

**Code availability**

The test approaches presented in this work are implemented in the R package `torustest`, available at https://github.com/gonzalez-delgado/torustest, together with the algorithms introduced in Appendix A. Empirical Wasserstein distances were computed using the R package `transport` [58].

## Appendix A: Geodesics on $\mathbb{T}^2$: practical considerations

This Section is devoted to address some practical questions that arise when defining the test proposed in Section 3.1. In Appendix A.1, we propose a sampling method to prevent the practitioner from explicitly choosing the $N_g$ geodesics, letting them be chosen randomly with respect to a given distribution. In Section A.2, we propose an algorithm to project a pair of samples on $\mathbb{T}^2$ to a given closed geodesic.

### *A.1. Sampling closed geodesics*

As the closed geodesics on $\mathbb{T}^2$ are given by the canonical projections of straight lines on $\mathbb{R}^2$ with rational slope, sampling from the set of all closed geodesics is equivalent to sampling from $\mathbb{Q}$, which is a countable set. This prevents the sampling to be uniform, in the sense that geodesics can not be equiprobable. Indeed, if $\mathbb{P}(q) = c$ for all $q \in \mathbb{Q}$, by countable additivity $\mathbb{P}(\mathbb{Q}) = \sum_{q \in \mathbb{Q}} c$, which is zero if $c = 0$ and $\infty$ otherwise. In consequence, as we have to assign different weights to rational slopes, we will opt for *simpler* geodesics to be more probable, in order to ease computational implementations. To achieve so, we can consider the random variable $Q = A/B$, studied in detail in [49], where $B$ follows a geometric distribution of parameter $p$ and, for a given denominator $B = b$, $A$ is uniform on $\{0, 1, \ldots, b\}$. Note that $Q$ maps into $\mathbb{Q} \cap [0, 1]$. As $p$ increases, $A$ and $B$ take smaller values and the corresponding geodesics revolt less over the torus. Conversely, when $p \to 0$, $\mathbb{P}(Q = q) \to 0$ for all $q \in \mathbb{Q} \cap [0, 1]$, and $Q$ is *asymptotically equiprobable* [49]. However, small values of $p$ yield extremely high values of $A$ and $B$ and, consequently, unmanageable geodesics with a too-big number of revolutions. The distribution of $Q$ for different values of the parameter $p$ is illustrated in Figure 5. Here, we will ask $p \geq 0.1$ for computational simplicity.

Note that rationals in $\mathbb{Q} \cap [0, 1]$ yield to straight lines in $\mathbb{R}^2$ whose director vector $(B, A)$ lies in the first (eq. fifth) octant. To cover all the set of closed geodesics, we uniformly assign an octant to each realization of $Q$ and transform its coordinates appropriately. As we would like all the $N_g$ $p$-values to be independent, we must only accept samples with $N_g$ different geodesics. This may be a problem if $N_g$ is too big, and might require decreasing the value of $p$.
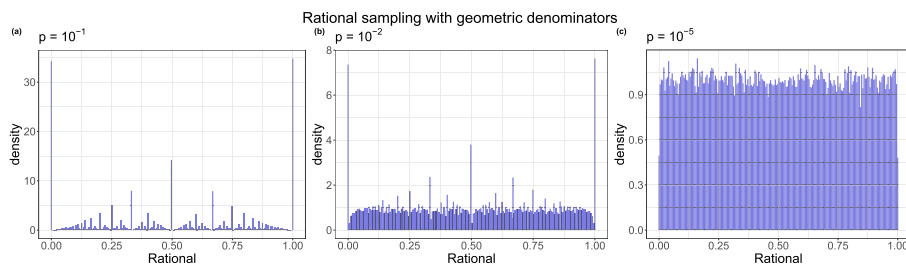
FIG 5. *Histograms representing the distribution of the random variable Q, for different values of the parameter p. For $p = 0.1$ (a), rationals with small values of A and B have more weight and, therefore, simpler geodesics are prioritized.*

Nevertheless, for a small number ($\lesssim 30$) of geodesics we can keep $p \sim 0.1$ and easily get samples with no repetitions. If one needs to perform the test for large values of $N_g$, we recommend to explicitly choose geodesics a priori to avoid this problem, leaving the sampling method for controlled values of $N_g$. The complete sampling procedure is described in Algorithm 1, which takes $N_g$ and $p$ as arguments and retrieves $N_g$ director vectors. In Algorithm 1, $\mathcal{M}_{N_g \times 2}(\mathbb{Z})$ denotes the set of $(N_g \times 2)$-matrices with entire entries and we define $\overset{\circ}{\gcd}$ as

$$\overset{\circ}{\gcd}(b, a) = \begin{cases} \gcd(b, a) & \text{if} & a \neq 0, \\ b & \text{otherwise}, \end{cases}$$

for $a, b \in \mathbb{Z}$ with $b \neq 0$.

---

**Algorithm 1** Geodesics sampling

---

**Require:** $N_g \in \mathbb{N}$, $p = 0.1$
**Ensure:** $G \in \mathcal{M}_{N_g \times 2}(\mathbb{Z})$
  $G \leftarrow 0 \in \mathcal{M}_{N_g \times 2}(\mathbb{Z})$
  **while** $|\{i = 1, \ldots, N_g : G_{ik} = G_{jk} \ \forall k \in \{1, 2\}$   for any $j \in \{1, \ldots, N_g\} \setminus \{i\} \}| > 0$ **do**
    **for** $i \leftarrow 1$ to $N_g$ **do**
      $b \leftarrow \mathcal{G}(p)$
      $a \leftarrow \mathcal{U}(\{0, 1, \ldots, b\})$
      $u \leftarrow (b, a)/\overset{\circ}{\gcd}(b, a)$                                   $\triangleright$ Director vector in $\mathbb{R}^2$.
      $o \leftarrow \mathcal{U}(\{1, 2, 3, 4\})$                          $\triangleright$ Octant of the upper semi-circle.
      **if** $o = 2$ **then**
        $u \leftarrow (a, b)/\overset{\circ}{\gcd}(b, a)$
      **else if** $o = 3$ **then**
        $u \leftarrow (-b, a)/\overset{\circ}{\gcd}(b, a)$
      **else if** $o = 4$ **then**
        $u \leftarrow (-a, b)/\overset{\circ}{\gcd}(b, a)$
      **end if**
      $G_{i\cdot} \leftarrow u$
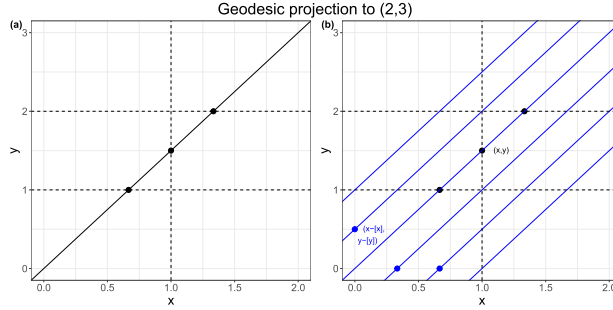    **end for**
  **end while**

---

FIG 6. *First steps of the projection algorithm for the closed geodesic corresponding to the straight line of director vector $u = (2,3)$. The three points in black constitute the ensemble $\mathcal{P}_u$. In (b), points of $\mathcal{P}_u$ are transferred to $[0,1] \times [0,1]$ by subtracting to their coordinates their integer parts. The blue lines are the elements of $\mathcal{L}_u$.*

### A.2. Projection to a closed geodesic

Let $a, b \in \mathbb{Z}$, with $b \neq 0$, and $u = (a, b)$ the director vector of a straight line $r_u^0$ containing the origin $(0,0)$. Let $I_a$ be the real interval $(\min(a,0), \max(a,0))$, being $I_b$ analogously defined. We aim to project a pair of samples into the geodesic given by the canonical projection of $r_u^0$. To do so, we first consider the finite set $\mathcal{P}_u$ of the points in $I_a \times I_b$ where $r_u^0$ cuts the lines $x = z_a$, $y = z_b$ for $z_a \in I_a \cap \mathbb{Z}$ and $z_b \in I_b \cap \mathbb{Z}$:

$$\mathcal{P}_u = \{(x, z) \,:\, x \in I_a, z \in \mathbb{Z}\} \cap \{(z, y) \,:\, y \in I_b, z \in \mathbb{Z}\} \cap r_u^0.$$

An example is presented in Figure 6a. Then, we consider the set $\mathcal{L}_u$ of straight lines of director vector $u$ and containing the points of $\mathcal{P}_u \cup \{(0,1), (1,0), (0,0)\}$ transferred to $[0,1] \times [0,1]$ by subtracting the integer part of its coordinates. If we denote $r_v^p$ the straight line containing $p = (p_x, p_y) \in \mathbb{R}^2$ and having $v$ as director vector, we can write $\mathcal{L}_u$ as follows

$$\mathcal{L}_u = \{r_u^q \,:\, q = (p_x - [p_x], p_y - [p_y]) \,, p \in \mathcal{P}_u \cup \{(0,1), (1,0), (0,0)\}\}.$$

This is illustrated in Figure 6b. In a first step, each point in $[0,1] \times [0,1]$ will be projected to the closest straight line in $\mathcal{L}_u$. Then, projections $(x_u, y_u)$ outside $[0,1] \times [0,1]$ will be replaced by the elements $(x'_u, y'_u) \in [0,1] \times [0,1]$ such that $(x_u, y_u)\mathcal{R}(x'_u, y'_u)$, where $\mathcal{R}$ is the one defined in the begging of Section 2. These two steps are depicted in Figure 7.

The last step is to relocate all the projections on $\mathbb{R}/\mathbb{Z}$. To do so, we put the segments $\mathcal{L}_u \cap ([0,1] \times [0,1])$ in order, following the spiral path. This corresponds to transfer back the points to the straight line $r_u^0$ of Figure 6a. Let $(x_u, y_u) \in r_u^p \in \mathcal{L}_u$. The element $t_u \in \mathbb{R}/\mathbb{Z}$ will be parameterized as

$$t_u = \frac{\|\tilde{p}\| + \|(x_u, y_u)\|}{\|u\|} \in [0,1),$$

where $\tilde{p} \in \mathcal{P}_u \cup \{(0,0)\}$ is the one such that $p_x = \tilde{p}_x - [\tilde{p}_x]$ and $p_y = \tilde{p}_y - [\tilde{p}_y]$.
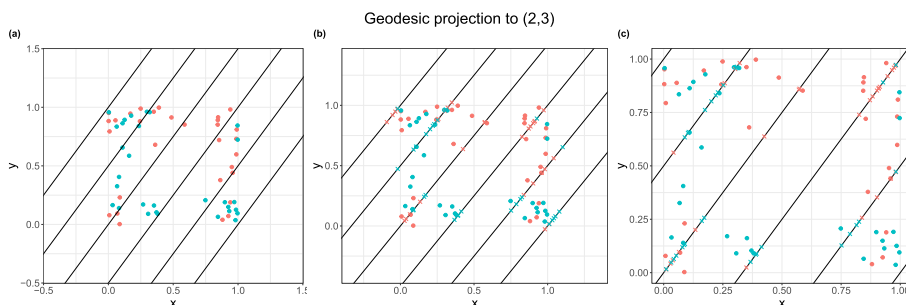
FIG 7. *Projection to the closed geodesic given by the director vector $u = (2,3)$ of a pair of samples of size $n = m = 30$ drawn from a uniform distribution on $\mathbb{T}^2$. Black lines are the elements of $\mathcal{L}_u$. In $(a)$, the given samples distinguished by colors. In $(b)$, their projections to the closest line in $\mathcal{L}_u$ are represented by colored crosses. In $(c)$, projections outside $[0,1] \times [0,1]$ are relocated in $[0,1] \times [0,1]$ according to the equivalence relation $\mathcal{R}$.*

## Appendix B: Proofs

### B.1. Proofs of Section 2

*Proof of Theorem 2.1.* Recall that we denote the interior of the support of a measure $\mu$ (over $\mathbb{T}^d$ or $\mathbb{R}^d$) as $\mathcal{X}_\mu$. Since $\mathbb{T}^d$ is a Polish space, Theorem 4.1 in [66] implies that there exists a solution $\pi^*$ of (2.2). Additionally, Theorem 5.10 in [66] establishes that $\mathrm{supp}(\pi^*)$ is $d^p$-cyclically monotone. More precisely, by Theorem 5.10 in [66], this support lies on the graph of the $d^p$-differential

$$\partial^{d^p} f(\bar{\mathbf{x}}) = \{\bar{\mathbf{y}} : \ f(\bar{\mathbf{z}}) \leq f(\bar{\mathbf{x}}) + d^p(\bar{\mathbf{z}}, \bar{\mathbf{y}}) - d^p(\bar{\mathbf{x}}, \bar{\mathbf{y}}), \text{ for all } \bar{\mathbf{z}} \in \mathbb{T}^d\}$$

of a function $f$ solving (2.3). Its graph is denoted by $\partial^{d^p} f = \{(\bar{\mathbf{x}}, \bar{\mathbf{y}}) : \ \bar{\mathbf{y}} \in \partial^{d^p} f(\bar{\mathbf{x}})\}$. These definitions of $d^p$-differential and $d^p$-concave functions apply verbatim to $\|\cdot\|^p$-differential and $\|\cdot\|^p$-concave functions with the obvious notation. Let $\Gamma$ be the set defined in (2.6), $\{(\mathbf{x}_k + \mathbf{p}_k, \mathbf{y}_k + \mathbf{p}_k)\}_{k=1}^n \subset \Gamma$ be a sequence and $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ be a bijection. Then, the definition of $\Gamma$ implies that

$$\begin{aligned}
\sum_{k=1}^n \|\mathbf{x}_k - \mathbf{y}_k\|^p &= \sum_{k=1}^n d^p(\bar{\mathbf{x}}_k, \bar{\mathbf{y}}_k) \\
&\leq \sum_{k=1}^n d^p(\bar{\mathbf{x}}_k, \bar{\mathbf{y}}_{\sigma(k)}) \\
&\leq \sum_{k=1}^n \|\mathbf{x}_k + \mathbf{p}_k - \mathbf{y}_{\sigma(k)} - \mathbf{p}_{\sigma(k)}\|^p,
\end{aligned}$$

which means that $\Gamma$ is $\|\cdot\|^p$-cyclically monotone. Therefore, $\Gamma \subset \partial^{\|\cdot\|^p}\varphi_p$, for some $\|\cdot\|^p$-concave function $\varphi_p$. Now, recall from Theorem 3.3 and Proposition 3.4 in [28], that

1. The set of differentiablity

$$\text{dom}(\nabla\varphi_p) = \left\{ \mathbf{x} \in \mathbb{R}^d : \ \partial^{\|\cdot\|^p}\varphi_p = \left\{ \mathbf{x} - \left( \frac{1}{p}\|\nabla\varphi_p(\mathbf{x})\| \right)^{\frac{2-p}{p-1}} \nabla\varphi_p(\mathbf{x}) \right\} \right\}$$

has full Lebesgue measure in $\text{dom}(\varphi_p) = \{ \mathbf{x} \in \mathbb{R}^d : \ \varphi_p(\mathbf{x}) \in \mathbb{R} \} \supset \mathcal{X}_{\mu_P}$,

2. The relation $\mathbf{S}_p(\mathbf{x}) = \mathbf{x} - \left( \frac{1}{p}\|\nabla\varphi_p(\mathbf{x})\| \right)^{\frac{2-p}{p-1}} \nabla\varphi_p(\mathbf{x})$ defines a Borel function in $\text{dom}(\nabla\varphi_p)$, and

3. The equality $\{\mathbf{S}_p(\mathbf{x})\} = \{\mathbf{y} : (\mathbf{x}, \mathbf{y}) \in \Gamma\}$ holds for all $\mathbf{x} \in \text{dom}(\nabla\varphi_p)$.

Since $\Gamma \subset \partial^{\|\cdot\|^p}\varphi_p$, this means that, for all $\mathbf{x} \in \text{dom}(\nabla\varphi_p)$, there exists an unique $\mathbf{y_x} = \mathbf{S}_p(\mathbf{x})$ such that $(\mathbf{x}, \mathbf{y_x}) \in \Gamma$. We observe that, due to the fact that $\mu_P \ll \ell_d$, the measure $\gamma^* = (\mathbf{Id} \times \mathbf{S}_p)\#\mu_P$ on $\mathbb{R}^d \times \mathbb{R}^d$ is well defined, its support is $\|\cdot\|^p$-cyclically monotone and its first marginal is $\mu_P$. We claim that the second marginal is $\mu_Q$. Let $(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in \pi^*$ be such that $\mathbf{x} + \mathbf{p} \in \text{dom}(\nabla\varphi_p)$, for all $\mathbf{p} \in \mathbb{Z}^d$. Then, for any representative pair, call it $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d$, there exist $\mathbf{p}, \mathbf{p}' \in \mathbb{Z}^d$ such that $(\mathbf{x} + \mathbf{p}, \mathbf{y} + \mathbf{p}') \in \Gamma$. Since

$$\begin{aligned}
\{\mathbf{S}_p(\mathbf{x}) + \mathbf{p}\} &= \{\mathbf{y} + \mathbf{p} : \ (\mathbf{x}, \mathbf{y}) \in \Gamma\} \\
&= \{\mathbf{y} : \ (\mathbf{x} + \mathbf{p}, \mathbf{y}) \in \Gamma\} \\
&= \{\mathbf{S}_p(\mathbf{x} + \mathbf{p})\} = \{\mathbf{y} + \mathbf{p}'\},
\end{aligned}$$

the relation $\bar{\mathbf{y}} = \overline{\mathbf{S}_p(\bar{\mathbf{x}})}$ holds. Since $\mathbf{x} + \mathbf{p} \in \text{dom}(\nabla\varphi_p)$, for all $\mathbf{p} \in \mathbb{Z}^d$, which is the intersection of sets of full $\mu_P$-measure, the relation $\bar{\mathbf{y}} = \overline{\mathbf{S}_p(\bar{\mathbf{x}})}$ happens $\mu_P$-a.e. This means that $\pi^* = (\mathbf{Id} \times \overline{\mathbf{S}_p(\cdot)})\#P$, which proves automatically the claim. Consequently, the existence is proven.

The uniqueness follows from the proof of Corollary 2.4. in [28]. Indeed, we can define the set

$$S = \bigcup_{\pi^* \text{ solving } (2.2)} \Gamma(\pi^*),$$

where $\Gamma(\pi^*)$ is defined as in (B.1) for each $\pi^*$ solving (2.2). Therefore, taking any finite sequence $\{(\mathbf{x}_k + \mathbf{p}_k, \mathbf{y}_k + \mathbf{p}_k)\}_{k=1}^n \subset S$, there exists at most $n$ different probability measures $\pi_k$, for $k = 1, \ldots, n$, such that $(\mathbf{x}_k + \mathbf{p}_k, \mathbf{y}_k + \mathbf{p}_k) \in \Gamma(\pi_k)$. As all of them are solutions of (2.2) we have, due to the linearity of the optimization in (2.2) and the convexity of the set $\Gamma(P, Q)$, that the mean $\pi_0 = \frac{1}{n}\sum_{k=1}^n \pi_k$ is also a solution. Then, its support is contained in a $d^p$-cyclically monotone set, and $\Gamma(\pi_0)$ is $\|\cdot\|^p$-cyclically monotone, since it contains the sequence $\{(\mathbf{x}_k + \mathbf{p}_k, \mathbf{y}_k + \mathbf{p}_k)\}_{k=1}^n \subset S$. Consequently, $S$ is $\|\cdot\|^p$-cyclically monotone.

To conclude, repeating the previous arguments, there exists a $\|\cdot\|^p$-concave function $f^S$ such that $S \subset \partial^d f^S$. Moreover, for any other $\varphi_p$, defined as before, it holds that $\partial^d \varphi_p \subset \partial^d f^S$. Then, the equality

$$\mathbf{x} - \left( \frac{1}{p}\|\nabla\varphi_p(\mathbf{x})\| \right)^{\frac{2-p}{p-1}} \nabla\varphi_p(\mathbf{x}) = \mathbf{x} - \left( \frac{1}{p}\|\nabla f^S(\mathbf{x})\| \right)^{\frac{2-p}{p-1}} \nabla f^S(\mathbf{x})$$

holds $\mu_P$-a.e. This proves the uniqueness of $\mathbf{S}_p$ and, consequently, the one of $\mathbf{T}_p$ ∎

*Proof of Theorem 2.2.* We set $(\mathbf{x}, \mathbf{y}) \in \Gamma$ and observe that $d(\bar{\mathbf{x}}, \bar{\mathbf{y}}) = \|\mathbf{x} - \mathbf{y}\|$. Since $(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in \text{supp}(\pi^*)$, Theorem 5.10 in [66] establishes that if $(f, g)$ solves (2.3), it holds

$$f(\bar{\mathbf{x}}) = \inf_{\bar{\mathbf{y}} \in \mathbb{T}^d} \{d(\bar{\mathbf{x}}, \bar{\mathbf{y}})^p - g(\bar{\mathbf{y}})\}.$$

Since, for each $(\mathbf{x}, \mathbf{y})$, there exists $\mathbf{p} \in \mathbb{Z}^d$ such that $d(\bar{\mathbf{x}}, \bar{\mathbf{y}}) = \|\mathbf{x} - \mathbf{y} - \mathbf{p}\|$, we can directly replace $\mathbf{y}$ by $\mathbf{y} + \mathbf{p}$ in the infimum without altering any of the terms. This yields the equality

$$f(\bar{\mathbf{y}}) = \inf_{z, \mathbf{y} \in \mathbb{R}^d} {}_{\bar{\mathbf{z}} = \bar{\mathbf{x}}} \{\|\mathbf{z} - \mathbf{y}\|^p - g(\bar{\mathbf{y}})\},$$

and allows to define the following periodic $\|\cdot\|^d$-concave function in $\mathbb{R}^d$:

$$\hat{\varphi}_p(\mathbf{x}) = \inf_{\mathbf{y} \in \mathbb{R}^d} \{\|\mathbf{x} - \mathbf{y}\|^p - g(\bar{\mathbf{y}})\} = f(\bar{\mathbf{x}}).$$

We claim that $\nabla \hat{\varphi}_p = \nabla \varphi_p$ for $\mu_P$-a.e., which implies the equality of both $\hat{\varphi}_p$ and $\varphi_p$, in each connected component of $\text{supp}(\mu_P)$. By assumption, $\text{supp}(\mu_P) = \bigcup_{p \in \mathbb{Z}^d} p + A$ is a union of connected sets. By periodicity we can restrict our study to the connected set $A$, where the claim yields $\nabla \hat{\varphi}_p = \nabla \varphi_p$ for $\ell_d$-a.e. We can apply Theorem 2.6 in [19] to conclude that $\varphi_p = \hat{\varphi}_p + C$ in $A$, thus in $\text{supp}(\mu_P)$. We prove now the claim. Let $\pi^*$ be a measure solving (2.2), we know (from Theorem 5.10 in [66]) that its support lies in the graph of $\partial^{d^p} f$. Therefore, we can define the following $\|\cdot\|^p$-cyclically monotone set (note that this is true by repeating the same arguments as for $\Gamma$):

$$\Gamma(\partial^{d^p} f) = \{(\mathbf{x} + p, \mathbf{y} + p) :$$
$$(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in \partial^{d^p} f, \ \mathbf{x} \in [0, 1]^d, \ d(\bar{\mathbf{x}}, \bar{\mathbf{y}}) = \|\mathbf{x} - \mathbf{y}\| \text{ and } p \in \mathbb{Z}^d\}, \quad \text{(B.1)}$$

which satisfies the relation $\Gamma(\pi^*) \subset \Gamma(\partial^{d^p} f)$, with the notation of the proof of Theorem 2.1. Recall that the relation $\Gamma(\pi^*) \subset \partial^{\|\cdot\|^p} \varphi_p$ also holds. Moreover, by definition we have $\Gamma(\partial^{d^p} f) \subset \partial^{\|\cdot\|^p} \hat{\varphi}_p$. Since $\mu_P$-a.e. the sets $\partial^{\|\cdot\|^p} \hat{\varphi}_p(\mathbf{x})$ and $\partial^{\|\cdot\|^p} \varphi_p(\mathbf{x})$ are singletons, and, for $\mu_P$-a.e. $\mathbf{x}$, there exists at least one $\mathbf{y} \in \mathbb{R}^d$ such that $(\mathbf{x}, \mathbf{y}) \in \Gamma(\pi^*)$, then $\partial^{\|\cdot\|^p} \hat{\varphi}_p(\mathbf{x}) = \partial^{\|\cdot\|^p} \varphi_p(\mathbf{x})$. This implies that the functions $\mathbf{S}_p(\mathbf{x}) = \mathbf{x} - \left(\frac{1}{p}\|\nabla \varphi_p(\mathbf{x})\|\right)^{\frac{2-p}{p-1}} \nabla \varphi_p(\mathbf{x})$ and $\hat{\mathbf{S}}_p(\mathbf{x}) = \mathbf{x} - \left(\frac{1}{p}\|\nabla \hat{\varphi}_p(\mathbf{x})\|\right)^{\frac{2-p}{p-1}} \hat{\varphi}_p(\mathbf{x})$ are equal $\mu_P$-a.e., which proves the claim. Note that, under continuity of the optimal transport potential, their uniqueness only need to be fulfilled $\mu_P$-a.e. ∎

*Proof of Lemma 2.4.* Set $\bar{\mathbf{x}}, \bar{\mathbf{z}} \in \text{dom}(f)$. Then, by definition

$$|f(\bar{\mathbf{x}}) - f(\bar{\mathbf{z}})| = |\inf_{\bar{\mathbf{y}} \in \mathbb{T}^d} \{d^p(\bar{\mathbf{x}}, \bar{\mathbf{y}}) - g(\bar{\mathbf{y}})\} - \inf_{\bar{\mathbf{y}} \in \mathbb{T}^d} \{d^p(\bar{\mathbf{z}}, \bar{\mathbf{y}}) - g(\bar{\mathbf{y}})\}|$$

$$= |\inf_{\bar{\mathbf{y}} \in \mathbb{T}^d} \{d^p(\bar{\mathbf{x}}, \bar{\mathbf{y}}) - g(\bar{\mathbf{y}})\} + \sup_{\bar{\mathbf{y}} \in \mathbb{T}^d} \{-d^p(\bar{\mathbf{z}}, \bar{\mathbf{y}}) + g(\bar{\mathbf{y}})\}|$$

$$\leq \sup_{\bar{\mathbf{y}} \in \mathbb{T}^d} |d^p(\bar{\mathbf{x}}, \bar{\mathbf{y}}) - d^p(\bar{\mathbf{z}}, \bar{\mathbf{y}})\}|.$$

The mean value theorem yields the inequality $a^p - b^p \leq p|a - b|(a^{p-1} + b^{p-1})$, which holds for any $a, b \geq 0$. Then, the triangle inequality for $d$ leads to

$$|f(\bar{\mathbf{z}}) - f(\bar{\mathbf{x}})| \leq p\, d(\bar{\mathbf{z}}, \bar{\mathbf{x}}) \sup_{\bar{\mathbf{y}} \in \mathbb{T}^d} |d^{p-1}(\bar{\mathbf{x}}, \bar{\mathbf{y}}) + d^{p-1}(\bar{\mathbf{z}}, \bar{\mathbf{y}})\}$$

$$\leq 2\, p\, d(\bar{\mathbf{z}}, \bar{\mathbf{x}}) \sup_{\bar{\mathbf{z}}, \bar{\mathbf{x}} \in \mathbb{T}^d} \left(d^{p-1}(\bar{\mathbf{z}}, \bar{\mathbf{x}})\right) \leq 2\, p\, d^{\frac{p-1}{2}}\, d(\bar{\mathbf{z}}, \bar{\mathbf{x}})$$

where the $d^{\frac{p-1}{2}}$ term comes from the trivial bound of the diameter of $\mathbb{T}^d$. This concludes the proof. $\qquad\square$

*Proof of Theorem 2.5.* Set $\bar{p} \in \mathcal{X}_P$ and assume that $f_p(\bar{p}) = 0$. Set $\epsilon_m \to 0$ and consider the sequence of balls $\mathbb{B}_{\epsilon_m}(\bar{p}) \subset \mathrm{supp}(P)$, centered at $\bar{p}$ with radius $\epsilon_n$. Since the ball is a continuity set of $P$, after Portmanteau's Theorem, $P_n \xrightarrow{w} P$ implies that for each $m$ there exists a $n_m$ such that $P_n$ gives mass to $\mathbb{B}_{\epsilon_m}(\bar{p})$ for all $n \geq m_n$. Then, we can extract a sequence $\bar{p}_n \to \bar{p}$ such that $\bar{p}_n \in \mathcal{X}_{P_n}$. As a consequence, we have that $f_n(\bar{p}_n) \in \mathbb{R}$, and we can set $a_n = -f_n(\bar{p}_n)$ and define $h_n = f_n + a_n$. Recall from Lemma 2.4 that all such functions are $L$-Lipschitz in their respective domains. Kirszbraun's Theorem (Theorem B in [36]) implies that, without loss of generality, we can consider that $h_n$ (resp. $f_p$) are $2p$-Lipschitz functions defined in the whole $\mathbb{T}^d$. The previous reasoning implies that $\{h_n\}_{n \in \mathbb{N}}$ is point-wise bounded for the compact sequence $\{\bar{p}_n\}_{n \in \mathbb{N}}$. Since all such functions are $2p$-Lipschitz, then Arzelá-Ascoli's Theorem concludes that every subsequence $\{h_{n_k}\}_{k \in \mathbb{N}}$ admits a convergent subsequence $\{h_{n_{k_j}}\}_{j \in \mathbb{N}}$. Let $h$ be one of those limits. Note that the $d^p$-conjugation is continuous in the sense that

$$|h_n^{d^p}(\bar{\mathbf{x}}) - h^{d^p}(\bar{\mathbf{x}})| = |\inf_{\bar{\mathbf{y}}} \{d^p(\bar{\mathbf{y}}, \bar{\mathbf{x}}) - h_n(\bar{\mathbf{x}})\} - \inf_{\bar{\mathbf{y}}} \{d^p(\bar{\mathbf{y}}, \bar{\mathbf{x}}) - h(\bar{\mathbf{x}})\}|$$

$$\leq \sup_{\bar{\mathbf{y}}} \{h_n(\bar{\mathbf{x}}) - h(\bar{\mathbf{x}})\} = \|h_n - h\|_\infty,$$

for all $\bar{\mathbf{y}} \in \mathbb{T}^d$. By assumption, we have

$$A_n = \int h_n d\alpha_n + \int h_n^{d^p} d\beta_n - \int f_p d\alpha - \int f_p^{d^p} d\beta \to 0,$$

and

$$\int h d\alpha + \int h^{d^p} d\beta =$$
$$\int h d(\alpha - \alpha_n) + \int h^{d^p} d(\beta - \beta_n) + \int (h_n - h) d\alpha_n + \int (h_n^{d^p} - h^{d^p}) d\beta_n.$$

Then, the inequality $|\int (h_n - h)d\alpha_n| \le \|h_n - h\|_\infty$ gives $\int h d\alpha + \int h^{d^p} d\beta = 0$. The function $h$ is thus an optimal transport potential. The uniqueness described in Theorem 2.2 and the fact that $\bar{p}_n \to \bar{p}$ and $h_n(\bar{p}_n) = f_p(\bar{p}) = 0$ conclude that $f_p$ is the unique possible limit of such subsequences in $\text{dom}(f_p)$. $\qquad\square$

*Proof of Theorem 2.6.* Note that as Theorem 2.5 holds, since probability measures are supported in a compact set, the torus, then the reasoning of [22] can be imitated. Here the main steps of the proof for the one-sample case are given. For further details about the proof we refer to the original text.

Efron-Stein inequality, see Chapter 3.1 in [10], states that if $(X_1', \ldots, X_n')$ is an independent copy of $(X_1, \ldots, X_n)$, then we have the bound

$$\text{Var}(f(X_1, \ldots, X_n)) \le \sum_{i=1}^n \mathbb{E}(f(X_1, \ldots, X_n)$$
$$- f(X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n).)_+^2.$$

Moreover, if $X_1, \ldots, X_n$ are i.i.d, such inequality can be written as

$$\text{Var}(f(X_1, \ldots, X_n)) \le n\mathbb{E}(f(X_1, \ldots, X_n) - f(X_1', \ldots, X_n))_+^2.$$

Set the empirical measures $P_n = \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$ and $P_n' = \frac{1}{n}(\delta_{X_1'} + \sum_{k=2}^n \delta_{X_k})$, and the values $R_n = \mathcal{T}_p(P_n, Q) - \int f_p dP_n$ and $R_n' = \mathcal{T}_p(P_n', Q) - \int f_p dP_n'$. Let $f_n$ and $f_n'$ be solutions of the dual problem (2.3) of $\mathcal{T}_p(P_n, Q)$ and $\mathcal{T}_p(P_n', Q)$ respectively. Then from (2.3) we derive that

$$(R_n - R_n')_+ \le \frac{1}{n}|f_n(X_1) - f_p(X_1) - f_n(X_1') + f_p(X_1')|$$
$$+ |f_n'(X_1) - f_p(X_1) - f_n'(X_1') + f_p(X_1')|,$$

which together with Theorem 2.5 yields

$$n(R_n - R_n')_+ \xrightarrow{a.s.} 0.$$

Since the probability measures are supported in the torus, which is compact, then $n^2\mathbb{E}(R_n - R_n')_+^2 \to 0$. Finally, we conclude by the so called Efron-Stein's inequality. $\qquad\square$

### B.2. Proofs of Section 3

*Proof of Lemma 3.1.* If $G$ is the distribution function of the uniform distribution on $\mathbb{R}/\mathbb{Z}$, we have that

$$(G - \alpha)^{-1}(t) = \inf\{s \ : \ s > t + \alpha\} = t + \alpha. \tag{B.2}$$

Plugging (B.2) in (3.6), we have

$$\mathcal{T}_2(P^c, U) = \inf_{\alpha \in \mathbb{R}} \int_0^1 \left(F^{-1}(t) - t - \alpha\right)^2 dt, \tag{B.3}$$

where the optimal value for $\alpha$ can be found by analytically minimizing the function

$$H(\alpha) = \int_0^1 \left(F^{-1}(t) - t - \alpha\right)^2 dt = \int_0^1 \left(F^{-1}(t) - t\right)^2 dt +$$

$$\alpha^2 - 2\alpha \int_0^1 \left(F^{-1}(t) - t\right) dt,$$

which satisfies $H'(\alpha) = 0 \Leftrightarrow \alpha = \int_0^1 \left(F^{-1}(t) - t\right) dt$. □

*Proof of Lemma 3.2.* Let $\mathcal{D}([0,1])$ denote the Banach space of right-continuous functions on $[0,1]$ with left limits. Donsker's Theorem [5, Theorem 14.3], states the weak convergence in $\mathcal{D}([0,1])$ of the empirical process $\sqrt{n}(F_n - F)$ for $n \to \infty$ to the standard Brownian bridge $\mathbb{B}(F(t))$. As the operator $h : \mathcal{D}([0,1]) \longrightarrow \mathbb{R}$ defined as

$$h(f) = \int_0^1 \left(f(t) - \int_0^1 f(s)\, ds\right)^2 dt = \int_0^1 f(t)^2\, dt - \left(\int_0^1 f(s)\, ds\right)^2 \quad \text{(B.4)}$$

is continuous, the continuous mapping Theorem [64, Theorem 1.3.6] yields that

$$n\mathcal{T}_2(P_n^c, U) \xrightarrow[n]{w} \int_0^1 \mathbb{B}(t)^2\, dt - \left(\int_0^1 \mathbb{B}(t)\, dt\right)^2, \quad \text{(B.5)}$$

when $P^c = U$, which concludes the proof. □

*Proof of Proposition 3.3.* Keeping the notation of the proof of Lemma 3.2, after Theorem 1 in [54] we have that, when $P^c = Q^c$,

$$\sqrt{\frac{nm}{n+m}} \left(G_m^{-1}(F_n) - \mathbb{I}\right) \xrightarrow[n,m]{w} \mathbb{B}(t), \quad \text{(B.6)}$$

in $\mathcal{D}([0,1])$, where $\mathbb{I}$ denotes the identity function. Finally, using the same arguments as in the proof of Lemma 3.2, the result is proved. □

*Proof of Proposition 3.4.* Note that

$$\mathbb{P}(\pi_{nm}^c = 1) = \mathbb{P}(T_{nm}^c \geq c_{nm}^c(\alpha)) = \mathbb{P}\left(\mathcal{T}_2(G_m \# P_n, U) \geq \frac{n+m}{nm} c_{nm}^c(\alpha)\right).$$

On one hand, we have that

$$c_{nm}^c(\alpha) = \inf\{t > 0 : F_{nm}^c(t) \geq 1 - \alpha\} = \inf\{t > 0 : \mathbb{P}_{H_0}(T_{nm}^c > t) \leq \alpha\} =$$

$$\frac{nm}{n+m} \inf\{t > 0 : \mathbb{P}_{H_0}(\mathcal{T}_2(G_m \# P_n, U) > t) \leq \alpha\}.$$

Under the null hypothesis, $G_m \# P_n \xrightarrow{w} U$. Thus, $\mathcal{T}_2(G_m \# P_n, U) \to 0$ in probability (recall Section 2.2). In consequence, for every $\varepsilon > 0$ and every $\alpha > 0$, we have

$$\frac{n+m}{nm} c_{nm}^c(\alpha) \leq \varepsilon \quad \text{(B.7)}$$

for sufficiently large $n, m$. On the other hand, when $P^c \neq Q^c$, $\mathcal{T}_2(G_m \# P_n, U) \to \mathcal{T}_2(G \# P, U) > 0$ in probability which, together with (B.7), proves the result. □

*Size of* ($N_g$-geod). Let us prove that ($N_g$-geod) controls the type I error at any significance level $\alpha > 0$. Indeed, if $H_0$ denotes the null hypothesis (3.1), we have

$$\mathbb{P}_{H_0}\left(\pi^g_{nm,N_g} = 0\right) = \mathbb{P}_{H_0}\left(\min_{i=1}^{N_g} p_i \leq \frac{\alpha}{N_g}\right) = \mathbb{P}_{H_0}\left(\bigcup_{i=1}^{N_g}\left\{p_i \leq \frac{\alpha}{N_g}\right\}\right) \leq$$

$$\sum_{i=1}^{N_g}\mathbb{P}_{H_0}\left(p_i \leq \frac{\alpha}{N_g}\right) = N_g\frac{\alpha}{N_g} = \alpha, \quad \text{(B.8)}$$

where the first equality in (B.8) is ensured as every $p_i$ follows a uniform distribution under the null. ∎

*Proof of Proposition 3.5.* Suppose that $P^c_j \neq Q^c_j$ w.l.o.g. for some $j \in \{1, \ldots, N_g\}$. After ($N_g$-geod), we have

$$\mathbb{P}(\pi^g_{nm,N_g} = 1) = \mathbb{P}\left(\min_{i=1}^{N_g} p_i \leq \frac{\alpha}{N_g}\right) \geq \mathbb{P}\left(p_j \leq \frac{\alpha}{N_g}\right).$$

Then, as the right side of the previous inequality tends to 1 after Proposition 3.4, so does the left side, which ends the proof. ∎

*Proof of Remark 3.6.* Suppose that $\mu_P \ll \ell_2$ and project with respect a given direction $\mathbf{e}_1$. As an immediate consequence of the monotone convergence theorem, $\ell_2(A \times \mathbb{R}) = 0$, for any Lebesgue null set $A \subset \mathbb{R}$. Consequently, by hypothesis $0 = \mu_P(A \times \mathbb{R}) = \mu^1_P(A)$. Here, $\mu^1_P$ is the projected measure of $\mu_P$ to the direction $\mathbf{e}_1$. Then, for any null set $B$ in $\mathbb{R}/\mathbb{Z}$ the leveraged set $\tilde{B} = \bigcup_{s \in \mathbb{Z}}(s + B)$ is a Lebesgue null set, so that $\mu^1_P(\tilde{B}) = 0$ and $P^c(B) = 0$. ∎

*Proof of Theorem 3.7.* Note that $\mathcal{T}_2(P_n, Q_m) = \mathcal{T}(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ is a function of $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$. For each $\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}_1, \ldots, \mathbf{y}_m \in \mathbb{T}^d$ and $\mathbf{x}' \in \mathbb{T}^d$ let $\pi$ and $\pi'$ be both joint measures such that

$$\mathcal{T} := \sum_{i,j} d(\mathbf{x}_i - \mathbf{y}_j)^2 \pi_{i,j} = \mathcal{T}(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}_1, \ldots, \mathbf{y}_m)$$

$$s.t. \quad \sum_{i,j} \pi_{i,j} = \frac{1}{n}, \quad j = 1, \ldots, m,$$

$$\sum_{i,j} \pi_{i,j} = \frac{1}{m}, \quad i = 1, \ldots, n,$$

and

$$\mathcal{T}' := \sum_{j} d(\mathbf{x}'_1 - \mathbf{y}_j)^2 \pi'_{1,j} + \sum_{i>1,j} d(\mathbf{x}_i - \mathbf{y}_j)^2 \pi'_{i,j} = \mathcal{T}(\mathbf{x}'_1, \ldots, \mathbf{x}_n, \mathbf{y}_1, \ldots, \mathbf{y}_m)$$

$$s.t. \quad \sum_{i,j} \pi'_{i,j} = \frac{1}{n}, \quad j = 1, \ldots, m,$$

$$\sum_{i,j} \pi'_{i,j} = \frac{1}{m}, \quad i = 1, \ldots, n.$$

Then we have that

$$\mathcal{T}' \leq \sum_j d(\mathbf{x}_1' - \mathbf{y}_j)^2 \pi_{1,j} + \sum_{i,j} d(\mathbf{x}_i - \mathbf{y}_j)^2 \pi_{i,j},$$

which implies

$$\mathcal{T}' - \mathcal{T} \leq \sum_j \left( d(\mathbf{x}_i - \mathbf{y}_j)^2 - d(\mathbf{x}_1 - \mathbf{y}_j)^2 \right) \pi_{1,j}$$

$$\leq \sum_j \frac{1}{2}\pi_{1,j} = \frac{1}{2n},$$

where the second inequality comes from the fact that $d^2(\mathbf{x}, \mathbf{y}) \leq 1/2$ in $\mathbb{T}^d$. By symmetry we also obtain the reverse inequality. Doing the same with $\mathbf{y}_1'$ and $\mathbf{y}_1$ we obtain the bound $\frac{1}{2m}$. By using McDiarmid's inequality, see [44], we derive that

$$\mathbb{P}\left(\mathcal{T}_2(P_n, Q_m) - \mathbb{E}\mathcal{T}_2(P_n, Q_m) > t\right) \leq \exp\left(-\frac{nm}{n+m}8t^2\right). \qquad \square$$

*Proof of Proposition 3.9.* Let $P = Q$. After Lemma 3.8, for every $\varepsilon > 0$ these exists $N_\varepsilon \in \mathbb{N}$ such that for all $n, m \geq N_\varepsilon$, $\mathbb{E}\mathcal{T}_2(P_n, Q_m) \leq \varepsilon$. Using explicitly the convergence speed, we can find the relationship between $\varepsilon$ and $N_\varepsilon$:

$$\frac{\log N_\varepsilon}{N_\varepsilon} = \frac{\varepsilon}{C}, \tag{B.9}$$

where $C > 0$ is an unspecified constant. Then, directly from Theorem 3.7, we can bound (3.11) as

$$\mathbb{P}_{H_0}\left(\mathcal{T}_2(P_n, Q_m) > t\right) \leq \exp\left(-\frac{nm}{n+m}8(t - \mathbb{E}\mathcal{T}_2(P_n, Q_m))^2\right) \leq$$

$$\exp\left(-\frac{nm}{n+m}8(t - \varepsilon)^2\right),$$

for all $n, m \geq N_\varepsilon$. $\qquad \square$

*Proof of Proposition 3.10.* Let us first prove that (UB) is asymptotically of level $\alpha$. Let $\varepsilon > 0$ and $N_\varepsilon \in \mathbb{N}$ such that for all $n, m \geq N_\varepsilon$, the test (UB) controls type I error. As we are taking the limit $n, m \to \infty$, we can choose $n, m$ large enough such that they surpass $N_\varepsilon$. Then, consistency is ensured by Proposition 3.9.

To conclude, we prove the consistency under fixed alternatives such that $\mathcal{T}_2(P, Q) > \varepsilon$. First, note that

$$\mathbb{P}(\pi_{nm,\varepsilon}^{ub} = 1) = \mathbb{P}\left(\mathcal{T}_2(P_n, Q_m) \geq \varepsilon + \sqrt{-\frac{n+m}{8nm}\log\alpha}\right)$$

$$= \mathbb{P}\left(\sqrt{\frac{mn}{n+m}}(\mathcal{T}_2(P_n, Q_m) - \mathcal{T}_2(P, Q)) \geq \sqrt{\frac{mn}{n+m}}(\varepsilon - \mathcal{T}_2(P, Q)) + \sqrt{-\frac{1}{8}\log\alpha}\right).$$

Now, (2.12) implies, under the alternative, the stochastically boundedness of the

left hand side. However, the right hand side is clearly unbounded if $\mathcal{T}_2(P,Q) > \epsilon$. Consequently,

$$\lim_{n,m\to\infty} \mathbb{P}(\pi_{nm,\varepsilon}^{ub} = 1) = 1,$$

which concludes the proof. □

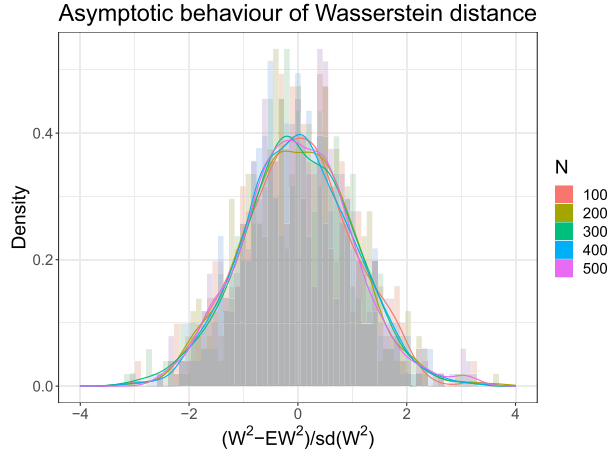## Appendix C: Supplementary figures



FIG 8. *Normalized asymptotic deviations from the mean of squared Wasserstein distance between two bivariate von Mises distributions of same means $(\mu,\nu) = (0,0)$ and different concentration parameters $(\kappa_1, \kappa_2, \kappa_3) = (0,0,0)$ and $(\kappa_1, \kappa_2, \kappa_3) = (2,2,0)$. The figures show the corresponding histograms and the associated kernel density estimates, for different sample sizes.*
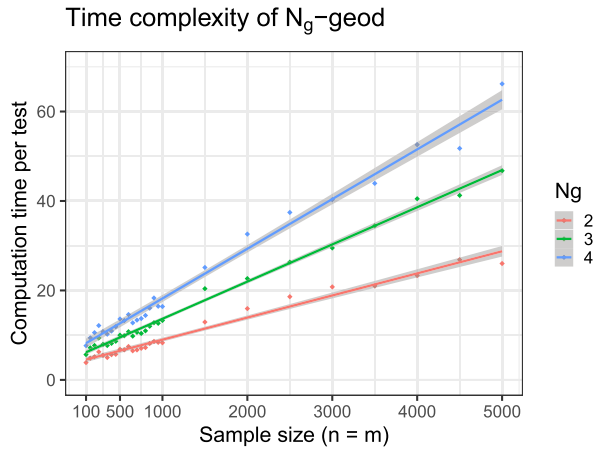


FIG 9. *Empirical time complexity of ($N_g$-geod) for $N_g = 2,3,4$. Each point corresponds to the average computation time per test among 200 repetitions of ($N_g$-geod) for two equally sized samples drawn from a bivariate von Mises distributions of equal means $(\mu,\nu) = (0,0)$ and different concentration parameters $(\kappa_1, \kappa_2, \kappa_3) = (0,0,0)$ and $(\kappa_1, \kappa_2, \kappa_3) = (1,1,0)$. The lines correspond to a linear regression performed for each value of $N_g$.*

## Acknowledgments

## References

[1] Ambrosio, L., Goldman, M. and Trevisan, D. (2021). On the quadratic random matching problem in two-dimensional domains. *Electronic Journal of Probability*. MR4416128

[2] Arjovsky, M., Chintala, S. and Bottou, L. (2017). Wasserstein Generative Adversarial Networks. In *Proceedings of the 34th International Conference on Machine Learning* (D. Precup and Y. W. Teh, eds.). *Proceedings of Machine Learning Research* **70** 214–223. PMLR.

[3] Berthet, P. and Fort, J.-C. (2019). Weak convergence of empirical Wasserstein type distances. arXiv:1911.02389v1.

[4] Betancourt, M. R. and Skolnick, J. (2004). Local Propensities and Statistical Potentials of Backbone Dihedral Angles in Proteins. *Journal of Molecular Biology* **342** 635–649.

[5] Billingsley, P. (1999). *Convergence of probability measures*, second ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. John Wiley & Sons Inc., New York. MR1700749

[6] Bonferroni, C. E. (1936). *Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del R. Istituto superiore di scienze economiche e commerciali di Firenze*. Libreria internazionale Seeber.

[7] Bonneel, N., Rabin, J., Peyré, G. and Pfister, H. (2014). Sliced and Radon Wasserstein Barycenters of Measures. *Journal of Mathematical Imaging and Vision* **51**. MR3300482

[8] Bonneel, N., van de Panne, M., Paris, S. and Heidrich, W. (2011). Displacement Interpolation Using Lagrangian Mass Transport. *ACM Transactions on Graphics (SIGGRAPH ASIA 2011)* **30**.

[9] Boothby, W. M. (1975). *An Introduction to Differentiable Manifolds and Riemannian Geometry. Pure and Applied Mathematics*. Academic Press, London. MR0426007

[10] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. OUP Oxford. MR3185193

[11] Brändén, C. I. and Tooze, J. (1998). *Introduction to Protein Structure (2nd ed.)*. Garland Science, New York.

[12] Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics* **44** 375–417. MR1100809

[13] Chernozhukov, V., Galichon, A., Hallin, M. and Henry, M. (2017). Monge–Kantorovich depth, quantiles, ranks and signs. *The Annals of Statistics* **45** 223–256. MR3611491

[14] CORDERO-ERAUSQUIN, D. (1999). Sur le transport de mesures périodiques. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics* **329** 199–202. MR1711060

[15] CUESTA, J. A. and MATRÁN, C. (1989). Notes on the Wasserstein metric in Hilbert spaces. *The Annals of Probability* 1264–1276. MR1009457

[16] CÁRCAMO, J., CUEVAS, A. and RODRÍGUEZ, L.-A. (2020). Directional differentiability for supremum-type functionals: Statistical applications. *Bernoulli* **26** 2143–2175. MR4091104

[17] DEL BARRIO, E., CUESTA-ALBERTOS, J. A., MATRAN, C. and RODRIGUEZ-RODRIGUEZ, J. M. (1999). Tests of Goodness of Fit Based on the L2-Wasserstein Distance. *The Annals of Statistics* **27** 1230–1239. MR1740113

[18] DEL BARRIO, E., GONZÁLEZ-SANZ, A. and HALLIN, M. (2020). A note on the regularity of optimal-transport-based center-outward distribution and quantile functions. *Journal of Multivariate Analysis* 104671. MR4147635

[19] DEL BARRIO, E., GONZÁLEZ-SANZ, A. and LOUBES, J.-M. (2021). Central Limit Theorems for General Transportation Costs. arXiv:2102.06379v2.

[20] DEL BARRIO, E., GONZÁLEZ-SANZ, A. and LOUBES, J.-M. (2022). Central Limit Theorems for Semidiscrete Wasserstein Distances. arXiv:2202.06380.

[21] DEL BARRIO, E., GORDALIZA, P. and LOUBES, J.-M. (2019). A central limit theorem for Lp transportation cost on the real line with application to fairness assessment in machine learning. *Information and Inference: A Journal of the IMA* **8**. MR4045479

[22] DEL BARRIO, E. and LOUBES, J.-M. (2019). Central limit theorems for empirical transportation cost in general dimension. *The Annals of Probability* **47** 926–951. MR3916938

[23] DELON, J., SALOMON, J. and SOBOLEVSKI, A. (2010). Fast transport optimization for Monge cost on the circle. *SIAM Journal on Applied Mathematics* **70** 2239–2258. MR2678036

[24] FASANO, G. and FRANCESCHINI, A. (1987). A multidimensional version of the Kolmogorov–Smirnov test. *Monthly Notices of the Royal Astronomical Society* **225** 155–170.

[25] FLORY, P. J. and VOLKENSTEIN, M. (1969). Statistical mechanics of chain molecules. *Biopolymers* **8** 699–700.

[26] FOURNIER, N. and GUILLIN, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields* **162** 707. MR3383341

[27] FREITAG, G., CZADO, C. and MUNK, A. (2007). A nonparametric test for similarity of marginals—With applications to the assessment of population bioequivalence. *Journal of Statistical Planning and Inference* **137** 697–711. Special Issue on Nonparametric Statistics and Related Topics: In honor of M.L. Puri. MR2301710

[28] GANGBO, W. and McCANN, R. J. (1996). The geometry of optimal transportation. *Acta Mathematica* **177** 113–161. MR1440931

[29] GIBRAT, J. F., GARNIER, J. and ROBSON, B. (1987). Further developments of protein secondary structure prediction using information theory:

New parameters and consideration of residue pairs. *Journal of Molecular Biology* **198** 425–443.

[30] HALLIN, M., MORDANT, G. and SEGERS, J. (2021). Multivariate goodness-of-fit tests based on Wasserstein distance. *Electronic Journal of Statistics* **15** 1328–1371. MR4255302

[31] HOLM, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* **6** 65–70. MR0538597

[32] HUNDRIESER, S., KLATT, M. and MUNK, A. (2022). The Statistics of Circular Optimal Transport. In *Forum for Interdisciplinary Mathematics* 57–82. Springer Nature Singapore. MR4461115

[33] HUNDRIESER, S., KLATT, M., STAUDT, T. and MUNK, A. (2022). A Unifying Approach to Distributional Limits for Empirical Optimal Transport. *arXiv preprint.*

[34] HUNDRIESER, S., STAUDT, T. and MUNK, A. (2022). Empirical Optimal Transport between Different Measures Adapts to Lower Complexity. arXiv:2202.10434.

[35] KABAT, E. and WU, T. (1973). The influence of nearest-neighbor amino acids on the conformation of the middle amino acid in proteins: comparison of predicted and experimental determination of -sheets in concanavalin A. *Proc Natl Acad Sci U S A.* **70** 1473–1477.

[36] LANG, U. and SCHROEDER, V. (1997). Kirszbraun's Theorem and Metric Spaces of Bounded Curvature. *Geometric & Functional Analysis GAFA* **7** 535–560. MR1466337

[37] LILJAS, A., LILJAS, L., PISKUR, J., LINDBLOM, G., NISSEN, P. and KJELDGAARD, M. (2009). *Textbook Of Structural Biology.* World Scientific Publishing, Singapore.

[38] LOVELL, S. C., DAVIS, I. W., ARENDALL III, W. B., DE BAKKER, P. I. W., WORD, J. M., PRISANT, M. G., RICHARDSON, J. S. and RICHARDSON, D. C. (2003). Structure validation by C$\alpha$ geometry: $\phi$, $\psi$ and C$\beta$ deviation. *Proteins: Structure, Function, and Bioinformatics* **50** 437–450.

[39] MANOLE, T., BALAKRISHNAN, S., NILES-WEED, J. and WASSERMAN, L. (2021). Plugin Estimation of Smooth Optimal Transport Maps.

[40] MANOLE, T. and NILES-WEED, J. (2021). Sharp Convergence Rates for Empirical Optimal Transport with Smooth Costs. *arXiv preprint.*

[41] MARDIA, K. V., TAYLOR, C. C. and SUBRAMANIAM, G. K. (2007). Protein Bioinformatics and Mixtures of Bivariate von Mises Distributions for Angular Data. *Biometrics* **63** 505–512. MR2370809

[42] MCCANN, R. (2001). Polar factorization of maps on Riemannian manifolds. *GAFA, Geom. funct. anal.* **11** 589–608. MR1844080

[43] MCCANN, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal* **80** 309–323. MR1369395

[44] MCDIARMID, C. (1989). *On the method of bounded differences.* In *Surveys in Combinatorics, 1989: Invited Papers at the Twelfth British Combinatorial Conference. London Mathematical Society Lecture Note Series* 148–188. Cambridge University Press. MR1036755

[45] Mena, G. and Weed, J. (2019). Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. In *NeurIPS*.

[46] Morris, A. L., MacArthur, M. W., Hutchinson, E. G. and Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. *Proteins: Structure, Function, and Bioinformatics* **12** 345–364.

[47] Munk, A. and Czado, C. (1998). Nonparametric validation of similar distributions and assessment of goodness of fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **60**. MR1625620

[48] Estaña, A., Sibille, N., Delaforge, E., Vaisset, M., Cortés, J. and Bernadó, P. (2019). Realistic Ensemble Models of Intrinsically Disordered Proteins Using a Structure-Encoding Coil Database. *Structure* **27** 381–391.e2.

[49] Petroni, N. C. (2019). Taking rational numbers at random. arXiv:1908.06944v1.

[50] Peyré, G. and Cuturi, M. (2019). Computational Optimal Transport: With Applications to Data Science. *Foundations and Trends® in Machine Learning* **11** 355–607.

[51] Phipson, B. and Smyth, G. K. (2010). Permutation P-values Should Never Be Zero: Calculating Exact P-values When Permutations Are Randomly Drawn. *Statistical Applications in Genetics and Molecular Biology* **9**. MR2746025

[52] Rabin, J., Delon, J. and Gousseau, Y. (2009). Transportation Distances on the Circle. *Journal of Mathematical Imaging and Vision* **41**. MR2824152

[53] Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology* **7** 95–99.

[54] Ramdas, A., Garcia, N. and Cuturi, M. (2015). On Wasserstein Two Sample Testing and Related Families of Nonparametric Tests. *Entropy* **19**. MR3608466

[55] Rata, I., Li, Y. and Jakobsson, E. (2010). Backbone statistical potential from local sequence-structure interactions in protein loops. *J Phys Chem B* **114** 1859–1869.

[56] Rockafellar, R. T. (1970). *Convex Analysis.* Princeton University Press. MR0274683

[57] Santambrogio, F. (2015). Optimal Transport for Applied Mathematicians. Calculus of Variations, PDEs and Modeling. MR3409718

[58] Schuhmacher, D., Bähre, B., Gottschlich, C., Hartmann, V., Heinemann, F. and Schmitzer, B. (2020). transport: Computation of Optimal Transport Plans and Wasserstein Distances R package version 0.12-2.

[59] Serrurier, M., Mamalet, F., Gonzalez-Sanz, A., Boissin, T., Loubes, J.-M. and del Barrio, E. (2021). Achieving Robustness in Classification Using Optimal Transport With Hinge Regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 505–514.

[60] Shen, Y., Roche, J., Grishaev, A. and Bax, A. (2018). Prediction of nearest neighbor effects on backbone torsion angles and NMR scalar coupling constants in disordered proteins. *Protein Science* **27** 146–158.

[61] Sommerfeld, M. and Munk, A. (2018). Inference for empirical Wasserstein distances on finite spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80** 219–238. MR3744719

[62] Staudt, T., Hundrieser, S. and Munk, A. (2022). On the Uniqueness of Kantorovich Potentials. *arXiv preprint*.

[63] Ting, D., Wang, G., Shapovalov, M., Mitra, R., Jordan, M. and Dunbrack, R. (2010). Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical Dirichlet process model. *PLoS computational biology* **6** e1000763. MR2652813

[64] Vaart, A. W. and Wellner, J. A. (1996). Weak convergence. In *Weak convergence and empirical processes* 16–28. Springer. MR1385671

[65] Villani, C. (2003). *Topics in Optimal Transportation.* American mathematical society, Providence, Rhode Island. MR1964483

[66] Villani, C. (2008). *Optimal Transport: Old and New.* Springer-Verlag Berlin Heidelberg. MR2459454

[67] Weed, J. and Bach, F. (2019). Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli* **25** 2620–2648. MR4003560