The Annals of Statistics
2022, Vol. 50, No. 2, 1012–1037
https://doi.org/10.1214/21-AOS2136

MULTIVARIATE RANKS AND QUANTILES USING OPTIMAL TRANSPORT: CONSISTENCY, RATES AND NONPARAMETRIC TESTING

BY PROMIT GHOSAL[1,a] AND BODHISATTVA SEN[2,b]

[1]*Department of Mathematics, Massachusetts Institute of Technology,* [a]*promit@mit.edu*
[2]*Department of Statistics, Columbia University,* [b]*bodhi@stat.columbia.edu*

In this paper, we study multivariate ranks and quantiles, defined using the theory of optimal transport, and build on the work of Chernozhukov et al. (*Ann. Statist.* **45** (2017) 223–256) and Hallin et al. (*Ann. Statist.* **49** (2021) 1139–1165). We study the characterization, computation and properties of the multivariate rank and quantile functions and their empirical counterparts. We derive the uniform consistency of these empirical estimates to their population versions, under certain assumptions. In fact, we prove a Glivenko–Cantelli type theorem that shows the asymptotic stability of the empirical rank map in any direction. Under mild structural assumptions, we provide global and local rates of convergence of the empirical quantile and rank maps. We also provide a sub-Gaussian tail bound for the global L_2-loss of the empirical quantile function. Further, we propose tuning parameter-free multivariate nonparametric tests—a two-sample test and a test for mutual independence—based on our notion of multivariate quantiles/ranks. Asymptotic consistency of these tests are shown and the rates of convergence of the associated test statistics are derived, both under the null and alternative hypotheses.

1. Introduction. Suppose that X is a random vector in $\mathbb{R}^d$, for $d \geq 1$, with distribution ν. When $d = 1$, the rank and quantile functions of X are defined as F and F^{-1} (the inverse[1] of F), respectively, where F is the cumulative distribution function of X. Moreover, when $d = 1$, quantile and rank functions and their empirical counterparts are ubiquitous in statistics and form the backbone of what is now known as classical nonparametrics (see, e.g., [50] and the references therein) and are important tools for inference (see, e.g., [40] and the references therein). In this paper, we study many properties of *multivariate* (empirical) ranks and quantiles defined using the theory of optimal transport (OT), as introduced in [18].

Unlike the real line, the d-dimensional Euclidean space $\mathbb{R}^d$, for $d \geq 2$, has no natural ordering. This has been a major impediment in defining analogues of quantiles and ranks in $\mathbb{R}^d$, for $d \geq 2$. Several notions of multivariate quantiles have been proposed in the statistical literature—some based on data depth ideas (see, e.g., [54, 60, 87]) and some based on geometric ideas (see, e.g., [17, 39, 49]); see [72] and [37] for recent surveys on this topic. However, most of these notions do not enjoy the numerous appealing properties that make univariate ranks and quantiles so useful. For example, most of these notions can lead to multivariate quantiles that may take values outside the support of the distribution ν.

To motivate the notions of ranks and quantiles based on the theory of OT (the subject of our study), let us first consider the case when $d = 1$. Suppose that $X \sim \nu$ has a continuous distribution function F. An important property of the one-dimensional rank function F is that

Received September 2020; revised September 2021.
MSC2020 subject classifications. Primary 62G30, 62G20; secondary 60F15, 35J96.
Key words and phrases. Brenier–McCann's theorem, convergence of subdifferentials of convex functions, Glivenko–Cantelli type theorem, Legendre–Fenchel dual, local uniform rate of convergence, semidiscrete optimal transport, testing mutual independence, two-sample goodness-of-fit testing.

[1] $F^{-1}(p) := \inf\{x \in \mathbb{R} : p \leq F(x)\}$.

$F(X) \sim \mu$ where $\mu \equiv \text{Uniform}([0,1])$, that is, F *transports* (see (6) for the formal definition) the distribution ν to μ. Similarly, the quantile function F^{-1} (which is the inverse of the rank map) transports μ to ν, that is, $F^{-1}(U) \sim X$ where $U \sim \mu$. In fact, it can be easily shown that the quantile function F^{-1} (or F) is the unique monotone nondecreasing map that transports μ to ν (or ν to μ). Moreover, if ν has finite second moment, it can be shown that F^{-1} is the almost everywhere (a.e.) unique map (on $[0,1]$) that transports μ to ν and minimizes the expected squared-error cost, that is,

$$F^{-1} = \arg \min_{T:T(U)\sim\nu} \mathbb{E}\big[\big(U - T(U)\big)^2\big], \quad \text{where } U \sim \mu \tag{1}$$

and the minimization is over all functions T that transport μ to ν (and thus the connection to OT); see Section 3 for the details. The rank function F also minimizes the expected squared-error cost where now one considers maps that transport ν to μ.

The multivariate quantile and rank functions using OT essentially extend the above properties of univariate rank and quantile functions. Now let μ be an absolutely continuous probability measure with respect to (w.r.t.) Lebesgue measure on $\mathbb{R}^d$ ($d \geq 1$) and supported on a compact convex set $\mathcal{S}$; for example, we can take μ to be $\text{Uniform}([0,1]^d)$ or uniform on the ball of radius one around $0 \in \mathbb{R}^d$. We often refer to μ as the *reference distribution* and will define quantiles relative to this reference measure (when $d = 1$ we usually take μ to be $\text{Uniform}([0,1])$). Let ν be another probability measure in $\mathbb{R}^d$ which we term as the *target distribution*; we think of ν as the population distribution of the observed data. We define the *multivariate quantile function* $Q : \mathcal{S} \to \mathbb{R}^d$ of ν w.r.t. μ as the solution to the following optimization problem:

$$Q := \arg \min_{T:T(U)\sim\nu} \mathbb{E}\big[\|U - T(U)\|^2\big], \quad \text{where } U \sim \mu, \tag{2}$$

and the minimization is over all functions $T : \mathcal{S} \to \mathbb{R}^d$ that transport μ to ν; cf. (1) and see Section 3 for the details. Here, $\|\cdot\|$ denotes the usual Euclidean norm in $\mathbb{R}^d$. Moreover, if ν does not have a finite second moment, the above optimization problem might not be meaningful but the notion of multivariate quantiles (using OT) can still be defined as follows. By the Brenier–McCann's theorem (see Theorem 2.2), there exists an a.e. unique map $Q : \mathcal{S} \to \mathbb{R}^d$—which we define as the quantile function of ν (w.r.t. the reference measure μ)—that is the gradient of a convex function and transports μ to ν; that is, $Q(U) \sim \nu$ where $U \sim \mu$. Further, it is known that when (2) is meaningful, the above two notions yield the same function Q. Note that when $d = 1$, the gradient of a convex function is a monotone nondecreasing function, and thus the above two characterizations of the quantile function Q are the exact analogues of the one-dimensional case described in the previous paragraph.

Although the rank function can be intuitively thought of as the inverse of the quantile function, such an inverse might not always exist—especially when ν is a discrete probability measure (which arises naturally when defining the empirical rank map). In Section 3, we tackle this issue and use the notion of the Legendre–Fenchel transform (see Section 2) to formally define the rank function. Indeed, if the reference and the target distributions are absolutely continuous, this notion of rank function is the inverse of the quantile function almost everywhere (a.e.). Furthermore, it can be shown that (see Proposition 3.1; also see [20], Theorem 1), under mild regularity conditions, the quantile and rank functions are continuous bijections (i.e., homeomorphisms) between the (interiors of the) supports of the reference and target distributions and they are inverses of each other. It is worth noting that when $d = 1$, a continuous bijective rank map corresponds to the distribution function being continuous and strictly increasing.

In Section 3, we describe some important properties of the defined multivariate quantile and rank functions; also see Section A.2 (of the Supplementary Material [33]). For example, in Lemma 3.2 we show that, under appropriate conditions, the rank map approaches a limit along every ray that depends only on the geometry of $\mathcal{S}$ (and not on ν); this plays a crucial role in proving the uniform convergence result for the empirical rank map in Theorem 4.1. Some useful properties of the multivariate quantile and rank functions, including: (i) equivariance under orthogonal transformations when the reference distribution is spherically symmetric, and (ii) decomposition/splitting into marginal quantile/rank functions when $X \sim \nu$ has mutually independent subvectors and $\mu = \text{Uniform}([0,1]^d)$, are given in Section A.3 of the Supplementary Material [33]. Thus the choice of the reference distribution μ affects the properties of the multivariate ranks/quantiles. In practice, this choice should be dictated by the application at hand; see Remark 3.11 for a discussion on this and some preliminary guidelines on the choice of μ. A complete study of the pros and cons of different reference distributions is an important and open research question beyond the scope of this paper.

Given n i.i.d. random vectors $X_1, \dots, X_n \sim \nu$ in $\mathbb{R}^d$, in Sections 3.1, we discuss the characterization and properties of the empirical quantile and rank maps, which are defined via (2) but with ν replaced by the empirical distribution of the data. Thus, the computation of the empirical quantile map reduces to a semidiscrete OT problem; see Section 3.2 for the details where we show that the empirical quantile map can be computed by solving a convex optimization problem with n variables. An attractive property of the empirical ranks, when $d = 1$, that makes ranks useful for statistical inference is that they are distribution-free. Lemma 3.4 shows that a distribution-free version of empirical multivariate ranks can be obtained by external randomization (also see Lemma 6.1). Although our approach of defining multivariate quantiles/ranks via the theory of OT has many similarities with those of [18, 37] and [11], there are subtle and important differences; in Section 3.3 we discuss these connections.

The main statistical contributions of this paper are divided in the three sections: Sections 4, 5 and 6. In the following, we highlight some of the main results in these sections and their novelties.

(*I*) *Uniform convergence of empirical quantile/rank maps*: In Section 4, we state our first main theoretical result on the almost sure (a.s.) uniform convergence of the empirical quantile and rank maps to their population counterparts. An informal statement of this result (Theorem 4.1) is given below. Suppose that μ is supported on a compact convex set $\mathcal{S} \subset \mathbb{R}^d$ with nonempty interior. Let $\mathcal{Y}$ be the support of ν and let $\{\widehat{\nu}_n\}_{n\geq 1}$ be a sequence of random probability distributions converging weakly to ν a.s. Suppose that the quantile map Q of ν (w.r.t. μ) is a continuous bijection from $\text{Int}(\mathcal{S})$ (the interior of $\mathcal{S}$) to $\text{Int}(\mathcal{Y})$. Then, with probability (w.p.) 1, the empirical quantile and rank maps corresponding to $\widehat{\nu}_n$ (w.r.t. μ)—$\widehat{Q}_n$ and $\widehat{R}_n$—converge uniformly to Q and $R \equiv Q^{-1}$, respectively, over compacts inside $\text{Int}(\mathcal{S})$ and $\text{Int}(\mathcal{Y})$. Moreover, if $\mathcal{S} \subset \mathbb{R}^d$ is a strictly convex set (see Definition 2.3) then $\widehat{R}_n$ converges uniformly to $R = Q^{-1}$ over the whole of $\mathbb{R}^d$ a.s.; furthermore, w.p. 1, the tail limit of $\widehat{R}_n$ stabilizes along any direction. We mention below two main novelties of the above result.

(*a*) One of the main consequences of Theorem 4.1 is the a.s. convergence of the empirical rank function $\widehat{R}_n$ on the whole of $\mathbb{R}^d$, under the strong convexity condition on the support $\mathcal{S}$ of μ. This can indeed be thought of as a generalization of the famous Glivenko–Cantelli theorem for rank functions when $d > 1$. Moreover, our result does not need any boundedness assumption on the support of ν and even applies when the second moment of ν is not finite. This is a major improvement over the corresponding results in [18], Theorem 3.1 and [11], Theorem 2.3. Furthermore, unlike in [37], μ can be any absolutely continuous distribution supported on a compact convex domain with minor restrictions on its boundary. Note that for Theorem 4.1 to hold we need to assume that Q is a homeomorphism; in particular, if ν has a convex support with a bounded density then the above holds; see, for example, Proposition 3.1.

(*b*) Our result (see (22) of Theorem 4.1) implies that when the population rank map is a homeomorphism, the tail limits of the estimated rank maps $\widehat{R}_n$ depend neither on ν nor on μ;

rather they depend on the geometry of $\mathcal{S}$—the support of the reference distribution μ. This is reminiscent of the case when $d = 1$ where the limits of the distribution (rank) function toward $-\infty$ and $+\infty$ are always 0 and 1, respectively (irrespective of ν).

(*II*) *Rate of convergence of empirical quantile/rank maps*: Theorem 4.1 naturally leads to the question: "What are the rates of convergence of the empirical quantile/rank maps—$\widehat{Q}_n$ and $\widehat{R}_n$"? We study this question in detail in Section 5. We first introduce the following notation:

$$r_{d,n} := \begin{cases} n^{-1/2} & d = 1, 2, 3, \\ n^{-1/2} \log n & d = 4, \\ n^{-2/d} & d > 4. \end{cases} \tag{3}$$

(*a*) In Theorem 5.2, we provide upper bounds on the L_2-global risk of the empirical quantile map $\widehat{Q}_n$. In particular, we show that, for all $n \geq 1$,

$$\mathbb{E}\left[\int \|\widehat{Q}_n - Q\|^2 \, d\mu\right] \leq C r_{d,n},$$

where $C > 0$ is a constant that depends only on μ and ν. This result is proved using Lemma 5.1, which is of independent interest, and gives a quantitative stability estimate for OT maps in the semidiscrete setting. Note that the rates obtained in Theorem 5.2 are strictly better than those obtained for OT maps in [81], Theorem 1.1, and [53], Section 4. Furthermore, in Theorem 5.2 we also give a sub-Gaussian tail bound for $\int \|\widehat{Q}_n - Q\|^2 \, d\mu$. We believe that Theorem 5.2 gives the exact rate of convergence for the empirical quantile map $\widehat{Q}_n$ when $d > 4$; see [43] where the conjecture of this optimality of the rate $n^{-2/d}$ (for $d > 4$) is made. Furthermore, Theorem 5.2 holds under minimal structural assumptions on ν—we only assume strong convexity of the underlying potential function (see (7) below).

(*b*) In Theorem 5.3, under appropriate assumptions, we give an upper bound on the risk of the sample rank map, that is, we show that, for all $n \geq 1$,

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \|\widehat{R}_n(X_i) - R(X_i)\|^2\right] \leq K r_{d,n},$$

where $K > 0$ is a constant that depends only on μ and ν. Deriving such a rate result for the multivariate sample rank map $\widehat{R}_n$ is a bit more tricky as $\widehat{R}_n$ is not an OT map per se, but is defined via the Legendre–Fenchel transform (see Section 3 for the details).

(*c*) We address the local uniform rate of convergence of the empirical quantile and rank maps in Theorem 5.4. The pointwise rate of convergence of empirical rank/quantile maps, defined via the theory of OT, is indeed a hard problem when $d > 1$ and not much is known in the literature. Under similar assumptions as in Theorems 5.2 and 5.3, we show that $\widehat{Q}_n$ and $\widehat{R}_n$ converge locally uniformly to Q and R, respectively, at the rate $r_{d,n}^{1/(d+2)}$. We consider Theorem 5.4 as a first step toward understanding the local behavior of transport maps. The proof of this result uses Theorem 5.2 and a correspondence result between the local uniform and local L_2 rates of convergence of the empirical rank and quantile functions (see Proposition F.1 in the Supplementary Material [33]) that could be of independent interest.

(*III*) *Applications to nonparametric testing*: In Section 6, we investigate two statistical applications of the multivariate rank and quantile functions studied in this paper—we propose methodology for multivariate two-sample goodness-of-fit testing (in Section 6.1) and testing for mutual independence (in Section 6.2). Both of the proposed tests are tuning parameter-free. Applying the uniform convergence results of Theorem 4.1, we prove the consistency of these proposed tests, that is, the power of these tests converges to 1 under fairly general assumptions on the underlying distributions (see Propositions 6.2 and 6.5). Moreover, using the results in Section 5 we provide rates of convergence of the test statistics (for both the testing problems), under both the null and alternative hypotheses; see Propositions 6.3, 6.4, 6.6 and 6.7. This leads to omnibus consistent nonparametric tests that are computationally feasible, and being rank based, do not depend on moment assumptions on the underlying distribution(s).

Although we state most of our results in terms of multivariate quantile and rank functions, many of the results have immediate implications in estimation of OT maps. Indeed, in recent years there has been a deluge of work at the intersection of statistics and the theory of OT; see, for example, [27, 28, 48, 62, 63, 65–67, 82] and the references therein.

The paper is organized as follows. We introduce notation and some basic notions from convex analysis and the theory of OT in Section 2. Section 3 defines the multivariate quantile and rank maps and their empirical counterparts and investigates some of their properties, including computation. The asymptotic results on the uniform a.s. convergence of the empirical quantile and rank maps are given in Section 4. Global and local rates of convergence of the empirical quantile/rank maps are given in Section 5. The two statistical applications in nonparametric testing are given in Section 6. Due to space constraints, all the proofs of the main results, additional (technical) results, further remarks and discussions are relegated to the Supplementary Material [33].

2. Preliminaries.

We start with some notation and recall some important concepts from convex analysis that will be relevant for the rest of the paper. For $u, v \in \mathbb{R}^d$, we use $\langle u, v\rangle$ to denote the dot product of u and v and $\|\cdot\|$ denotes the usual Euclidean norm in $\mathbb{R}^d$. For $y_1, \ldots, y_k \in \mathbb{R}^d$, we write $\mathrm{Conv}(y_1, \ldots, y_k)$ to denote the convex hull of $\{y_1, \ldots, y_k\} \subset \mathbb{R}^d$. A *convex polyhedron* is the intersection of finitely many closed half-spaces. A *convex polytope* is the convex hull of a finite set of points. The interior, closure and boundary of a set $\mathcal{X} \subset \mathbb{R}^d$ will be denoted by $\mathrm{Int}(\mathcal{X})$, $\mathrm{Cl}(\mathcal{X})$, and $\mathrm{Bd}(\mathcal{X})$, respectively. The Dirac delta measure at x is denoted by δ_x. For $\delta > 0$ and $x \in \mathbb{R}^d$, $B_\delta(x) := \{y \in \mathbb{R}^d : \|y - x\| < \delta\}$ denotes the open ball of radius δ around x. The set of natural numbers will be denoted by $\mathbb{N}$.

The *domain* of a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, denoted by $\mathrm{dom}(f)$, is the set $\{x \in \mathbb{R}^d : f(x) < +\infty\}$. A function f is called *proper* if $\mathrm{dom}(f) \neq \varnothing$. A function $f \in L^\infty(\mathcal{S})$, where $\mathcal{S} \subset \mathbb{R}^d$, if $\sup_{x\in\mathcal{S}} |f(x)| < \infty$. We say that f is lower semicontinuous (l.s.c.) at $x_0 \in \mathbb{R}^d$ if $\liminf_{x\to x_0} f(x) \geq f(x_0)$. For a proper function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, the *Legendre–Fenchel dual* (or convex conjugate or simply the dual) of f is the proper function $f^* : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f^*(y) := \sup_{x\in\mathbb{R}^d} \{\langle x, y\rangle - f(x)\}, \quad \text{for all } y \in \mathbb{R}^d.$$

It is well known that f^* is a proper, l.s.c. convex function. The Legendre–Fenchel duality theorem says that for a proper l.s.c. convex function f, $(f^*)^* = f$. Throughout the paper, we will assume that all the convex functions that we will be dealing with are l.s.c.

Given a convex function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, we define the *subdifferential* set of f at $x \in \mathrm{dom}(f)$ by

$$\partial f(x) := \{\xi \in \mathbb{R}^d : f(x) + \langle y - x, \xi\rangle \leq f(y), \text{ for all } y \in \mathbb{R}^d\}.$$

Any element in $\partial f(x)$ is called a *subgradient* of f at x. The subdifferential $\partial f(x)$ is empty if $f(x) = +\infty$ and nonempty if $x \in \mathrm{Int}(\mathrm{dom}(f))$. If f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$. A convex function is a.e. differentiable (w.r.t. Lebesgue measure) on $\mathrm{Int}(\mathrm{dom}(f))$. As a consequence, a convex function is continuous in the interior of its domain. For a convex function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, we sometimes just write $\nabla f(x)$ to denote the (sub)differential of f at x with the understanding that when f is not differentiable at x we can take $\nabla f(x)$ to be any point in the set $\partial f(x)$. This avoids the need to deal with the set-valued function ∂f. However, sometimes we will need to view ∂f as a multivalued mapping, that is, a mapping from $\mathbb{R}^d$ into the power set of $\mathbb{R}^d$, and we will use the notation ∂f in that case. We will find the following results useful (see, e.g., [79], Proposition 2.4).

LEMMA 2.1 (Characterization of subdifferential). *Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a proper l.s.c. convex function. Then for all $x, y \in \mathbb{R}^d$,*

$$\langle x, y\rangle = f(x) + f^*(y) \quad \Longleftrightarrow \quad y \in \partial f(x) \quad \Longleftrightarrow \quad x \in \partial f^*(y). \tag{4}$$

Lemma 2.1 shows a one-to-one relation between the subdifferential set of a convex function and its Legendre–Fenchel dual.

DEFINITION 2.1 (Strongly convex function). A function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is *strongly convex* with parameter $\lambda > 0$ if for all $x, y \in \mathrm{dom}(f)$,

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{\lambda}{2}\|y - x\|^2.$$

DEFINITION 2.2 (Set convergence). Let $K_1 \subset K_2 \subset \dots$ be an increasing sequence of sets in $\mathbb{R}^d$. We say that K_n *increases* to $K \subset \mathbb{R}^d$, and write $K_n \uparrow K$, if for any compact set $A \subset \mathrm{Int}(K)$ there exists $n_0 = n_0(A) \in \mathbb{N}$ such that $A \subseteq K_n$ for all $n \geq n_0$.

The above notion is slightly stronger than just assuming $K_1 \subset K_2 \subset \dots$ and $\liminf_{n\to\infty} K_n = K$.

A supporting hyperplane of a closed convex set $S \subset \mathbb{R}^d$ is a hyperplane that has both of the following two properties: (i) S is entirely contained in one of the two closed half-spaces bounded by the hyperplane and (ii) S intersected with the hyperplane is nonempty.

DEFINITION 2.3 (Strictly convex set). A convex set $S \subset \mathbb{R}^d$ is said to be strictly convex if any supporting hyperplane to $\mathrm{Cl}(S)$ touches $\mathrm{Cl}(S)$ at only one point.

Let μ and ν be two Borel probability measures supported on $\mathcal{S} \subset \mathbb{R}^d$ and $\mathcal{Y} \subset \mathbb{R}^d$, respectively. The goal of OT (Monge's problem), under the squared Euclidean loss, is to find a measurable transport map $T \equiv T_{\mu;\nu} : \mathcal{S} \to \mathcal{Y}$ solving the (constrained) minimization problem

$$\inf_T \int \|u - T(u)\|^2 \, d\mu(u) \quad \text{subject to } T\#\mu = \nu \tag{5}$$

where the minimization is over T (a *transport map*), a measurable map from $\mathcal{S}$ to $\mathcal{Y}$, and $T\#\mu$ is the *push forward* of μ by T, that is,

$$T\#\mu(B) = \mu\big(T^{-1}(B)\big), \quad \text{for all } B \subset \mathcal{Y} \text{ Borel.} \tag{6}$$

A map $T_{\mu;\nu}$ that attains the infimum in (5) is called an OT map from μ to ν. We state an important result in this theory, namely Brenier–McCann's theorem ([12, 56]). This result will be very useful to us; see Section A.1 in the Supplementary Material [33] for a brief introduction to the field of OT.

THEOREM 2.2 (Brenier–McCann theorem). *Let μ and ν be two Borel probability measures on $\mathbb{R}^d$. Suppose further that μ has a Lebesgue density. Then there exists a convex function $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ whose gradient $G = \nabla\psi : \mathbb{R}^d \to \mathbb{R}^d$ pushes μ forward to ν. In fact, there exists only one such G that arises as the gradient of a convex function, that is, G is unique μ-a.e. Moreover, if μ and ν have finite second moments, G uniquely minimizes Monge's problem* (5).

3. Quantile and rank maps in $\mathbb{R}^d$ when $d \geq 1$. Suppose that $X \sim \nu$ is supported on $\mathcal{Y} \subset \mathbb{R}^d$. Let μ be a known absolutely continuous distribution on $\mathbb{R}^d$ (i.e., μ has a density w.r.t. Lebesgue measure on $\mathbb{R}^d$) with support $\mathcal{S}$—a compact convex subset of $\mathbb{R}^d$ with nonempty interior; for example, we can take μ to be Uniform($[0, 1]^d$). Other natural choices of μ are the uniform distribution on the unit ball $B_1(0)$ in $\mathbb{R}^d$ [18], and the spherical uniform distribution (V has the spherical uniform distribution if $V = L\varphi$ where φ is uniformly distributed on the unit sphere around $0 \in \mathbb{R}^d$ and $L \sim$ Uniform([0, 1]), and L and φ are mutually independent); see [30, 37].

In the following, we define the multivariate *quantile* and *rank* maps for ν w.r.t. the distribution μ using the theory of OT. We first define the quantile function for ν and then use it to define the rank map. Our approach is essentially the same as outlined in [18] although there are some important and subtle differences; see Section 3.3 for a discussion.

DEFINITION 3.1 (Quantile function). The quantile function of the probability measure ν (w.r.t. μ) is defined as the μ-a.e. unique map $Q : \mathcal{S} \to \mathbb{R}^d$, which pushes μ to ν and has the form

$$Q := \nabla \psi, \tag{7}$$

where $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is convex. We call ψ a potential function.

REMARK 3.2 (Uniqueness of Q). As the convex function ψ in Definition 3.1 need not be differentiable everywhere, there is a slight ambiguity in the definition of Q. When ψ is not differentiable, say at $u \in \mathcal{S}$, we can define $Q(u)$ to be any element of the subdifferential set $\partial\psi(u)$ (see Section 2 for its formal definition). As a convex function is differentiable a.e. (on its domain) this convention does not affect the μ-a.e. uniqueness of Q. Further, this convention bypasses the need to define quantiles as a multivalued map.

The existence and μ-a.e. uniqueness of the quantile map $Q(\cdot)$, for any probability measure ν on $\mathbb{R}^d$, is guaranteed by the Brenier–McCann theorem (see Theorem 2.2). Further, by Theorem 2.2, if ν has finite second moment, then $Q(\cdot)$ can be expressed as in (2). As discussed in the Introduction, the above notion of quantiles obviously extend our usual definition of quantiles when $d = 1$; see Section A.4 of the Supplementary Material [33] for a more detailed discussion.

REMARK 3.3 (Nonuniqueness of ψ). Although Q is μ-a.e. unique it is easy to see that ψ (as in Definition 3.1) is not unique; in fact, $\psi(\cdot) + c$ where $c \in \mathbb{R}$ is a constant would also suffice (as $\partial(\psi + c) = \partial\psi$). Further, we can change $\psi(\cdot)$ outside the set $\mathcal{S}$ and this does not change Q (as Q has domain $\mathcal{S}$). For this reason, we will consider

$$\psi(u) = +\infty, \quad \text{for } u \in \mathbb{R}^d \setminus \mathcal{S}. \tag{8}$$

The above convention will be useful in the subsequent discussion.

DEFINITION 3.4 (Rank map). Recall the convex potential function $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ whose gradient yields the quantile map (see (7); also see (8)). We define the rank function $R : \mathbb{R}^d \to \mathcal{S}$ of ν (w.r.t. μ) as

$$R := \nabla \psi^*, \tag{9}$$

where $\psi^* : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is the Legendre–Fenchel dual of the convex function ψ, that is, $\psi^*(x) := \sup_{u \in \mathbb{R}^d} \{\langle x, u\rangle - \psi(u)\}$, for $x \in \mathbb{R}^d$. Note that ψ^* is also referred to as the dual potential of $Q \equiv \nabla\psi$.

A few remarks are in order now.

REMARK 3.5 (The domain of R). The rank map $R(x)$ is finite for all $x \in \mathbb{R}^d$; cf. the quantile map $Q(\cdot)$, which is μ-a.e. uniquely defined. This follows from the fact that $\psi^*(x) < \infty$ for every $x \in \mathbb{R}^d$; see Lemma A.3 in the Supplementary Material [33]. If ψ^* is not differentiable at x (say), we can define $R(x)$ to be any element in the subdifferential set $\partial\psi^*(x)$.

REMARK 3.6 (The range of the rank map). Using Lemma 2.1, one can argue that $R(x) \in \mathcal{S}$ for every $x \in \mathbb{R}^d$. This follows from the fact that $R(x) \in \partial\psi^*(x)$ exists for every $x \in \mathbb{R}^d$, and by (4), $u \in \partial\psi^*(x) \Leftrightarrow x \in \partial\psi(u)$. Note that as $\partial\psi(u)$ exists we must have $\psi(u) < +\infty$, which in turn implies that $u \in \mathcal{S}$ (as $\psi(u) = +\infty$, for $u \in \mathbb{R}^d \setminus \mathcal{S}$ by (8)).

REMARK 3.7 (When ψ^* is not differentiable). As $\psi^*(x) < \infty$ for every $x \in \mathbb{R}^d$, ψ^* has a gradient a.e. Thus, $R(x)$ is uniquely defined for a.e. x. For $x \in \mathbb{R}^d$, where $\psi^*(x)$ is not differentiable, $R(x)$ is not uniquely defined. Although for such an x we can define $R(x)$ to be any element in the subdifferential set $\partial\psi^*(x)$ (as was done in [18]), in Section 3.1.3 we give a randomized choice of $R(x)$ that leads to the map R having appealing statistical properties.

Absolute continuity of ν is a sufficient condition for the rank map R and the quantile map Q to be the essential inverses of one another, that is,

$$R \circ Q(u) = u, \quad \text{for } \mu\text{-a.e. } u, \quad \text{and} \quad Q \circ R(x) = x, \quad \text{for } \nu\text{-a.e. } x,$$

and $R\#\nu = \mu$ (see, e.g., [79], Theorem 2.12 and Corollary 2.3). This justifies the definition of R via (9). Observe that the rank map, as in Definition 3.4, clearly extends the notion of the distribution function beyond $d = 1$. There is an intimate connection between the quantile map and the celebrated Monge–Ampère differential equation; see, for example, [79], Lemma 4.6, (also see [13, 20, 21]). In Section A.2 (in the Supplementary Material [33]), we discuss a few other interesting properties of quantile/rank and potential functions.

Although we know that $R = Q^{-1}$ μ-a.e. when ν is absolutely continuous, one may ask if the equality holds everywhere (as opposed to a.e.). Several results have been obtained in this direction that provide sufficient conditions for such an equality. Caffarelli (see [13–15]) showed that when $\mathcal{S}$ and $\mathcal{Y}$ are two bounded convex sets in $\mathbb{R}^d$ and μ and ν are absolutely continuous with positive densities (on their supports), then the corresponding OT maps $T : \mathcal{S} \to \mathcal{Y}$ (such that $T\#\mu = \nu$) and $T^* : \mathcal{Y} \to \mathcal{S}$ (such that $T^*\#\nu = \mu$) are continuous homeomorphisms and $T^* = T^{-1}$ everywhere in $\mathcal{Y}$; see [80], pp. 317–323, for other sufficient conditions. In Proposition 3.1 below, we give another such sufficient condition that is particularly useful in statistical applications. As pointed out by an anonymous referee, the main result in the recent paper [20] implies Proposition 3.1(a); Proposition 3.1(b) can then be derived as a consequence. See [20] for a proof of the following result.

PROPOSITION 3.1 ([20], Theorem 1). *Let $\mathcal{S} \subset \mathbb{R}^d$ be a compact convex set and $\mathcal{Y} \subset \mathbb{R}^d$ be a convex set. Let μ and ν be two probability distributions supported on $\mathcal{S}$ and $\mathcal{Y}$, respectively, with Lebesgue densities $p_{\mathcal{S}}$ and $p_{\mathcal{Y}}$. Suppose that $p_{\mathcal{S}}, p_{\mathcal{S}}^{-1} \in L^\infty(\mathcal{S})$ and $p_{\mathcal{Y}} \in L^\infty(\mathcal{Y} \cap B_R)$ for any $R > 0$, where B_R is the ball of radius R centered at* 0. *Then*:

(a) $\nabla\psi^*$, *restricted to* $\mathrm{Int}(\mathcal{Y})$, *is a homeomorphism from* $\mathrm{Int}(\mathcal{Y})$ *to* $\mathrm{Int}(\mathcal{S})$.

(b) $\nabla\psi$ *is a homeomorphism from* $\mathrm{Int}(\mathcal{S})$ *to* $\mathrm{Int}(\mathcal{Y})$. *Furthermore, we have* $\nabla\psi = (\nabla\psi^*)^{-1}$ *in* $\mathrm{Int}(\mathcal{S})$.

REMARK 3.8 (Convexity of $\mathcal{S}$ and $\mathcal{Y}$). Convexity of the domains, $\mathcal{S}$ and $\mathcal{Y}$, is one of the important conditions for the existence of continuous transport maps. Cafarrelli constructed a counterexample (see, e.g., [80], pp. 283–285) where he showed that the transport map may fail to be continuous when the two measures are absolutely continuous with bounded densities on two smooth and simply connected nonconvex domains.

REMARK 3.9. The condition on the density $p_{\mathcal{Y}}$ in Proposition 3.1 does not necessarily require $\mathcal{Y}$ to be compact. For example, any unimodal density supported on a convex domain $\mathcal{Y} \subset \mathbb{R}^d$ belongs to $L^\infty(\mathcal{Y} \cap B_R)$ for any $R > 0$; in particular, this includes the family of all absolutely continuous multivariate normal distributions.

Similar to the univariate distribution function, the one-dimensional projection of the rank map R along any direction, is nondecreasing (see Lemma A.4 in the Supplementary Material [33] for a formal statement of this result). A univariate distribution function is not only nondecreasing but takes the value 0 or 1 as one approaches $-\infty$ or $+\infty$, irrespective of ν. Under mild assumptions on $\mathcal{S}$ and R, we show in the following lemma (proved in Section C.1 of [33]) that $R(\cdot)$ is continuous on the whole of $\mathbb{R}^d$ and it approaches a limit along every ray that depends only on the geometry of $\mathcal{S}$ and not on the measure ν.

LEMMA 3.2. *Let $\mathcal{S} \subset \mathbb{R}^d$ be a strictly convex compact set* (*as in Definition* 2.3). *Let μ and ν be two probability measures on $\mathcal{S}$ and $\mathcal{Y} \subset \mathbb{R}^d$, respectively, where $\mathcal{Y}$ has nonempty interior. Let R be the rank map of ν w.r.t. μ. Suppose that R is a homeomorphism from* $\operatorname{Int}(\mathcal{Y})$ *to* $\operatorname{Int}(\mathcal{S})$. *Then, for any $x \in \mathbb{R}^d$*, $\lim_{\lambda\to+\infty} R(\lambda x) = \arg\max_{v\in\mathcal{S}}\langle x, v\rangle$.

Note that the above required condition on $\mathcal{S}$ is certainly satisfied, for example, when $\mathcal{S}$ is the unit ball in $\mathbb{R}^d$, that is, $\mathcal{S} = B_1(0)$; unfortunately when $\mathcal{S} = [0, 1]^d$, the condition is not satisfied. Lemma 3.2 has a simple interpretation when $\mathcal{S} = B_1(0)$—in this case, $\lim_{\lambda\to+\infty} R(\lambda x) = \arg\max_{v\in\mathcal{S}}\langle x, v\rangle = \frac{x}{\|x\|}$ if $x \neq 0$; cf. [26], Corollary 3.1. This generalizes the fact that for a distribution function F on $\mathbb{R}$, $F(x) \to 1$ as $x \to +\infty$ and $F(x) \to 0$ as $x \to -\infty$.

3.1. *The sample quantile and rank maps.* As before, we fix an absolutely continuous distribution μ with compact convex support $\mathcal{S} \subset \mathbb{R}^d$. Given a random sample $X_1, \dots, X_n$ from a distribution ν (on $\mathbb{R}^d$), we now consider estimating the population quantile and rank maps Q and R, respectively (w.r.t. μ). We simply define the sample versions of the quantile and rank maps as those obtained by replacing the unknown distribution ν with its empirical counterpart $\widehat{\nu}_n$—the empirical distribution of the data, that is,

$$\widehat{\nu}_n(A) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}(A), \quad \text{for any Borel set } A \subset \mathbb{R}^d.$$

3.1.1. *Empirical quantile function.* By Theorem 2.2, there exists an μ-a.e. unique map $\widehat{Q}_n$, which pushes μ to $\widehat{\nu}_n$ and can be expressed as

$$\widehat{Q}_n = \nabla\widehat{\psi}_n, \tag{10}$$

where $\widehat{\psi}_n : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is a convex function. Further, by Theorem 2.2, $\widehat{Q}_n$ can be computed via

$$\widehat{Q}_n = \arg\min_{T} \int \|u - T(u)\|^2\, d\mu(u) \quad \text{subject to} \quad T\#\mu = \widehat{\nu}_n. \tag{11}$$

Note that $\widehat{Q}_n \equiv \nabla\widehat{\psi}_n$ is μ-a.e. unique; when $\widehat{\psi}_n$ is not differentiable, say at u, we can define $\widehat{Q}_n(u)$ to be any point in $\partial\widehat{\psi}_n(u)$. As $\widehat{Q}_n \equiv \nabla\widehat{\psi}_n$ pushes μ to $\widehat{\nu}_n$, $\widehat{\psi}_n$ is a convex function whose gradient takes μ-a.e. finitely many values—in the set $\mathcal{X} := \{X_1, \dots, X_n\}$. Thus $\widehat{\psi}_n$ is piecewise linear (affine), and hence, there exists $\widehat{h} = (\widehat{h}_1, \dots, \widehat{h}_n) \in \mathbb{R}^n$ (unique up to adding a scalar multiple of $(1, \dots, 1) \in \mathbb{R}^n$) such that $\widehat{\psi}_n : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ can be represented as

$$\widehat{\psi}_n(u) := \begin{cases} \max_{i=1,\dots,n} \{u^\top X_i + \widehat{h}_i\} & \text{for } u \in \mathcal{S}, \\ +\infty & \text{for } u \notin \mathcal{S}. \end{cases} \tag{12}$$

The vector $\widehat{h}$ can be computed by solving a convex optimization problem; see Section 3.2 for the details.

REMARK 3.10 (Form of the subdifferential set $\partial\widehat{\psi}_n(u)$). As $\widehat{\psi}_n$ is piecewise linear (affine) and convex (and thus a finite pointwise maximum of affine functions), we can explicitly write its subdifferential, that is, for any $u \in \mathcal{S}$,

$$\partial\widehat{\psi}_n(u) = \mathrm{Conv}\big(\{X_i : \langle u, X_i\rangle + \widehat{h}_i = \widehat{\psi}_n(u)\}\big).$$

As $\widehat{Q}_n(u) \in \partial\widehat{\psi}_n(u)$, for any $u \in \mathrm{Int}(\mathcal{S})$, $\widehat{Q}_n(u)$ belongs to the convex hull of the data. The function $\widehat{Q}_n = \nabla\widehat{\psi}_n$ induces a cell decomposition of $\mathcal{S}$: Each cell is a convex set and is defined as

$$W_i(\widehat{h}) := \{u \in \mathcal{S} : \nabla\widehat{\psi}_n(u) = X_i\}. \tag{13}$$

In defining $W_i(\widehat{h})$, we only consider points $u \in \mathcal{S}$ where $\widehat{\psi}_n$ is differentiable; see [36] for more details. Note that, for a.e. sequence $X_1, \dots, X_n$, each cell $W_i(\widehat{h})$ has μ measure $1/n$ and $\bigcup_{i=1}^n W_i(\widehat{h}) \subset \mathcal{S}$. Figure 1 illustrates this with four points X_1, X_2, X_3 and X_4 and $\mu = \mathrm{Uniform}([0, 1]^2)$. Each point in the four cells (labeled 1, 2, 3 and 4) is mapped to the corresponding data point (X_1, X_2, X_3 and X_4) by the sample quantile function $\widehat{Q}_n \equiv \nabla\widehat{\psi}_n$. The convex function $\widehat{\psi}_n$ is not differentiable at the boundary of the 4 cells (marked by the black lines in the right panel of Figure 1). Remark A.3 in the Supplementary Material [33] illustrates the above ideas when $d = 1$ and $\mu = \mathrm{Uniform}([0, 1])$.

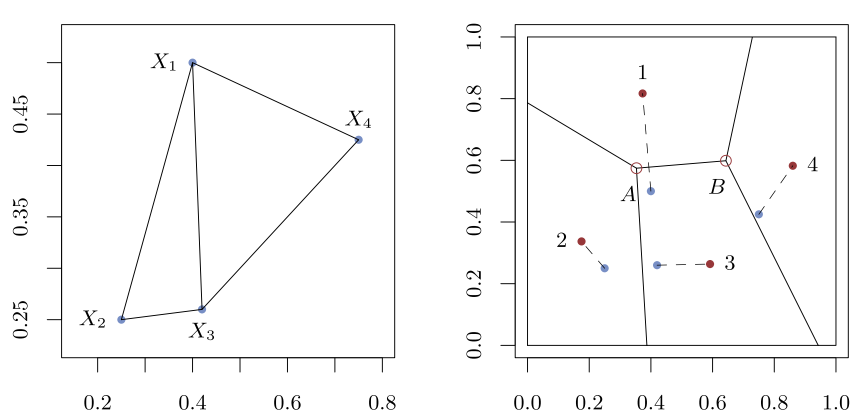

FIG. 1. *The left plot shows a data set with four 2-dimensional points X_1, X_2, X_3 and X_4. The right plot shows the four cells (each with area 1/4) marked 1, 2, 3, 4, and the four data points in blue (appropriately scaled to lie in $[0, 1]^2$) along with dashed lines connecting each data point to the centroid (in red) of the corresponding cell. The two points A and B in the right plot correspond to the intersection of three cells—1, 2, 3 and 1, 3, 4.*

3.1.2. *Empirical rank map.* Let us define $\widehat{\psi}_n^* : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ to be the Legendre–Fenchel dual of $\widehat{\psi}_n$, that is,

$$\widehat{\psi}_n^*(x) := \sup_{y \in \mathbb{R}^d} \{\langle x, y\rangle - \widehat{\psi}_n(y)\} = \sup_{u \in \mathcal{S}} \{\langle x, u\rangle - \widehat{\psi}_n(u)\}, \quad \text{for } x \in \mathbb{R}^d. \tag{14}$$

We define the *multivariate sample rank function* $\widehat{R}_n : \mathbb{R}^d \to \mathcal{S}$ as

$$\widehat{R}_n := \nabla \widehat{\psi}_n^*. \tag{15}$$

Lemma A.6 in the Supplementary Material [33] gives an alternate expression for $\widehat{R}_n$, which was used in [18], Definition 3.1. Further, Remark A.4 in [33] shows that when $d = 1$, $\widehat{R}_n$ is not defined uniquely at the data points. Note that the nonuniqueness of the rank function when $d = 1$ was finessed by enforcing right continuity, which is hard to do as we go beyond $d = 1$. Indeed, for any $d \geq 1$, $\widehat{R}_n(X_i)$ could be defined as any element in the (closure of the) cell $W_i(\widehat{h})$; this follows from Lemma 2.1. Figure 1 illustrates this when $\mu = \text{Uniform}([0, 1]^2)$. We see in Figure 1 that any point in the interior of the triangle formed by X_1, X_2 and X_3 (or X_1, X_3 and X_4) is mapped to the point A (or B) by the sample rank map $\widehat{R}_n$. The following result (Lemma 3.3; proved in Section C.3 of the Supplementary Material [33]) formalizes this observation when $d \geq 2$ and provides a way of finding the empirical rank map at any given point.

LEMMA 3.3. *Fix $x \in \mathbb{R}^d$. Suppose that for $i_1, \ldots, i_{d+1} \subset \{1, \ldots, n\}$:* (i) $x \in \text{Int}(\text{Conv}(X_{i_1}, \ldots, X_{i_{d+1}}))$ *and* (ii) *there exists a unique $u \in \mathcal{S}$ such that $u = \text{Cl}(W_{i_1}(\widehat{h})) \cap \cdots \cap \text{Cl}(W_{i_{d+1}}(\widehat{h}))$ (see (13)). Then u is the unique point in $\mathcal{S}$ such that $x \in \partial\widehat{\psi}_n(u)$. Furthermore, $\partial\widehat{\psi}_n^*(x) = u = \widehat{R}_n(x)$.*

3.1.3. *Empirical ranks.* By the "ranks" of the data points, we mean the rank function evaluated at the data points. When $d = 1$ and the underlying distribution is continuous, the usual ranks, that is, $\{\mathbb{F}_n(X_i)\}_{i=1}^n$ (here, $\mathbb{F}_n$ is the empirical distribution function), are identically distributed on the discrete set $\{1/n, 2/n, \ldots, n/n\}$ with probability $1/n$ each. As a consequence, the usual ranks are *distribution-free* (in $d = 1$), that is, the distribution of $\mathbb{F}_n(X_i)$ does not depend on the distribution of X_i. We may ask: "Does a similar property hold for the multivariate ranks $\widehat{R}_n(X_i)$?"

From the discussion in Section 3.1.2, it is clear that the multivariate ranks $\widehat{R}_n(X_i)$ are nonunique. In fact, we can choose $\widehat{R}_n(X_i)$ to be any point in the set $W_i(\widehat{h})$ (see (13)). In the sequel, we will use a special choice of $\widehat{R}_n(X_i)$, which will lead to a distribution-free notion. We define $\widehat{R}_n(X_i)$ as a random point drawn from the distribution $\widehat{\mu}_i$, that is, for $i \in \{1, \ldots, n\}$,

$$\widehat{R}_n(X_i) | X_1, \ldots, X_n \sim \widehat{\mu}_i \tag{16}$$

where

$$\widehat{\mu}_i : B \mapsto n\mu(W_i(\widehat{h}) \cap B) \quad \text{for any Borel } B \subset \mathbb{R}^d.$$

Note that $\widehat{\mu}_i$ is a Borel probability measure supported on the cell $W_i(\widehat{h})$ as $\mu(W_i(\widehat{h})) = n^{-1}$. When μ is the uniform distribution on $\mathcal{S}$, an equivalent representation of (16) is $\widehat{R}_n(X_i)|X_1, \ldots, X_n \sim \text{Uniform}(W_i(\widehat{h}))$. Thus, our choice of the empirical ranks $\{\widehat{R}_n(X_i)\}_{i=1}^n$ is random. However, this external randomization leads to the following interesting consequence—the multivariate ranks are marginally distribution-free. This is formalized in the following lemma (proved in Section C.4 of the Supplementary Material [33]).

LEMMA 3.4. *Suppose that $X_1, \ldots, X_n$ are i.i.d. ν, an absolutely continuous distribution on $\mathbb{R}^d$. Then, for any $i = 1, \ldots, n$, $\widehat{R}_n(X_i) \sim \mu$.*

Compare Lemma 3.4 with the result $R(X) \sim \mu$ where R is the (population) rank map of $X \sim \nu$ (as R pushes forward ν to μ). If we do not want a randomized choice of ranks, then we can define $\widehat{R}_n(X_i) := \max_{u \in \mathrm{Cl}(W_i(\widehat{h}))} \|u\|$; the above choice is convenient for computational purposes.

REMARK 3.11 (Choice of the reference distribution μ). As may have been clear from the above discussion, the concept of multivariate (empirical) ranks and quantiles, based on OT, depends on the choice of the reference distribution μ. For example, the choice of a spherically symmetric μ (e.g., Uniform($B_1(0)$)) leads to quantile maps that are equivariant under orthogonal transformations (which can be a useful property when studying multivariate depth, outlyingness, etc.), whereas the choice of $\mu = \text{Uniform}([0, 1]^d)$ guarantees factorization into lower dimensional marginals under independence (that may be more appropriate for measuring association/independence between the marginals of ν); see Section A.3 in the Supplementary Material [33] for formal results in this regard. We would like to point out here that $\mathcal{S}$, the support of μ, can play an important role in determining the behavior of the rank map $\widehat{R}_n(\cdot)$; we will see in Theorem 4.1 that the choice $\mathcal{S} = [0, 1]^d$ could lead to inconsistent estimation of $\widehat{R}_n(x)$ for x near the boundary of $\mathcal{Y}$ (see Remark 4.1 for further discussion).

The two plots in Figure 2 show the cell decompositions corresponding to the uniform measures on $[0, 1]^2$ and $B_1(0) \subset \mathbb{R}^2$, respectively. Note that when μ is the uniform measure on any set $\mathcal{S} \subset \mathbb{R}^d$, the volume of each cell is $1/n$; moreover, when $\mu = \text{Uniform}([0, 1]^d)$ the cells are convex polyhedrons, which can be especially easy to visualize and implement in a computer. As in $d = 1$, we believe that the use of appropriate score functions can mitigate the dependence of multivariate rank-based procedures on the reference distribution μ; see, for example, [78], Chapters 13 and 15, for the usefulness of a score-based approach when $d = 1$. Indeed, the recent papers [23, 74] illustrate the flexibility and power of incorporating score functions in the definition of multivariate rank-based tests. In particular, in the recent paper [23] the authors show that different reference distributions lead to different asymptotic efficiencies for certain tests in the two-sample problem. Thus, when $d > 1$, the choice of μ should be guided by various considerations, as alluded to above. We expect that future research will shed more light on this important question of the choice of the reference distribution.

3.2. *Computation of the sample quantile and rank functions*. The computation of the empirical quantile function $\widehat{Q}_n$ (in (11)) reduces to a semidiscrete OT problem. There are several methods proposed in the literature to solve the semidiscrete OT problem. Oliker and Prüssner [61] proposed one of the earliest algorithms in this regard relying on coordinatewise increments; also see [16, 58]. Although this algorithm has convergence guarantees (see [45]), it is quite slow in practice. Recently, fast algorithms for solving (11) have been proposed that typically rely on the formulation of the semidiscrete OT problem as an unconstrained convex optimization problem which is then solved using a (damped) Newton or quasi-Newton method; see, for example, [2, 47, 52, 57]. See [58] for a detailed account of many of the algorithms cited above.

In the following, we outline our approach to computing $\widehat{Q}_n$ (see the R package https://github.com/Francis-Hsu/testOTM [85]). We use Newton-type algorithms proposed in the papers [47, 52, 57] and implemented in the Geogram[2] package. These algorithms are experimentally efficient and converge globally with linear rate; see [47]. Our approach is different from the "gradient algorithm" used in [18], Section 4, to solve the semidiscrete problem.

[2]http://alice.loria.fr/software/geogram/doc/html/

The computation of $\widehat{Q}_n$ leads to a "partition" of $\mathcal{S}$ into n convex sets, each with volume $1/n$ (i.e., the $W_i(\widehat{h})$'s in (13)), and involves what is usually called the power diagram [1]—a type of weighted Voronoi diagram. Recall that $\mathcal{X} := \{X_1, \dots, X_n\}$. Let $w = (w_1, \dots, w_n) \in \mathbb{R}^n$ be a given (weight) vector. The *power diagram* of $(\mathcal{X}, w)$ is the decomposition of the set $\mathcal{S}$ into a finite number of cells, one for each element in $\mathcal{X}$, defined by (for $i = 1, \dots, n$)

$$\mathrm{Vor}^w_{\mathcal{X}}(i) := \{u \in \mathcal{S} : \|u - X_i\|^2 - w_i \le \|u - X_j\|^2 - w_j, \text{ for all } j \ne i\}.$$

Note that if the weights are all zero, this coincides with the usual Voronoi diagram. The computation of the power diagram is a classical problem in computational geometry, for which there exists very efficient software, such as CGAL[3] or Geogram. Two-dimensional power diagrams can be constructed by an algorithm that runs in time $O(n \log n)$. More generally, d-dimensional power diagrams (for $d > 2$) may be constructed by an algorithm with worst case complexity $O(n^{\lceil d/2 \rceil})$; see, for example, [1], [3], Chapter 6.

Given the power diagram of $(\mathcal{X}, w)$ we can define the *power map* $T^w_{\mathcal{X}} : \mathcal{S} \to \mathcal{X}$ such that $T^w_{\mathcal{X}}(u) := X_i$ if $u \in \mathrm{Vor}^w_{\mathcal{X}}(i)$. This map is well defined μ-a.e. (except on the boundary of the power cells). A weight vector $w \in \mathbb{R}^n$ is called *adapted* to $(\mu, \widehat{\nu}_n)$ if for every $i = 1, \dots, n$, one has $\mu(\mathrm{Vor}^w_{\mathcal{X}}(i)) = \int_{\mathrm{Vor}^w_{\mathcal{X}}(i)} d\mu(u) = n^{-1}$. [57], Theorem 2, shows that finding a weight vector adapted to $(\mu, \widehat{\nu}_n)$ amounts to finding the global minimum of the convex function

$$L(w) := -\sum_{i=1}^{n} \left[\frac{w_i}{n} + \int_{\mathrm{Vor}^w_{\mathcal{X}}(i)} (\|u - X_i\|^2 - w_i)\, d\mu(u) \right], \quad \text{for } w \in \mathbb{R}^n; \tag{17}$$

also see [2]. Moreover, [57], Theorem 2, shows that the power map $T^{\widehat{w}}_{\mathcal{X}}$, where $\widehat{w}$ is the global minimizer of $L(\cdot)$, is the OT map between μ and $\widehat{\nu}_n$, that is,

$$\widehat{Q}_n = T^{\widehat{w}}_{\mathcal{X}}.$$

Thus, we have to minimize $L(\cdot)$ in (17) to obtain $\widehat{w}$, which will yield $\widehat{Q}_n$. Note that the gradient and Hessian of $L(\cdot)$ can be easily computed; see, for example, [52]. This makes Newton-type algorithms especially attractive in computing $\widehat{w}$.

The potential function $\widehat{\psi}_n$, as defined in (12), can also be recovered from the above optimization problem. Let $\widehat{h}_i := \frac{1}{2}(\widehat{w}_i - \|X_i\|^2)$, for $i = 1, \dots, n$. Now, we can easily see that the convex function thus defined by (12) has gradients that coincides with the quantile map $\widehat{Q}_n$; see, for example, [57], Section 3.4. The computation of the dual potential $\widehat{\psi}^*_n$ (as defined in (14)) and the empirical rank map $\widehat{R}_n$ now follows easily; see Remark A.2 (in Section A.2 in the Supplementary Material [33]) for the details.

REMARK 3.12 (Computation of the sample ranks). For computing the sample rank $\widehat{R}_n(X_i)$, for $i = 1, \dots, n$, we advocate the use of a randomized choice where we define $\widehat{R}_n(X_i)$ as any point in the set $W_i(\widehat{h})$, chosen according to the probability measure in (16). When μ is the uniform distribution on a convex polytope $\mathcal{S}$ (e.g., $[0, 1]^d$), this computation is especially simple as then $W_i(\widehat{h})$ is also a convex polytope whose vertices are already provided by our algorithm, and thus, uniform sampling can be carried out easily (e.g., via rejection sampling on the smallest hyperrectangle containing $W_i(\widehat{h})$). In fact, the above approach is much more broadly applicable, as in practice the computer always approximates $\mathcal{S}$ by a convex polytope (see, e.g., Figure 2).

The two plots in Figure 2 show the cell decompositions of $[0, 1]^2$ and $B_1(0) \subset \mathbb{R}^2$, obtained from the semidiscrete OT problem; see Section A.5 in the Supplementary Material [33] for

[3] http://www.cgal.org

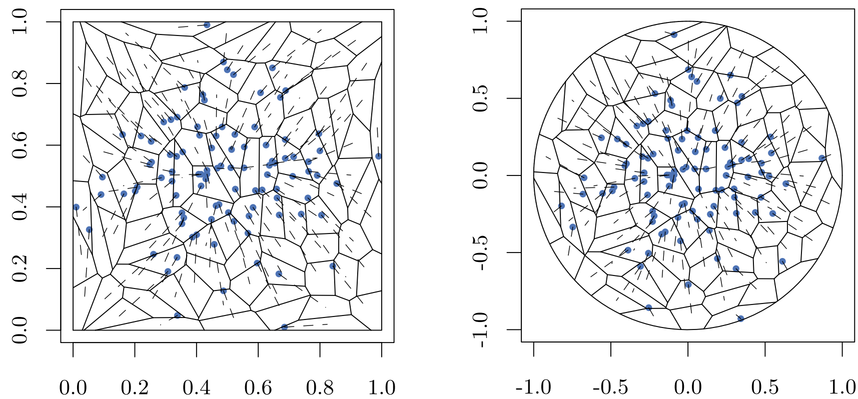

FIG. 2. *Left plot*: *Shows the cell decomposition of* $\mathcal{S} = [0, 1]^2$ (*each with area* $1/n$ *where* $n = 100$) *induced by the estimated quantile function* $\widehat{Q}_n$ (*w.r.t.* $\mu = \text{Uniform}([0, 1]^2)$) *where the data points are drawn i.i.d. from* $N_2((0, 0), I_2)$ (*and appropriately scaled to lie in* $[0, 1]^2$) *along with dashed lines indicating which cell corresponds to which data point*. *Right plot*: *Shows the corresponding cell decomposition of* $\mathcal{S} = B_1(0)$—*the ball of radius* 1 *around* $(0, 0) \in \mathbb{R}^2$—*induced by* $\widehat{Q}_n$ (*w.r.t.* $\mu = \text{Uniform}(B_1(0))$).

more plots of this kind. In particular, we can directly visualize the empirical quantile maps for the two settings. As the (empirical) rank function (taking values in $\mathbb{R}^d$) is a bit difficult to visualize, in Figure 3 we plot the estimated depth functions for the banana-shaped distribution when $n = 1000$; cf. [18], Figure 2, where the authors motivate the use of multivariate ranks/quantiles based on OT using this data. The banana-like geometry of the data cloud is correctly picked up by the nonconvex contours in Figure 3. We also provide depth function plots for other distributions in Section A.5 of the Supplementary Material [33].

3.3. *Comparison with Chernozhukov et al*. [18], *Hallin et al*. [37] *and Boeckel et al*. [11]. In the papers [18, 37] and [11], the authors use ideas from the theory of OT to define multivariate quantiles and ranks. Although our approach is similar in spirit to that of [18] there are subtle and important differences. As opposed to [18] and [11], the quantile map here is defined based on McCann's result (see Theorem 2.2), which extended Brenier's theorem to general probabilities, without the need for a second moment. Whereas [18] studied multivariate quantiles and ranks to obtain notions of statistical depth, we study quantiles and ranks to aid us to construct nonparametric goodness-of-fit and mutual independence tests.

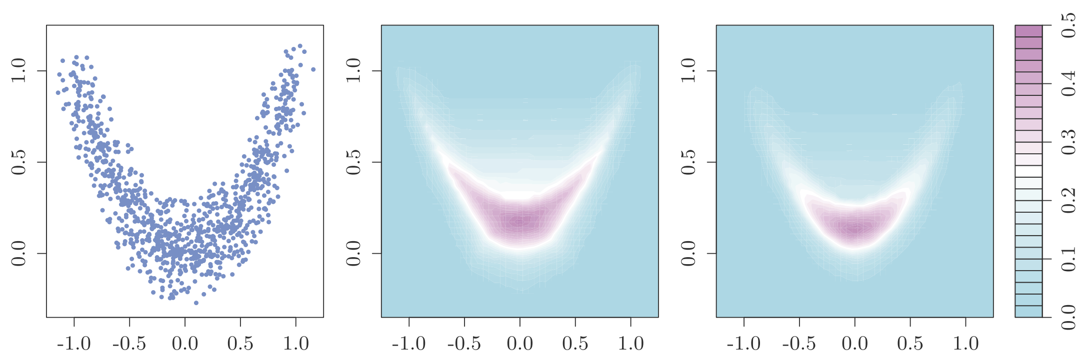

FIG. 3. *Left panel*: *A random sample of size* $n = 1000$ *drawn from the banana-shaped distribution*. *Middle panel*: *The estimated depth function—defined as* $\widehat{D}_n(x) := 1/2 - \|\widehat{R}_n(x) - (1/2)\mathbf{1}\|_\infty$ *for* $x \in \mathbb{R}^d$ (*see* [18]), *where* $\mathbf{1} = (1, 1, \dots, 1) \in \mathbb{R}^d$—*using* $\mu = \text{Uniform}([0, 1]^2)$. *Right panel*: *The estimated depth function—defined as* $\widehat{D}_n(x) := \pi^{-1}(\theta - \cos\theta \sin\theta)$ *where* $\theta = \arccos(\|\widehat{R}_n(x)\|)$—*w.r.t.* $\mu = \text{Uniform}(B_1(0))$; *see* [69], *Section* 5.6.

The approach to defining multivariate ranks and quantiles proposed and studied in [37] and [11] are quite different from ours. In the papers [37] and [11], the authors solve a discrete-discrete OT problem, compared to our semidiscrete approach (further, in [11] the authors only consider target distributions supported on a compact subset of $\mathbb{R}^d$). Thus, to define the empirical rank map this approach involves the choice of n representative points inside the set $\mathcal{S}$ (that approximates the measure μ) to solve the discrete-discrete OT problem (between the sample data points and the n chosen points). Thus, the "ranks" of the data points are forced to be the points in the chosen grid. This approach immediately leads to many attractive features for the empirical ranks, for example, the distribution-freeness of the ranks. However, this approach does not automatically give rise to a quantile function (or quantile contours) and special smoothing interpolation is required. In comparison, our approach has the drawback of leading to nonunique ranks at the data points. In a sense, our approach yields an elegant and useful notion of quantiles while the approach of [37] (and [11]) yields a notion of ranks with attractive properties.

4. Uniform convergence of empirical quantile and rank maps. The rank and quantile functions in one dimension enjoy many interesting asymptotic properties. For example, if $X_1, \dots, X_n \sim \nu$, where ν is a distribution on $\mathbb{R}$, then by the Glivenko–Cantelli theorem, the empirical rank function (which is the empirical distribution function when $d = 1$) converges uniformly to the population rank function a.s. Similarly, for $d = 1$, the empirical quantile function converges uniformly (on compacts $[a, b] \subset (0, 1)$) to the population quantile function, when the underlying distribution function is continuous and strictly increasing. One may wonder if such results also hold for the multivariate empirical quantile/rank maps studied in this paper. In Theorem 4.1 below, we show that this is indeed the case.

Suppose that ν is absolutely continuous with support $\mathcal{Y} \subset \mathbb{R}^d$; here, ν is the target distribution. Let μ be an absolutely continuous distribution supported on a compact convex set $\mathcal{S} \subset \mathbb{R}^d$. Let Q and R be the quantile and rank maps of ν (w.r.t. μ); as in (7) and (9), respectively. Let $X_1, X_2, \dots, X_n \overset{i.i.d.}{\sim} \nu$. Let $\{\widehat{\nu}_n\}_{n \geq 1}$ be a sequence of random probability distributions (computed from $X_1, \dots, X_n$) such that $\widehat{\nu}_n$ converges weakly to ν a.s., that is,

$$\widehat{\nu}_n \overset{d}{\to} \nu \quad \text{a.s.} \tag{18}$$

We can take $\widehat{\nu}_n$ to be the empirical distribution obtained from the first n data points, that is, $\widehat{\nu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$; in this case, we know that (18) holds (see, e.g., [29], Theorem 11.4.1). Denote the multivariate quantile/rank functions for $\widehat{\nu}_n$ by $\widehat{Q}_n$ and $\widehat{R}_n$. In particular, when the underlying potential functions (see Definition 3.1) are not differentiable, we define $\widehat{Q}_n$ and $\widehat{R}_n$ to be any point in the corresponding subdifferential set. The following is a main result of this paper (see Section D.2 of the Supplementary Material [33] for its proof).

THEOREM 4.1. *Consider the notation introduced above and suppose that* (18) *holds. Suppose that* $Q : \mathrm{Int}(\mathcal{S}) \to \mathrm{Int}(\mathcal{Y})$ *is a homeomorphism.*[4] *Let* $K_1 \subset \mathrm{Int}(\mathcal{S})$ *and* $K_2 \subset \mathrm{Int}(\mathcal{Y})$ *be any two compact sets.*

(a) *Then*

$$\sup_{u \in K_1} \|\widehat{Q}_n(u) - Q(u)\| \overset{a.s.}{\to} 0. \tag{19}$$

(b) *Further,*

$$\sup_{x \in K_2} \|\widehat{R}_n(x) - R(x)\| \overset{a.s.}{\to} 0. \tag{20}$$

[4] See Proposition 3.1 for sufficient conditions.

(c) *Suppose that* $\mathcal{S}$ *is a strictly convex compact set* (*as in Definition* 2.3). *Let* $\{\lambda_n\}_{n\geq 1} \subset \mathbb{R}$ *be a sequence such that* $\lambda_n \to \infty$ *as* $n \to \infty$. *Then*

$$\sup_{x\in\mathbb{R}^d} \|\widehat{R}_n(x) - R(x)\| \stackrel{a.s.}{\to} 0, \quad \textit{and} \tag{21}$$

$$\lim_{\lambda_n\to\infty} \widehat{R}_n(\lambda_n x) \stackrel{a.s.}{=} \arg\max_{v\in\mathcal{S}} \langle x, v\rangle, \quad \textit{for all } x \in \mathbb{R}^d. \tag{22}$$

Theorem 4.1(a) (i.e., (19)) extends the uniform convergence of the empirical quantile function (on compacts in the interior of $[0, 1]$) beyond $d = 1$. Theorem 4.1(b) (i.e., (20)) shows the uniform convergence of the estimated rank map on any compact set inside Int($\mathcal{Y}$). One may notice that Theorem 4.1(a) and (b) improve upon the result of [18], Theorem 3.1, where the authors prove a similar convergence result for the estimated quantile/rank maps under the assumption of compactness of $\mathcal{Y}$. In [37], Theorem 2.2.1, a result similar to (21) is given for the empirical rank map arising from a discrete-discrete OT problem, when the reference measure is the spherical uniform distribution; also see [11], Theorem 2.3, for a similar result where the authors only consider a compactly supported ν. In [86], Proposition 6, the authors prove a local uniform convergence result for the empirical quantile map, under additional finite second moment assumptions on ν. Theorem 4.1(c) (see (21)) can be thought of as the proper generalization of the Glivenko–Cantelli theorem beyond $d = 1$ where we show the a.s. convergence of the estimated rank map uniformly over the whole of $\mathbb{R}^d$.

To prove Theorem 4.1, one needs to develop tools that deal with convergence of (sub)gradients of a sequence of convex functions and their Legendre–Fenchel duals. These tools are summarized in three deterministic lemmas in Section D of the Supplementary Material [33]—Lemmas D.1, D.2 and D.3—and could be of independent interest.

REMARK 4.1 (On the sufficient condition for (21)). In (21), we show that the empirical rank map converges to the population rank function uniformly on $\mathbb{R}^d$, under the strict convexity assumption on $\mathcal{S}$. This sufficient condition is certainly satisfied, for example, when $\mathcal{S}$ is the unit ball in $\mathbb{R}^d$, that is, $\mathcal{S} = B_1(0)$. Unfortunately, when $\mathcal{S} = [0, 1]^d$, this condition is not satisfied.

REMARK 4.2 (Necessity of Q to be a homeomorphism). One of the main assumptions in Theorem 4.1 is that the population quantile Q is a homeomorphism; for $d = 1$, this corresponds to assuming that the distribution function is continuous and strictly increasing. It is actually a necessary condition for showing the uniform convergence of $\widehat{Q}_n$ (the sample quantile function) to Q; in fact, more generally, for a sequence of (sub)gradients of convex functions. To see this, consider the example of a sequence of convex functions $\phi_n : \mathbb{R} \to \mathbb{R}$ defined as $\phi_n(x) := (x^2 + n^{-1})^{1/2}$. As $n \to \infty$, $\phi_n(x)$ converges pointwise to $\phi(x) := |x|$. However, the subdifferential set of the function $\phi(x)$ at $x = 0$ is equal to $[-1, 1]$ whereas $\phi_n'(0) = 0$ for all $n \geq 1$. Hence, $\phi_n'(\cdot)$ does not converge uniformly on any compact set containing 0.

REMARK 4.3 (When is Q a homeomorphism?). In Proposition 3.1, we provide a sufficient condition on the density of ν, supported on a convex set, which ensures that the quantile map Q will be a homeomorphism; also see Remarks 3.8 and 3.9. Recently, in [46], Proposition 4.5 and Corollary 4.6, some results are provided that show that Q can be a homeomorphism even when the support of ν is a union of convex domains.

REMARK 4.4 (Connection to [22]). The recent paper [22] implies "graphical convergence" of the estimated quantile maps (see [22], Theorem 4.2 and Corollary 4.4). Their result does not need absolute continuity of ν and no restrictions are placed on the supports of the measures μ and ν. However, graphical convergence, which implies a form of local uniform convergence, is weaker than uniform convergence on compacta stated in Theorem 4.1. Moreover, since the sample rank map $\widehat{R}_n$ is not strictly a transport map, it is not clear if [22] implies any notion of convergence for $\widehat{R}_n$.

5. Rate of convergence of empirical quantile/rank maps. In this section, we study the global and local rates of convergence of the empirical quantile/rank maps. Section 5.1 provides upper bounds on the global L_2-risk of the empirical quantile map whereas Section 5.2 provides analogous results for the empirical rank map. In Section 5.3, we provide a result that gives a local uniform rate of convergence for the empirical quantile/rank maps.

5.1. *Global rate of convergence for the empirical quantile map $\widehat{Q}_n$.* We first state a lemma (see Lemma 5.1 below; proved in Section E.1 of [33]) that upper bounds the L_2-distance between two OT maps using the difference of the corresponding 2-Wasserstein distances (and a remainder term). Note that the *2-Wasserstein distance* between μ and ν is defined as

$$W_2(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)}\left(\int \|u-x\|^2\,d\pi(u,x)\right)^{1/2},$$

where $\Pi(\mu,\nu)$ denotes the collection of all joint distributions (couplings) π with marginal distributions μ and ν; see Section A.1 of [33] for more details.

LEMMA 5.1. *Let μ, ν and $\widetilde{\nu}$ be three probability measures on $\mathbb{R}^d$ such that $\int\|x\|^2\,d\mu(x) < +\infty$, $\int\|x\|^2\,d\nu(x) < +\infty$ and $\int\|x\|^2\,d\widetilde{\nu}(x) < +\infty$. Also, let ψ and $\widetilde{\psi}$ be two convex functions such that $\nabla\psi\#\mu = \nu$ and $\nabla\widetilde{\psi}\#\mu = \widetilde{\nu}$, respectively. Suppose $\psi^* : \mathbb{R}^d \to \mathbb{R}\cup\{+\infty\}$, the Legendre–Fenchel dual of ψ, is strongly convex with parameter $\lambda > 0$. Then, letting $g(x) := \frac{\|x\|^2}{2} - \psi^*(x)$,*

$$\int \|\nabla\widetilde{\psi} - \nabla\psi\|^2\,d\mu \le \frac{1}{\lambda}\left[\{W_2^2(\mu,\widetilde{\nu}) - W_2^2(\mu,\nu)\} + 2\int g\,d(\nu-\widetilde{\nu})\right].$$

The above lemma, which is of independent interest, gives a quantitative stability estimate for OT maps in the semidiscrete setting. Although the stability of OT maps has recently been studied by many authors (see, e.g., [34, 43, 53, 81]), we could not find such an explicit upper bound, under such minimal assumptions on the underlying distributions. Moreover, as we illustrate in Theorem 5.2 below (proved in Section E.2 of the Supplementary Material [33]), Lemma 5.1 can be used to obtain rates for OT maps that are strictly better than those obtained in [81], Theorem 1.1, and [53], Section 4. It is worth pointing out that the starting point of the proof of Lemma 5.1 is based on an observation in [34], Proposition 3.3.

THEOREM 5.2. *Let μ be an absolutely continuous probability measure supported on a compact convex set $\mathcal{S}\subset\mathbb{R}^d$. Let $X_1,\ldots,X_n \overset{i.i.d.}{\sim} \nu$, where ν is an absolutely continuous distribution on $\mathbb{R}^d$ with population quantile map $Q \equiv \nabla\psi$ (see (7)); here, ψ is a convex function. Suppose that $\psi^* : \mathbb{R}^d \to \mathbb{R}\cup\{+\infty\}$ (the Legendre–Fenchel dual of ψ) is strongly convex. Then, for all $n\ge 1$, with $\widehat{Q}_n$ being the empirical quantile map (see (10)),*

$$\mathbb{E}\left[\int \|\widehat{Q}_n - Q\|^2\,d\mu\right] \le C r_{d,n}, \tag{23}$$

where $C \equiv C(\mu,\nu) > 0$ is a constant that depends on μ and ν, and $r_{d,n}$ is defined in (3). Furthermore, there exists $c > 0$ such that, for all $s \ge 0$,

$$\mathbb{P}\left(\int \|\widehat{Q}_n - Q\|^2\,d\mu \ge C r_{d,n} + n^{-1/2}s\right) \le \exp(-cs^2).$$

We believe that the above result gives the exact rate of convergence for the $\widehat{Q}_n$ when $d > 4$; see [43] where the authors mention "... In this case, one formally recovers the rate $n^{-2/d}$ and we conjecture that this is the minimax rate of estimation in the context where the transport map T_0 is only assumed to be the gradient of a strongly convex function with Lipschitz gradient..." Note that in Theorem 5.2 we just assume strong convexity of the dual potential associated with the quantile map Q. We would also like to point out here that, even when $d = 1$, without some assumptions on ν it is impossible to derive rates of convergence for $\widehat{Q}_n$ as in (23); see, for example, [10, 25].

The left-hand side in (23) is obviously an upper bound for $\mathbb{E}[W_2^2(\widehat{\nu}_n, \nu)]$ and, as a consequence,

$$\mathbb{E}[W_2^2(\widehat{\nu}_n, \nu)] \le \mathbb{E}\left[\int \|\widehat{Q}_n - Q\|^2 \, d\mu\right] \le C r_{d,n}.$$

Compare this with [31], Theorem 1, which yields $\mathbb{E}[W_2^2(\widehat{\nu}_n, \nu)] \le C r_{d,n}$, when ν has a finite moment of order $q > 4$. Thus, Theorem 5.2 is an improvement of the result in [31], under the strong convexity assumption on the potential function ψ^*.

The proof of Theorem 5.2 utilizes the stability result of the empirical quantile map $\widehat{Q}_n$ obtained in Lemma 5.1. As one may note, Lemma 5.1 bounds the L_2-loss of $\widehat{Q}_n$ by the difference (up to a smaller order term) between two 2-Wasserstein distances, under minimal structural assumptions on the dual potential of Q. The rate of convergence in Theorem 5.2 is then obtained by analyzing the expected value of the difference of the Wasserstein distances using empirical process theory; see, for example, the proof of [19], Theorem 2.

5.2. *Global rate of convergence for the empirical rank map* $\widehat{R}_n$. Deriving a rate of convergence for the multivariate sample rank $\widehat{R}_n$ map is a bit more tricky. Note that $\widehat{R}_n$ is not an OT map per se, but is defined via the Legendre–Fenchel dual of the potential function $\widehat{\psi}_n$ (see (15)). Also, the sample ranks $\widehat{R}_n(X_i)$'s are not uniquely defined (see Section 3.1.3). In this subsection, we consider the randomized choice of the empirical ranks (as in (16)). In the following result, we give an upper bound on the risk of the sample rank map (see Section E.3 of the Supplementary Material [33] for its proof).

THEOREM 5.3. *Let μ be an absolutely continuous probability measure supported on a compact convex set $\mathcal{S} \subset \mathbb{R}^d$. Let $X_1, \dots, X_n$ be i.i.d. from an absolutely continuous distribution ν on $\mathbb{R}^d$ with compact support and rank map $R \equiv \nabla\psi^*$, where $\psi : \mathcal{S} \to \mathbb{R}$ is assumed to be strongly convex. For $i = 1, \dots, n$, let $\widehat{R}_n(X_i)$ be defined as in* (16). *Then, for all $n \ge 1$,*

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \|\widehat{R}_n(X_i) - R(X_i)\|^2\right] \le K r_{d,n}, \tag{24}$$

where $K \equiv K(\mu, \nu) > 0$ depends on μ and ν, and $r_{d,n}$ is defined in (3).

The expectation on the left-hand side of (24) averages over the external randomization in the definition of the empirical ranks. Note that for Theorem 5.3 to hold we need to assume that ν has compact support, in addition to the strong convexity of ψ. Although a formal result on the optimality of the upper bound in Theorem 5.3 is beyond the scope of this paper, we believe that the obtained bounds are optimal when $d > 4$.

5.3. *Local uniform rate of convergence.* Theorem 5.4 below (proved in Section F.1 of the Supplementary Material [33]), provides a local uniform rate of convergence of the empirical quantile/rank maps. In the following result, when the underlying potential functions are not differentiable, we define $\widehat{Q}_n$ and $\widehat{R}_n$ to be any point in the corresponding subdifferential sets.

THEOREM 5.4. *Let μ be an absolutely continuous distribution with a bounded nonvanishing density supported on a compact convex set $\mathcal{S} \subset \mathbb{R}^d$. Let $X_1, \ldots, X_n$ be i.i.d. ν absolutely continuous and supported on a convex set $\mathcal{Y} \subset \mathbb{R}^d$ with population quantile map $Q \equiv \nabla\psi$ (see (7)), where ψ is a convex function. Suppose that Q is a homeomorphism from* $\mathrm{Int}(\mathcal{S})$ *to* $\mathrm{Int}(\mathcal{Y})$. *Assume that ψ^* and ψ are strongly convex inside $\mathcal{Y}$ and $\mathcal{S}$, respectively. Fix $u_0 \in \mathrm{Int}(\mathcal{S})$ and $\delta_0 \equiv \delta_0(u_0) > 0$ such that $B_{\delta_0}(u_0) \subset \mathcal{S}$ and $B_{\delta_0}(\nabla\psi(u_0)) \subset \mathcal{Y}$. Then there exists a constant $C \equiv C(\mu, \nu, u_0) > 0$, depending only on μ, ν and u_0, such that, for all $n \geq 1$,*

$$\mathbb{E}\Big[\sup_{u \in B_{\delta_0/3}(u_0)} \|\widehat{Q}_n(u) - Q(u)\|\Big] \leq C r_{d,n}^{\frac{1}{d+2}},$$

and

$$\mathbb{E}\Big[\sup_{x \in B_{\delta_0/6}(\nabla\psi(u_0))} \|\widehat{R}_n(x) - R(x)\|\Big] \leq C r_{d,n}^{\frac{1}{d+2}}.$$

The proof of Theorem 5.4 is built on Proposition F.1, stated in the Supplementary Material [33], which connects the local uniform rate of convergence of $\widehat{Q}_n$ and $\widehat{R}_n$ with the local L_2-rate of convergence of $\widehat{Q}_n$. Theorem 5.2 is then used to upper bound this local L_2-rate of convergence.

To the best of our knowledge, the above result is the first attempt to study the local uniform behavior of transport maps. However, it is not clear to us whether the above bounds are tight when $d \geq 2$. We believe that it may be possible to improve our rate of convergence result under further assumptions on ν. We hope to address this in future work.

6. Applications to nonparametric testing.

6.1. *Two-sample goodness-of-fit testing in* $\mathbb{R}^d$. Suppose that we observe $X_1, \ldots, X_m$ i.i.d. ν_X and $Y_1, \ldots, Y_n$ i.i.d. ν_Y, where $m, n \geq 1$ and ν_X and ν_Y are unknown absolutely continuous distributions on $\mathbb{R}^d$. We also assume that both the samples are drawn mutually independently. In this section, we consider the two-sample equality of distribution hypothesis testing problem:

$$H_0 : \nu_X = \nu_Y \quad \text{versus} \quad H_1 : \nu_X \neq \nu_Y. \tag{25}$$

The two-sample problem for multivariate data has been extensively studied, beginning with the works of [7, 84]. Several graph based methods have been proposed in the literature for this problem; see, for example, [6, 32, 68, 70] and the references therein. Also see [4, 71, 76] for distance and kernel based methods for the two-sample problem. Recently, the theory of OT and Wasserstein distances have been used to construct goodness-of-fit tests for (25); see, for example, [11, 24, 28, 38, 65]. In the following we propose a tuning-free method that uses the (estimated) multivariate quantile/rank maps defined in Section 3.

Let μ be an absolutely continuous distribution supported on a compact convex set $\mathcal{S} \subset \mathbb{R}^d$ having a density (w.r.t. Lebesgue measure), for example, $\mu = \mathrm{Uniform}([0,1]^d)$ or $\mu = \mathrm{Uniform}(B_1(0))$. Let $\widehat{Q}_X$ and $\widehat{Q}_Y$ be the sample quantile maps estimated from the X_i's and Y_j's, respectively (w.r.t. μ). Let $\widehat{R}_{X,Y}$ be the empirical rank map of the pooled sample $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ (w.r.t. μ). As in Section 3.1.3, we define the rank at any data point as a randomized value (as in (16)). We use the following test statistic for testing (25):

$$T_{X,Y} := \int_{\mathcal{S}} \|\widehat{R}_{X,Y}(\widehat{Q}_X(u)) - \widehat{R}_{X,Y}(\widehat{Q}_Y(u))\|^2 \, d\mu(u). \tag{26}$$

Exactly computing $T_{X,Y}$ is possible as the above integral reduces to a finite sum; see Section H.1 of the Supplementary Material [33] for the details. One can also easily approximate $T_{X,Y}$ using Monte Carlo.

We reject H_0 when $T_{X,Y}$ is large. To motivate the form of the above test statistic, consider the one-sample Cramér–von Mises statistic when $d = 1$. Let $\mathbb{F}_n$ be the empirical distribution of the data (when $d = 1$) and F be the true distribution function (assumed to be absolutely continuous). Then the Cramér–von Mises statistic can be written as

$$\int_{\mathbb{R}} \{\mathbb{F}_n(x) - F(x)\}^2 \, dF(x) = \int_0^1 \{\mathbb{F}_n(F^{-1}(u)) - u\}^2 \, du.$$

Indeed, (26) is similar to the right-hand side of the above display; however, as we are now in the two-sample case, F^{-1} is unknown and is replaced by the sample quantile function.

The connection to the Cramér–von Mises statistic immediately raises the following question: Is $T_{X,Y}$ distribution-free under H_0 (as the Cramér–von Mises statistic when $d = 1$)? Unfortunately, we do not know the exact answer to this question. In the following lemma (proved in Section G.1 of the Supplementary Material [33]), we show that $\widehat{R}_{X,Y}(\widehat{Q}_X(U))$ and $\widehat{R}_{X,Y}(\widehat{Q}_Y(U))$ (as in (26)) are both marginally distribution-free and distributed as μ under H_0.

LEMMA 6.1. *Suppose that $\nu_X = \nu_Y$. Then $\widehat{R}_{X,Y}(\widehat{Q}_X(U)) \sim \mu$ and $\widehat{R}_{X,Y}(\widehat{Q}_Y(U)) \sim \mu$, and hence their distributions do not depend on $\nu_X \equiv \nu_Y$.*

REMARK 6.1 (Finding the critical value of $T_{X,Y}$). Although we have shown (in Lemma 6.1) that $\widehat{R}_{X,Y}(\widehat{Q}_X(U)) \sim \mu$ and $R_{X,Y}(\widehat{Q}_Y(U)) \sim \mu$ (and thus both quantities are distribution-free), it is not immediately clear if the test statistic $T_{X,Y}$ in (26) is distribution-free, under H_0. In Section H.1 of [33], we provide simulation evidence that suggests that a properly normalized version of $T_{X,Y}$ may be asymptotically distribution-free, at least when $d = 2$. In any case, the critical value of the test can always be computed by conditioning on the observed samples and using the following permutation principle: Under H_0, $X_1, \dots, X_m, Y_1, \dots, Y_n$ are i.i.d., and thus we can consider any partition of the $m + n$ data points into two sets of sizes m and n and recompute our test statistic to simulate its null distribution. This is indeed the most common approach in these nonparametric testing problems as it avoids the need to use asymptotic distributions and leads to exact tests; see, for example, [41, 44, 51].

The following result (proved in Section G.2 of the Supplementary Material [33]) shows that our proposed test has asymptotic power 1 when $\nu_X \neq \nu_Y$.

PROPOSITION 6.2 (Consistency). *Suppose that $H_0 : \nu_X = \nu_Y \equiv \nu$ holds. Assume that ν is supported on a domain $\mathcal{Y} \subset \mathbb{R}^d$ such that the quantile map $Q : \mathrm{Int}(\mathcal{S}) \to \mathrm{Int}(\mathcal{Y})$ is a homeomorphism. Also, assume that $m, n \to \infty$ such that $\frac{m}{m+n} \to \theta \in (0, 1)$. Then, under H_0, as $m, n \to \infty$,*

$$T_{X,Y} \xrightarrow{a.s.} 0.$$

Now, suppose that $X_1, \dots, X_m \overset{\text{i.i.d.}}{\sim} \nu_X$ and $Y_1, \dots, Y_n \overset{\text{i.i.d.}}{\sim} \nu_Y$ where $\nu_X \neq \nu_Y$ are two distinct probability measures supported on domains $\mathcal{Y}_X$ and $\mathcal{Y}_Y$, respectively. Denote the quantile maps of the distributions ν_X, ν_Y and $\theta\nu_X + (1-\theta)\nu_Y$ by Q_X, Q_Y and $Q_{X,Y}$, respectively. Assume that Q_X, Q_Y and $Q_{X,Y} : \mathrm{Int}(\mathcal{S}) \to \mathrm{Int}(\mathcal{Y}_X \cup \mathcal{Y}_Y)$ are homeomorphisms. Then, as $m, n \to \infty$,

$$T_{X,Y} \xrightarrow{a.s.} c := \int_{\mathcal{S}} \| R_{X,Y}(Q_X(u)) - R_{X,Y}(Q_Y(u)) \|^2 \, d\mu(u),$$

where $c > 0$ and $R_{X,Y}$ is the rank function for the measure $\theta\nu_X + (1-\theta)\nu_Y$.

The following two results (proved in Sections G.3 and G.4 of [33]) provide rates of convergence of $T_{X,Y}$ under the null and alternative hypotheses. The proofs of these results are built on Theorems 5.2 and 5.3.

PROPOSITION 6.3. *Suppose that $H_0 : \nu_X = \nu_Y \equiv \nu$ holds. Assume that ν is absolutely continuous and supported on a compact domain $\mathcal{Y} \subset \mathbb{R}^d$. Further, we assume that the convex potential ψ of the quantile map Q of ν (w.r.t. μ) is strongly convex. Under H_0, if* $\min\{m, n\}/(m + n) \geq \theta \in (0, 1)$, *then*

$$\mathbb{E}[T_{X,Y}] \leq C r_{d,m+n},$$

where $C \equiv C(\mu, \nu, \theta) > 0$ depends on μ, ν and θ, and $r_{d,n}$ is defined in (3).

PROPOSITION 6.4. *Suppose that $\nu_X \neq \nu_Y$, where ν_X and ν_Y are compactly supported. Recall the notation from Proposition* 6.2. *For convenience, we will assume that the pooled sample size N is fixed and that $m|N \sim$* Binomial(N, θ), *where $\theta \in (0, 1)$. Further, we assume that the convex potential functions ψ_X, ψ_Y, $\psi_{X,Y}$ of the quantile maps Q_X, Q_Y, $Q_{X,Y}$ are strongly convex. Then we have*

$$\mathbb{E}\big[|T_{X,Y} - c|\big] \leq C r_{d,N}^{1/2},$$

where $C \equiv C(\mu, \nu_X, \nu_Y, \theta) > 0$, and $r_{d,n}$ is defined in (3).

A detailed study of the finite sample performance and the asymptotic weak limit of the above test in beyond the scope of the present paper. We plan to pursue this in a future paper. As mentioned before, $T_{X,Y}$ is inspired by the form of the Cramér–von Mises (one-sample) goodness-of-fit statistic. One can, of course, use other test statistics based on the empirical quantile/rank maps for testing (25). A key observation for constructing such tests is to realize that, under H_0, $\widehat{R}_{X,Y}(X_1), \dots, \widehat{R}_{X,Y}(X_m), \widehat{R}_{X,Y}(Y_1), \dots, \widehat{R}_{X,Y}(Y_n)$ are *exchangeable* and are all marginally distributed as μ.

6.2. *Mutual independence testing*. Let $X = (X^{(1)}, \dots, X^{(k)}) \sim \nu$ be a random vector in $\mathbb{R}^d$ where $k \geq 2$ and $X^{(j)} \sim \nu_j$ is a random vector in $\mathbb{R}^{d_j}$, for $j = 1, \dots, k$, with $\sum_{j=1}^k d_j = d$. In this subsection, we consider the problem of testing the mutual independence of $X^{(1)}, \dots, X^{(k)}$. Specifically, we consider testing whether ν is equal to the product measure $\nu_1 \otimes \cdots \otimes \nu_k$, for some $\nu_1, \dots, \nu_k$, that is,

$$H_0 : \nu = \nu_1 \otimes \cdots \otimes \nu_k \quad \text{versus} \quad H_1 : \nu \neq \nu_1 \otimes \cdots \otimes \nu_k, \tag{27}$$

when we observe i.i.d. data from ν. This is again a fundamental problem in statistics and there has been many approaches investigated in the literature; see, for example, [8, 9], [42], Chapter 8, [24] and the references therein. The use of kernel (see, e.g., [35, 55, 64, 71]) and distance covariance (see e.g., [24, 59, 73, 75–77]) based methods have become very popular for this problem. Also see [5, 83] and the references therein for some recent other approaches to testing (27). We use our multivariate quantile and rank functions to construct a tuning parameter-free consistent test for (27).

For simplicity of notation, let us assume that $k = 2$. As we will see, the extension to $k > 2$ is straightforward. Let $\{Z_i \equiv (X_i, Y_i) : 1 \leq i \leq n\}$ be i.i.d. ν, assumed to be absolutely continuous on $\mathbb{R}^{d_X} \times \mathbb{R}^{d_Y}$; here, $d_X, d_Y \geq 1$ and $d_X + d_Y = d$. Further, we assume that $X \sim \nu_X$ and $Y \sim \nu_Y$. We want to test the hypothesis of mutual independence between X and Y, that is,

$$H_0 : \nu = \nu_X \otimes \nu_Y \quad \text{versus} \quad H_1 : \nu \neq \nu_X \otimes \nu_Y. \tag{28}$$

Let $\mu_X = \text{Uniform}([0,1]^{d_X})$, $\mu_Y = \text{Uniform}([0,1]^{d_Y})$ and let $\mu := \mu_X \otimes \mu_Y = \text{Uniform}([0,1]^d)$. We define $\widehat{R} : \mathbb{R}^d \to [0,1]^d$ and $\widehat{Q} : [0,1]^d \to \mathbb{R}^d$ to be the empirical rank and quantile maps of the joint sample $(X_1, Y_1), \dots, (X_n, Y_n)$. Let $\widehat{R}_X : \mathbb{R}^{d_X} \to \mathbb{R}^{d_X}$ be the empirical rank map of $X_1, \dots, X_n$; similarly, let $\widehat{R}_Y : \mathbb{R}^{d_Y} \to \mathbb{R}^{d_Y}$ be the sample rank map obtained from $Y_1, \dots, Y_n$. Define $\widetilde{R} := (\widehat{R}_X, \widehat{R}_Y) : \mathbb{R}^d \to [0,1]^d$. We consider the following test statistic:

$$T_n := \int_{[0,1]^d} \|\widehat{R}(\widehat{Q}(u)) - \widetilde{R}(\widehat{Q}(u))\|^2 \, du = \frac{1}{n} \sum_{i=1}^{n} \|\widehat{R}(Z_i) - \widetilde{R}(Z_i)\|^2.$$

Note that the above integral reduces to a finite average as $\widehat{Q}(\cdot)$ can only take n distinct values a.s. We reject the null hypothesis in (28) when T_n is large. As in Section 6.1, the critical value of the test can be computed using the permutation principle: We take a random permutation of σ of $\{1, \dots, n\}$ and consider the permuted data set $\{(X_i, Y_{\sigma(i)})\}_{i=1}^n$. The (conditional) null distribution of T_n can be computed by considering the permutation distribution of T_n (i.e., computed from the data $\{(X_i, Y_{\sigma(i)})\}_{i=1}^n$, as σ varies).

The following result, proved in Section G.5 of the Supplementary Material [33], describes the asymptotic behavior of the proposed test statistic under the null and alternative hypotheses; in particular, it shows that the power of the test converges to 1, as the sample size n increases.

PROPOSITION 6.5 (Consistency). *We have $\widehat{R}(\widehat{Q}(U)) \sim \mu$, where $U \sim \mu = \text{Uniform}([0,1]^d)$. Suppose H_0 holds in (28), that is, $\nu = \nu_X \otimes \nu_Y$. Then $\widetilde{R}(\widehat{Q}(U)) \sim \mu$. Assume further that ν_X and ν_Y are two probability measures supported on the domains $\mathcal{Y}_X \subset \mathbb{R}^{d_X}$ and $\mathcal{Y}_Y \subset \mathbb{R}^{d_Y}$ respectively. Denote the quantile maps of the measures ν_X, ν_Y and ν w.r.t. the measures $\text{Uniform}([0,1]^{d_X})$, $\text{Uniform}([0,1]^{d_Y})$ and $\text{Uniform}([0,1]^d)$ by Q_X, Q_Y and Q, respectively, where $d = d_X + d_Y$. Assume that $Q_X : (0,1)^{d_X} \to \text{Int}(\mathcal{Y}_X)$, $Q_X : (0,1)^{d_Y} \to \text{Int}(\mathcal{Y}_Y)$ and $Q : (0,1)^d \to \text{Int}(\mathcal{Y}_X \times \mathcal{Y}_Y)$ are homeomorphisms. Then, under H_0, as $n \to \infty$,*

$$T_n \xrightarrow{a.s.} 0.$$

Now suppose that $\nu \neq \nu_X \otimes \nu_Y$. Let $\overline{R} = (R_X, R_Y)$ where R_X and R_Y are the rank maps of ν_X and ν_Y, respectively. Then

$$T_n \xrightarrow{a.s.} c := \int_{[0,1]^d} \|u - \overline{R}(Q(u))\|^2 \, du, \quad \text{as } n \to \infty.$$

The following two results (proved in Sections G.6 and G.7 of [33]) provide rates of convergence of T_n under the null and alternative hypotheses.

PROPOSITION 6.6. *Suppose $H_0 : \nu = \nu_X \otimes \nu_Y$ holds, where ν_X and ν_Y are compactly supported absolutely continuous distributions on $\mathbb{R}^{d_X}$ and $\mathbb{R}^{d_Y}$ with quantile maps Q_X and Q_Y. Further, assume that the convex potentials ψ_X and ψ_Y of Q_X and Q_Y are strongly convex. Then, for $d = d_X + d_Y$,*

$$\mathbb{E}[T_n] \leq C r_{d,n},$$

where $C \equiv C(\mu, \nu) > 0$ depends on μ and ν, and $r_{d,n}$ is defined in (3).

PROPOSITION 6.7. *Suppose that $\nu \neq \nu_X \otimes \nu_Y$, where ν_X and ν_Y are compactly supported. Recall the notation from Proposition 6.5. Further, we assume that the convex potential functions ψ_X, ψ_Y, ψ of the quantile maps Q_X, Q_Y, Q are strongly convex functions. Then we have*

$$\mathbb{E}[|T_n - c|] \leq C r_{d,n}^{1/2},$$

where $C \equiv C(\mu, \nu) > 0$ and $r_{d,n}$ is defined in (3).

Acknowledgments. The authors are extremely grateful to Peng Xu for creating the R-package https://github.com/Francis-Hsu/testOTM (see [85]) for the computation of all the estimators studied in this paper. In particular, all of the plots in the paper are obtained from his R-package. The authors would like to thank Nabarun Deb, Adityanand Guntuboyina, Marc Hallin and Johan Segers for helpful discussions. The authors also acknowledge the many insightful comments by the two anonymous referees that helped improve the paper.

Funding. The second author was supported by NSF Grant DMS-2015376.

SUPPLEMENTARY MATERIAL

Supplement to "Multivariate ranks and quantiles using optimal transport: Consistency, rates and nonparametric testing" (DOI: 10.1214/21-AOS2136SUPP; .pdf). This supplementary material [33] contains proofs of all results in the main paper; other auxiliary results (with their proofs) alluded to in the main paper and further discussions.

REFERENCES

[1] AURENHAMMER, F. (1987). Power diagrams: Properties, algorithms and applications. *SIAM J. Comput.* **16** 78–96. MR0873251 https://doi.org/10.1137/0216006

[2] AURENHAMMER, F., HOFFMANN, F. and ARONOV, B. (1998). Minkowski-type theorems and least-squares clustering. *Algorithmica* **20** 61–76. MR1483422 https://doi.org/10.1007/PL00009187

[3] AURENHAMMER, F., KLEIN, R. and LEE, D.-T. (2013). *Voronoi Diagrams and Delaunay Triangulations*. World Scientific Co. Pte. Ltd., Hackensack, NJ. MR3186045 https://doi.org/10.1142/8685

[4] BARINGHAUS, L. and FRANZ, C. (2004). On a new multivariate two-sample test. *J. Multivariate Anal.* **88** 190–206. MR2021870 https://doi.org/10.1016/S0047-259X(03)00079-4

[5] BERRETT, T. B. and SAMWORTH, R. J. (2019). Nonparametric independence testing via mutual information. *Biometrika* **106** 547–566. MR3992389 https://doi.org/10.1093/biomet/asz024

[6] BHATTACHARYA, B. B. (2020). Asymptotic distribution and detection thresholds for two-sample tests based on geometric graphs. *Ann. Statist.* **48** 2879–2903. MR4152627 https://doi.org/10.1214/19-AOS1913

[7] BICKEL, P. J. (1968). A distribution free version of the Smirnov two sample test in the p-variate case. *Ann. Math. Stat.* **40** 1–23. MR0256519 https://doi.org/10.1214/aoms/1177697800

[8] BLOMQVIST, N. (1950). On a measure of dependence between two random variables. *Ann. Math. Stat.* **21** 593–600. MR0039190 https://doi.org/10.1214/aoms/1177729754

[9] BLUM, J. R., KIEFER, J. and ROSENBLATT, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Stat.* **32** 485–498. MR0125690 https://doi.org/10.1214/aoms/1177705055

[10] BOBKOV, S. and LEDOUX, M. (2019). One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *Mem. Amer. Math. Soc.* **261** v+126. MR4028181 https://doi.org/10.1090/memo/1259

[11] BOECKEL, M., SPOKOINY, V. and SUVORIKOVA, A. (2018). Multivariate Brenier cumulative distribution functions and their application to non-parametric testing. Preprint. Available at arXiv:1809.04090.

[12] BRENIER, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44** 375–417. MR1100809 https://doi.org/10.1002/cpa.3160440402

[13] CAFFARELLI, L. A. (1992). The regularity of mappings with a convex potential. *J. Amer. Math. Soc.* **5** 99–104. MR1124980 https://doi.org/10.2307/2152752

[14] CAFFARELLI, L. A. (1992). Boundary regularity of maps with convex potentials. *Comm. Pure Appl. Math.* **45** 1141–1151. MR1177479 https://doi.org/10.1002/cpa.3160450905

[15] CAFFARELLI, L. A. (1996). Boundary regularity of maps with convex potentials. II. *Ann. of Math.* (2) **144** 453–496. MR1426885 https://doi.org/10.2307/2118564

[16] CAFFARELLI, L. A., KOCHENGIN, S. A. and OLIKER, V. I. (1999). On the numerical solution of the problem of reflector design with given far-field scattering data. In *Monge Ampère Equation: Applications to Geometry and Optimization* (*Deerfield Beach, FL*, 1997). *Contemp. Math.* **226** 13–32. Amer. Math. Soc., Providence, RI. MR1660740 https://doi.org/10.1090/conm/226/03233

[17] CHAUDHURI, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* **91** 862–872. MR1395753 https://doi.org/10.2307/2291681

[18] CHERNOZHUKOV, V., GALICHON, A., HALLIN, M. and HENRY, M. (2017). Monge-Kantorovich depth, quantiles, ranks and signs. *Ann. Statist.* **45** 223–256. MR3611491 https://doi.org/10.1214/16-AOS1450

[19] CHIZAT, L., ROUSSILLON, P., LÉGER, F., VIALARD, F.-X. and PEYRÉ, G. (2020). Faster Wasserstein distance estimation with the Sinkhorn divergence. *Adv. Neural Inf. Process. Syst.* **33**.

[20] CORDERO-ERAUSQUIN, D. and FIGALLI, A. (2019). Regularity of monotone transport maps between unbounded domains. *Discrete Contin. Dyn. Syst.* **39** 7101–7112. MR4026183 https://doi.org/10.3934/dcds.2019297

[21] DE PHILIPPIS, G. and FIGALLI, A. (2014). The Monge-Ampère equation and its link to optimal transportation. *Bull. Amer. Math. Soc.* (*N.S.*) **51** 527–580. MR3237759 https://doi.org/10.1090/S0273-0979-2014-01459-4

[22] DE VALK, C. and SEGERS, J. (2018). Tails of optimal transport plans for regularly varying probability measures. Preprint. Available at arXiv:1811.12061.

[23] DEB, N., BHATTACHARYA, B. B. and SEN, B. (2021). Efficiency lower bounds for distribution-free Hotelling-type two-sample tests based on optimal transport. Preprint. Available at arXiv:2104.01986.

[24] DEB, N. and SEN, B. (2019). Multivariate rank-based distribution-free nonparametric testing using measure transportation. Preprint. Available at arXiv:1909.08733.

[25] DEL BARRIO, E., GINÉ, E. and UTZET, F. (2005). Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli* **11** 131–189. MR2121458 https://doi.org/10.3150/bj/1110228245

[26] 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. *J. Multivariate Anal.* **180** 104671, 13. MR4147635 https://doi.org/10.1016/j.jmva.2020.104671

[27] DEL BARRIO, E., GORDALIZA, P., LESCORNEL, H. and LOUBES, J.-M. (2019). Central limit theorem and bootstrap procedure for Wasserstein's variations with an application to structural relationships between distributions. *J. Multivariate Anal.* **169** 341–362. MR3875604 https://doi.org/10.1016/j.jmva.2018.09.014

[28] DEL BARRIO, E. and LOUBES, J.-M. (2019). Central limit theorems for empirical transportation cost in general dimension. *Ann. Probab.* **47** 926–951. MR3916938 https://doi.org/10.1214/18-AOP1275

[29] DUDLEY, R. M. (2002). *Real Analysis and Probability*. *Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358 https://doi.org/10.1017/CBO9780511755347

[30] FIGALLI, A. (2018). On the continuity of center-outward distribution and quantile functions. *Nonlinear Anal.* **177** 413–421. MR3886582 https://doi.org/10.1016/j.na.2018.05.008

[31] FOURNIER, N. and GUILLIN, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** 707–738. MR3383341 https://doi.org/10.1007/s00440-014-0583-7

[32] FRIEDMAN, J. H. and RAFSKY, L. C. (1979). Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *Ann. Statist.* **7** 697–717. MR0532236

[33] GHOSAL, P. and SEN, B. (2022). Supplement to "Multivariate ranks and quantiles using optimal transport: Consistency, rates and nonparametric testing." https://doi.org/10.1214/21-AOS2136SUPP

[34] GIGLI, N. (2011). On Hölder continuity-in-time of the optimal transport map towards measures along a curve. *Proc. Edinb. Math. Soc.* (2) **54** 401–409. MR2794662 https://doi.org/10.1017/S001309150800117X

[35] GRETTON, A., HERBRICH, R., SMOLA, A., BOUSQUET, O. and SCHÖLKOPF, B. (2005). Kernel methods for measuring independence. *J. Mach. Learn. Res.* **6** 2075–2129. MR2249882

[36] GU, X., LUO, F., SUN, J. and YAU, S.-T. (2016). Variational principles for Minkowski type problems, discrete optimal transport, and discrete Monge-Ampere equations. *Asian J. Math.* **20** 383–398. MR3480024 https://doi.org/10.4310/AJM.2016.v20.n2.a7

[37] HALLIN, M., DEL BARRIO, E., CUESTA-ALBERTOS, J. and MATRÁN, C. (2021). Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *Ann. Statist.* **49** 1139–1165. MR4255122 https://doi.org/10.1214/20-aos1996

[38] HALLIN, M., MORDANT, G. and SEGERS, J. (2021). Multivariate goodness-of-fit tests based on Wasserstein distance. *Electron. J. Stat.* **15** 1328–1371. MR4255302 https://doi.org/10.1214/21-ejs1816

[39] HALLIN, M., PAINDAVEINE, D. and ŠIMAN, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth. *Ann. Statist.* **38** 635–669. MR2604670 https://doi.org/10.1214/09-AOS723

[40] HALLIN, M. and WERKER, B. J. M. (2003). Semi-parametric efficiency, distribution-freeness and invariance. *Bernoulli* **9** 137–165. MR1963675 https://doi.org/10.3150/bj/1068129013

[41] HOEFFDING, W. (1952). The large-sample power of tests based on permutations of observations. *Ann. Math. Stat.* **23** 169–192. MR0057521 https://doi.org/10.1214/aoms/1177729436

[42] HOLLANDER, M. and WOLFE, D. A. (1999). *Nonparametric Statistical Methods*, 2nd ed. *Wiley Series in Probability and Statistics*: *Texts and References Section*. Wiley, New York. MR1666064

[43] HÜTTER, J.-C. and RIGOLLET, P. (2021). Minimax estimation of smooth optimal transport maps. *Ann. Statist.* **49** 1166–1194. MR4255123 https://doi.org/10.1214/20-aos1997

[44] KIM, I., BALAKRISHNAN, S. and WASSERMAN, L. (2020). Minimax optimality of permutation tests. Preprint. Available at arXiv:2003.13208.

[45] KITAGAWA, J. (2014). An iterative scheme for solving the optimal transportation problem. *Calc. Var. Partial Differential Equations* **51** 243–263. MR3247388 https://doi.org/10.1007/s00526-013-0673-x

[46] KITAGAWA, J. and MCCANN, R. (2019). Free discontinuities in optimal transport. *Arch. Ration. Mech. Anal.* **232** 1505–1541. MR3928755 https://doi.org/10.1007/s00205-018-01348-3

[47] KITAGAWA, J., MÉRIGOT, Q. and THIBERT, B. (2019). Convergence of a Newton algorithm for semi-discrete optimal transport. *J. Eur. Math. Soc. (JEMS)* **21** 2603–2651. MR3985609 https://doi.org/10.4171/JEMS/889

[48] KLATT, M., TAMELING, C. and MUNK, A. (2020). Empirical regularized optimal transport: Statistical theory and applications. *SIAM J. Math. Data Sci.* **2** 419–443. MR4105566 https://doi.org/10.1137/19M1278788

[49] KOLTCHINSKII, V. I. (1997). M-estimation, convexity and quantiles. *Ann. Statist.* **25** 435–477. MR1439309 https://doi.org/10.1214/aos/1031833659

[50] LEHMANN, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks. Holden-Day Series in Probability and Statistics*. Holden-Day, San Francisco, CA; McGraw-Hill, New York–Düsseldorf. With the special assistance of H. J. M. d'Abrera. MR0395032

[51] LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. MR2135927

[52] LÉVY, B. (2015). A numerical algorithm for L_2 semi-discrete optimal transport in 3D. *ESAIM Math. Model. Numer. Anal.* **49** 1693–1715. MR3423272 https://doi.org/10.1051/m2an/2015055

[53] LI, W. and NOCHETTO, R. H. (2021). Quantitative stability and error estimates for optimal transport plans. *IMA J. Numer. Anal.* **41** 1941–1965. MR4286252 https://doi.org/10.1093/imanum/draa045

[54] LIU, R. Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405–414. MR1041400 https://doi.org/10.1214/aos/1176347507

[55] LYONS, R. (2013). Distance covariance in metric spaces. *Ann. Probab.* **41** 3284–3305. MR3127883 https://doi.org/10.1214/12-AOP803

[56] MCCANN, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.* **80** 309–323. MR1369395 https://doi.org/10.1215/S0012-7094-95-08013-2

[57] MÉRIGOT, Q. (2011). A multiscale approach to optimal transport. In *Computer Graphics Forum* **30** 1583–1592. Wiley, New York.

[58] MÉRIGOT, Q. and THIBERT, B. (2021). Optimal transport: Discretization and algorithms. In *Geometric Partial Differential Equations. Part II. Handb. Numer. Anal.* **22** 133–212. Elsevier/North-Holland, Amsterdam. MR4254135 https://doi.org/10.1016/bs.hna.2020.10.001

[59] MÓRI, T. F. and SZÉKELY, G. J. (2019). Four simple axioms of dependence measures. *Metrika* **82** 1–16. MR3897521 https://doi.org/10.1007/s00184-018-0670-3

[60] OJA, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.* **1** 327–332. MR0721446 https://doi.org/10.1016/0167-7152(83)90054-8

[61] OLIKER, V. I. and PRUSSNER, L. D. (1988). On the numerical solution of the equation $(\partial^2 z/\partial x^2)(\partial^2 z/\partial y^2) - ((\partial^2 z/\partial x \partial y))^2 = f$ and its discretizations. I. *Numer. Math.* **54** 271–293. MR0971703 https://doi.org/10.1007/BF01396762

[62] PANARETOS, V. M. and ZEMEL, Y. (2019). Statistical aspects of Wasserstein distances. *Annu. Rev. Stat. Appl.* **6** 405–431. MR3939527 https://doi.org/10.1146/annurev-statistics-030718-104938

[63] PEYRÉ, G., CUTURI, M. et al. (2019). Computational optimal transport: With applications to data science. *Found. Trends Mach. Learn.* **11** 355–607.

[64] PFISTER, N., BÜHLMANN, P., SCHÖLKOPF, B. and PETERS, J. (2018). Kernel-based tests for joint independence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 5–31. MR3744710 https://doi.org/10.1111/rssb.12235

[65] RAMDAS, A., GARCÍA TRILLOS, N. and CUTURI, M. (2017). On Wasserstein two-sample testing and related families of nonparametric tests. *Entropy* **19** Paper No. 47, 15. MR3608466 https://doi.org/10.3390/e19020047

[66] RIGOLLET, P. and WEED, J. (2018). Entropic optimal transport is maximum-likelihood deconvolution. *C. R. Math. Acad. Sci. Paris* **356** 1228–1235. MR3907589 https://doi.org/10.1016/j.crma.2018.10.010

[67] RIGOLLET, P. and WEED, J. (2019). Uncoupled isotonic regression via minimum Wasserstein deconvolution. *Inf. Inference* **8** 691–717. MR4045481 https://doi.org/10.1093/imaiai/iaz006

[68] ROSENBAUM, P. R. (2005). An exact distribution-free test comparing two multivariate distributions based on adjacency. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 515–530. MR2168202 https://doi.org/10.1111/j.1467-9868.2005.00513.x

[69] ROUSSEEUW, P. J. and RUTS, I. (1999). The depth function of a population distribution. *Metrika* **49** 213–244. MR1731769

[70] SCHILLING, M. F. (1986). Multivariate two-sample tests based on nearest neighbors. *J. Amer. Statist. Assoc.* **81** 799–806. MR0860514

[71] SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41** 2263–2291. MR3127866 https://doi.org/10.1214/13-AOS1140

[72] SERFLING, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardisation. *J. Nonparametr. Stat.* **22** 915–936. MR2738875 https://doi.org/10.1080/10485250903431710

[73] SHI, H., DRTON, M. and HAN, F. (2019). Distribution-free consistent independence tests via Hallin's multivariate rank. Preprint. Available at arXiv:1909.10024.

[74] SHI, H., HALLIN, M., DRTON, M. and HAN, F. (2020). Rate-optimality of consistent distribution-free tests of independence based on center-outward ranks and signs. Preprint. Available at arXiv:2007.02186.

[75] SZÉKELY, G. J. and RIZZO, M. L. (2009). Brownian distance covariance. *Ann. Appl. Stat.* **3** 1236–1265. MR2752127 https://doi.org/10.1214/09-AOAS312

[76] SZÉKELY, G. J. and RIZZO, M. L. (2013). Energy statistics: A class of statistics based on distances. *J. Statist. Plann. Inference* **143** 1249–1272. MR3055745 https://doi.org/10.1016/j.jspi.2013.03.018

[77] SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665 https://doi.org/10.1214/009053607000000505

[78] VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

[79] VILLANI, C. (2003). *Topics in Optimal Transportation. Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483 https://doi.org/10.1090/gsm/058

[80] VILLANI, C. (2009). *Optimal Transport: Old and new. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **338**. Springer, Berlin. MR2459454 https://doi.org/10.1007/978-3-540-71050-9

[81] WANG, X.-J. (2017). Monge-Ampère equation and optimal transportation. In *Proceedings of the Sixth International Congress of Chinese Mathematicians. Vol. i. Adv. Lect. Math. (ALM)* **36** 153–172. Int. Press, Somerville, MA. MR3702074

[82] 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 https://doi.org/10.3150/18-BEJ1065

[83] WEIHS, L., DRTON, M. and MEINSHAUSEN, N. (2018). Symmetric rank covariances: A generalized framework for nonparametric measures of dependence. *Biometrika* **105** 547–562. MR3842884 https://doi.org/10.1093/biomet/asy021

[84] WEISS, L. (1960). Two-sample tests for multivariate distributions. *Ann. Math. Stat.* **31** 159–164. MR0119305 https://doi.org/10.1214/aoms/1177705995

[85] XU, P. (2019). testOTM: Multivariate Ranks and Quantiles using Optimal Transportation. R package version 1.00.0.

[86] ZEMEL, Y. and PANARETOS, V. M. (2019). Fréchet means and Procrustes analysis in Wasserstein space. *Bernoulli* **25** 932–976. MR3920362 https://doi.org/10.3150/17-bej1009

[87] ZUO, Y. (2003). Projection-based depth functions and associated medians. *Ann. Statist.* **31** 1460–1490. MR2012822 https://doi.org/10.1214/aos/1065705115