Monge-Kantorovich depth, quantiles, ranks and signs

Victor Chernozhukov
Alfred Galichon
Marc Hallin
Marc Henry

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP04/15

MONGE-KANTOROVICH DEPTH, QUANTILES, RANKS AND SIGNS

VICTOR CHERNOZHUKOV, ALFRED GALICHON, MARC HALLIN, AND MARC HENRY

ABSTRACT. We propose new concepts of statistical depth, multivariate quantiles, ranks and signs, based on canonical transportation maps between a distribution of interest on $\mathbb{R}^d$ and a reference distribution on the d-dimensional unit ball. The new depth concept, called *Monge-Kantorovich depth*, specializes to halfspace depth in the case of elliptical distributions, but, for more general distributions, differs from the latter in the ability for its contours to account for non convex features of the distribution of interest. We propose empirical counterparts to the population versions of those Monge-Kantorovich depth contours, quantiles, ranks and signs, and show their consistency by establishing a uniform convergence property for empirical transport maps, which is of independent interest.

AMS 1980 subject classification: 62M15, 62G35.

Keywords: Statistical depth, vector quantiles, vector ranks, multivariate signs, empirical transport maps, uniform convergence of empirical transport.

1. INTRODUCTION

The concept of statistical depth was introduced in order to overcome the lack of a canonical ordering in $\mathbb{R}^d$ for $d > 1$, hence the absence of the related notions of quantile and distribution functions, ranks, and signs. The earliest and most popular depth concept is *halfspace depth*, the definition of which goes back to Tukey [48]. Since then, many other concepts have been considered: *simplicial depth* [33], *majority depth* ([46] and [36]), *projection depth* ([34], building on [47]

Date: January 27, 2015.

Chernozhukov's research was supported by the NSF. Galichon's research was funded by the European Research Council under the European Unions Seventh Framework Programme (FP7/2007-2013), ERC grant agreement no 313699. Hallin's research was supported by the IAP research network grant P7/06 of the Belgian government (Belgian Science Policy) and the Discovery grant DP150100210 of the Australian Research Council. Henry's research was supported by SSHRC Grant 435-2013-0292 and NSERC Grant 356491-2013.

and [11], [54]), *Mahalanobis depth* ([37], [34], [36]), *Oja depth* [41], *zonoid depth* ([30] and [29]), *spatial depth* ([32], [40], [6], [50]), L^p depth [55], among many others. An axiomatic approach, aiming at unifying all those concepts, was initiated by Liu [33] and Zuo and Serfling [55], who list four properties that are generally considered desirable for any statistical depth function, namely affine invariance, maximality at the center, linear monotonicity relative to the deepest points, and vanishing at infinity (see Section 2.2 for details). Halfspace depth $\mathrm{D}^{\mathrm{Tukey}}$ is the prototype of a depth concept satisfying the Liu-Zuo-Serfling axioms for the family $\mathcal{P}$ of all absolutely continuous distributions on $\mathbb{R}^d$.

An important feature of halfspace depth is the convexity of its contours, which thus satisfy the star-convexity requirement embodied in the linear monotonicity axiom. That feature is shared by most existing depth concepts and might be considered undesirable for distributions with non convex supports or level contours, and multimodal ones. Proposals have been made, under the name of *local depths*, to deal with this, while retaining the spirit of the Liu-Zuo-Serfling axioms: see [7], [27], [1], and [42] who provide an in-depth discussion of those various attempts. In this paper, we take a totally different and more agnostic approach, on the model of the discussion by Serfling in [45]: if the ultimate purpose of statistical depth is to provide, for each distribution P, a P-related ordering of $\mathbb{R}^d$ producing adequate concepts of quantile and distribution functions, ranks and signs, the relevance of a given depth function should be evaluated in terms of the relevance of the resulting ordering, and the quantiles, ranks and signs it produces.

Now, the concepts of quantiles, ranks and signs are well understood in two particular cases, essentially, that should serve as benchmarks. The first case is that of the family $\mathcal{P}^1$ of all distributions with nonvanishing Lebesgue densities over the real line. Here, the concepts of quantile and distribution functions, ranks, and signs are related to the "classical" univariate ones. The second case is that of the family $\mathcal{P}^d_{\mathrm{ell}}$ of all full-rank elliptical distributions over $\mathbb{R}^d$ $(d > 1)$ with nonvanishing radial densities. There, elliptical contours with P-probability contents τ provide a natural definition of *τ-quantile contours*, while the ranks and unit vectors associated with sphericized observations have proven to be adequate concepts of multivariate ranks and signs, as shown in [19], [20], [21] and [22]: call them *elliptical quantiles, ranks* and *signs.* In both cases, the relevance of ranks and signs, whether traditional or elliptical, is related to their role as *maximal* invariants under a group of transformations *minimally* generating $\mathcal{P}$, of which distribution-freeness is just a by-product, as explained in [23]. We argue that an adequate depth function, when restricted to those two particular cases, should lead to the same well-established concepts: classical quantiles, ranks and signs for $\mathcal{P}^1$, elliptical ones for $\mathcal{P}^d_{\mathrm{ell}}$.

A closer look at halfspace depth in those two particular cases reveals that the halfspace depth contours are the images of the hyperspheres with radii $\tau \in [0, 1]$

centered at the origin, by a map Q that is the gradient of a convex function. That mapping Q actually is the essentially unique gradient of a convex function that transports the spherical uniform distribution U_d on the unit ball $\mathbb{S}^d$ of $\mathbb{R}^d$, (i.e., the distribution of a random vector $r\varphi$, where r is uniform on $[0,1]$, φ is uniform on the unit sphere $\mathcal{S}^{d-1}$, and r and φ are mutually independent) into the univariate or elliptical distribution of interest P. By McCann's [38] extension of Brenier's celebrated Polar Factorization Theorem [4], such gradient of a convex function Q_P transporting U_d into P exists, and is an essentially unique, for any distribution P on $\mathbb{R}^d$—not just the elliptical ones. Moreover, when P has finite moments of order two, that mapping Q_P coincides with the L^2-optimal transport map, in the sense of measure transportation, of the spherical distribution U_d to P. This suggests a new concept of statistical depth, which we call the *Monge-Kantorovich depth* D^{MK}, the contours and signs of which are obtained as the images by Q_P of the hyperspheres with radius $\tau \in [0,1]$ centered at the origin and their unit rays. When restricted to $\mathcal{P}^1$ or $\mathcal{P}^d_{\mathrm{ell}}$, Monge-Kantorovich and halfspace depths coincide, and affine-invariance is preserved. For $P \in \mathcal{P}^d \setminus \mathcal{P}^d_{\mathrm{ell}}$ with $d > 1$, the two concepts are distinct, and Monge-Kantorovich depth is no longer affine-invariant.

Under suitable regularity conditions due to Caffarelli (see [51], Section 4.2.2), Q_P is a homeomorphism, and its inverse $\mathrm{R}_P := \mathrm{Q}_P^{-1}$ is also the gradient of a convex function; the Monge-Kantorovich depth contours are continuous and the corresponding depth regions are nested, so that Monge-Kantorovich depth indeed provides a center-outward ordering of $\mathbb{R}^d$, namely,

$$x_2 \geq_{\mathrm{D}_P^{\mathrm{MK}}} x_1 \text{ if and only if } \|\mathrm{R}_P(x_2)\| \leq \|\mathrm{R}_P(x_1)\|.$$

Thus, our approach based on the theory of measure transportation allows us to define

(a) a *vector quantile* map Q_P, and the associated *quantile* correspondence, which maps $\tau \in [0,1]$ to $\mathrm{Q}_P(\mathcal{S}(\tau))$,
(b) a *vector rank* (or *signed rank*) function R_P, which can be decomposed into a *rank* function from $\mathbb{R}^d$ to $[0,1]$, with $r_P(x) := \|\mathrm{R}_P(x)\|$, and a *sign* function u_P, mapping $x \in \mathbb{R}^d$ to $u_P(x) := \mathrm{R}_P(x)/\|\mathrm{R}_P(x)\| \in \mathcal{S}^{d-1}$.

We call them *Monge-Kantorovich quantiles, ranks* and *signs.*

To the best of our knowledge, this is the first proposal of measure transportation-based depth concept—hence the first attempt to provide a measure-driven ordering of $\mathbb{R}^d$ based on measure transportation theory. That ordering, namely $\geq_{\mathrm{D}_P^{\mathrm{MK}}}$, is canonical in the sense that it is invariant under shifts, multiplication by a non zero scalar, orthogonal transformations, and combinations thereof; so are the Monge-Kantorovitch ranks. Previous proposals have been made, however, of measure transportation-based vector quantile functions in Ekeland, Galichon and

Henry [15] and Galichon and Henry [16]. Carlier, Chernozhukov and Galichon [5] extended the notion to vector quantile regression, creating a vector analogue of Koenker and Basset's [28] scalar quantile regression. More recently, Decurninge [9] proposed a new concept of multivariate L^p moments based upon the same notion. In these contributions, however, the focus is not statistical depth and the associated quantiles and ranks, and the leading case for the reference distribution is uniform on the unit hypercube in $\mathbb{R}^d$, as opposed to the spherical uniform distribution U_d we adopt here as leading case, while pointing out that other reference distributions may be entertained, such as the standard Gaussian distribution on $\mathbb{R}^d$ or the uniform on the hypercube $[0,1]^d$ as mentioned above.

Empirical versions of Monge-Kantorovich vector quantiles are obtained as the essentially unique gradient $\hat{\mathrm{Q}}_n$ of a convex function from (some estimator of) the reference distribution to some estimator $\hat{P}_n$ of the distribution of interest P. In case of smooth estimators, where $\hat{P}_n$ satisfies Caffarelli regularity conditions, empirical ranks, depth and depth contours are defined identically to their theoretical counterparts, and possess the same properties. In case of discrete estimators, such as the empirical distribution of a sample drawn from P, $\hat{\mathrm{Q}}_n$ is not invertible, and empirical vector ranks and depth can be defined as multi-valued mappings or selections from the latter. In all cases, we prove uniform convergence of Monge-Kantorovich empirical depth and quantile contours, vector quantiles and vector ranks, ranks and signs to their theoretical counterparts, as a special case of a new result on uniform convergence of optimal transport maps, which is of independent interest.

Notation, conventions and preliminaries. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be some probability space. Throughout, $\mathcal{P}$ denotes a class of probability distributions over $\mathbb{R}^d$. Unless otherwise specified, it is the class of all Borel probability measures on $\mathbb{R}^d$. Denote by $\mathbb{S}^d := \{x \in \mathbb{R}^d : \|x\| \leq 1\}$ the unit ball, and by $\mathcal{S}^{d-1} := \{x \in \mathbb{R}^d : \|x\| = 1\}$ the unit sphere, in $\mathbb{R}^d$. For $\tau \in (0,1]$, $\mathbb{S}(\tau) := \{x \in \mathbb{R}^d : \|x\| \leq \tau\}$ is the ball, and $\mathcal{S}(\tau) := \{x \in \mathbb{R}^d : \|x\| = \tau\}$ the sphere, of radius τ. Let P_X stand for the distribution of the random vector X. Following Villani [51], we denote by $g\#\mu$ the *image measure* (or *push-forward*) of a measure $\mu \in \mathcal{P}$ by a measurable map $g : \mathbb{R}^d \to \mathbb{R}^d$. Explicitly, $g\#\mu(A) := \mu(g^{-1}(A))$ for any Borel set A. For a Borel subset $\mathbb{D}$ of a vector space equipped with the norm $\|\cdot\|$ and $f : \mathbb{D} \mapsto \mathbb{R}$, let

$$\|f\|_{\mathrm{BL}(\mathbb{D})} := \sup_x |f(x)| \vee \sup_{x \neq x'} |f(x) - f(x')| \|x - x'\|^{-1}.$$

For two probability distributions P and P' on a measurable space $\mathbb{D}$, define the bounded Lipschitz metric as

$$d_{\mathrm{BL}}(P, P') := \|P - P'\|_{\mathrm{BL}} := \sup_{\|f\|_{\mathrm{BL}(\mathbb{D})} \leq 1} \int f d(P - P'),$$

which metrizes the topology of weak convergence. A *convex* function ψ on $\mathbb{R}^d$ refers to a function $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ for which $\psi((1-t)x+tx') \leq (1-t)\psi(x)+t\psi(x')$ for any (x, x') such that $\psi(x)$ and $\psi(x')$ are finite and for any $t \in (0,1)$. Such a function is continuous on the interior of the convex set dom $\psi := \{x \in \mathbb{R}^d : \psi(x) < \infty\}$, and differentiable Lebesgue almost everywhere in dom ψ. Write $\nabla\psi$ for the gradient of ψ. For any function $\psi : \mathbb{R}^d \mapsto \mathbb{R} \cup \{+\infty\}$, the *conjugate* $\psi^* : \mathbb{R}^d \mapsto \mathbb{R} \cup \{+\infty\}$ of ψ is defined for each $y \in \mathbb{R}^d$ by $\psi^*(y) = \sup_{z \in \mathbb{R}^d} y^\top z - \psi(z)$. The conjugate ψ^* of ψ is a convex lower-semi-continuous function. Let $\mathcal{U}$ and $\mathcal{Y}$ be convex, closed subsets of $\mathbb{R}^d$. We shall call *conjugate pair* a pair of functions $\mathcal{U} \to \mathbb{R} \cup \{+\infty\}$ that are conjugates of each other. The transpose of a matrix A is denoted $A^\top$. Finally, we call *weak order* a complete reflexive and transitive binary relation.

Outline of the paper. Section 2 introduces and motivates a new notion of statistical depth, vector quantiles and vector ranks based on optimal transport maps. Section 3 describes estimators of depth contours, quantiles and ranks, and proves consistency of these estimators. Additional results and proofs are collected in the appendix.

2. Statistical depth and vector ranks and quantiles

2.1. Statistical depth, regions and contours.

The notion of statistical depth serves to define a center-outward ordering of points in the support of a distribution on $\mathbb{R}^d$, for $d > 1$. As such, it emulates the notion of quantile for distributions on the real line. We define it as a real-valued index on $\mathbb{R}^d$ as follows.

Definition (Statistical depth index and ordering)**.** A depth function is a mapping

$$\begin{array}{rccc} \mathrm{D} : & \mathbb{R}^d \times \mathcal{P} & \longrightarrow & \mathbb{R}_+ \\ & (x, P) & \longmapsto & \mathrm{D}_P(x), \end{array}$$

and $\mathrm{D}_P(x)$ is called the *depth of* x *relative to* P. For each $P \in \mathcal{P}$, the *depth ordering* $\geq_{\mathrm{D}_P}$ *associated with* D_P is the weak order on $\mathbb{R}^d$ defined for each $(x_1, x_2) \in \mathbb{R}^{2d}$ by

$$x_1 \geq_{\mathrm{D}_P} x_2 \text{ if and only if } \mathrm{D}_P(x_1) \geq \mathrm{D}_P(x_2),$$

in which case x_1 is said to be *deeper* than x_2 relative to P.

The depth function thus defined allows graphical representations of the distribution P through depth contours, which are collections of points of equal depth relative to P.

Definition (Depth regions and contours)**.** Let D_P be a depth function relative to distribution P on $\mathbb{R}^d$.

(1) The *region of depth* d (hereafter d*-depth region*) associated with D_P is defined as $\mathbb{C}_P(d) = \{x \in \mathbb{R}^d : \ \mathrm{D}_P(x) \geq d\}$.
(2) The *contour of depth* d (hereafter d*-depth contour*) associated with D_P is defined as $\mathcal{C}_P(d) = \{x \in \mathbb{R}^d : \ \mathrm{D}_P(x) = d\}$.

By construction, the depth regions relative to any distribution P are nested, i.e.,

$$\forall (d, d') \in \mathbb{R}^2_+, \ d' \geq d \implies \mathbb{C}_P(d') \subseteq \mathbb{C}_P(d).$$

Hence, the depth ordering qualifies as a *center-outward ordering* of points in $\mathbb{R}^d$.

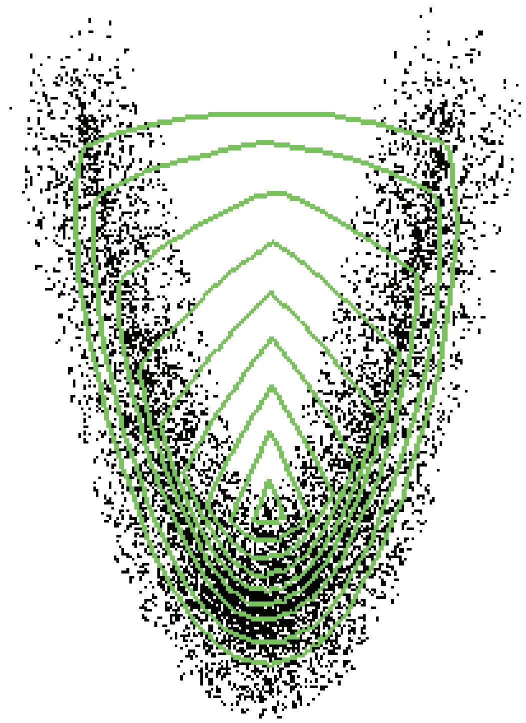

FIGURE 1. Tukey Halfspace depth contours for a banana-shaped distribution, produced with the algorithm of Paindaveine and Šiman [43] from a sample of 9999 observations. The banana-like geometry of the data cloud is not picked by the convex contours, and the deepest point is close to the boundary of the support.

2.2. **Liu-Zuo-Serfling axioms and halfspace depth.** The four axioms proposed by Liu [33] and Zuo and Serfling [55] to unify the diverse depth functions proposed in the literature are the following.

(A1) (Affine invariance) $\mathrm{D}_{P_{AX+b}}(Ax+b) = \mathrm{D}_{P_X}(x)$ for any $x \in \mathbb{R}^d$, any full-rank $d \times d$ matrix A, and any $b \in \mathbb{R}^d$.
(A2) (Maximality at the *center*) If x_0 is a *center of symmetry* for P (symmetry here can be either *central*, *angular* or *halfspace* symmetry), it is *deepest*, that is, $\mathrm{D}_P(x_0) = \max_{x\in\mathbb{R}^d} \mathrm{D}_P(x)$.
(A3) (Linear monotonicity relative to the deepest points) If $\mathrm{D}_P(x_0)$ is equal to $\max_{x\in\mathbb{R}^d} \mathrm{D}_P(x)$, then $\mathrm{D}_P(x) \leq \mathrm{D}_P((1-\alpha)x_0 + \alpha x)$ for all $\alpha \in [0,1]$ and $x \in \mathbb{R}^d$: depth is monotonically decreasing along any straight line running through a deepest point.
(A4) (Vanishing at infinity) $\lim_{\|x\|\to\infty} \mathrm{D}_P(x) = 0$.

The earliest and most popular depth function is *halfspace depth* proposed by Tukey [48]:

Definition (Halfspace depth)**.** The halfspace depth $\mathrm{D}_P^{\mathrm{Tukey}}(x)$ of a point $x \in \mathbb{R}^d$ with respect to the distribution P_X of a random vector X on $\mathbb{R}^d$ is defined as

$$\mathrm{D}_{P_X}^{\mathrm{Tukey}}(x) := \min_{\varphi\in\mathcal{S}^{d-1}} \mathbb{P}[(X-x)^\top\varphi \geq 0].$$

Halfspace depth relative to any distribution with nonvanishing density on $\mathbb{R}^d$ satisfies (A1)-(A4). The appealing properties of halfspace depth are well known and well documented: see Donoho and Gasko [12], Mosler [39], Koshevoy [29], Ghosh and Chaudhuri [17], Cuestas-Albertos and Nieto-Reyes [8], Hassairi and Regaieg [26], to cite only a few. Halfspace depth takes values in $[0, 1/2]$, and its contours are continuous and convex; the corresponding regions are closed, convex, and nested as d decreases. Under very mild conditions, halfspace depth moreover fully characterizes the distribution P. For somewhat less satisfactory features, however, see Dutta et al. [13]. An important feature of halfspace depth is the convexity of its contours, which implies that halfspace depth contours cannot pick non convex features in the geometry of the underlying distribution, as illustrated in Figure 1.

We shall propose below a new depth concept, the Monge-Kantorovich (MK) depth, that relinquishes the affine equivariance and star convexity of contours imposed by Axioms (A1) and (A3) and recovers non convex features of the underlying distribution. As a preview of the concept, without going through any definitions, we illustrate in Figure 2 (using the same example as in Figure 1) the ability of the MK depth to capture non-convexities. In what follows, we characterize these abilities more formally. We shall emphasize that this notion comes in a package with new, interesting notions of vector ranks and quantiles, based on optimal transport, which reduce to classical notions in univariate and multivariate elliptical cases.

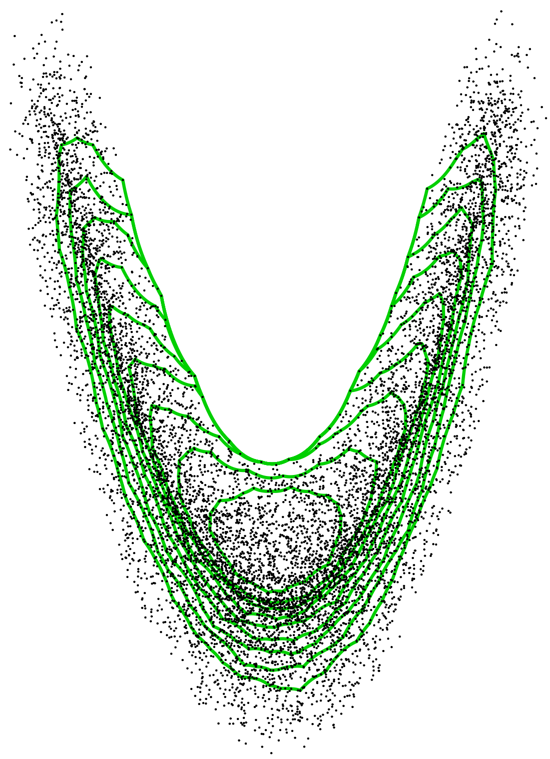

FIGURE 2. The Monge-Kantorovich depth contours for the same banana-shaped distribution from a sample of 9999 observations, as in Figure 1. The banana-like geometry of the data cloud is correctly picked up by the non convex contours.

2.3. **Monge-Kantorovich depth.** The principle behind the notion of depth we define here is to map the depth regions and contours relative to a well chosen reference distribution F, into depth contours and regions relative to a distribution of interest P on $\mathbb{R}^d$, using a well chosen mapping. The mapping proposed here is the *optimal transport plan* from F to P for quadratic cost.

Definition. Let P and F be two distributions on $\mathbb{R}^d$ with finite variance. An optimal transport plan from F to P for quadratic cost is a map $\mathrm{Q} : \mathbb{R}^d \longrightarrow \mathbb{R}^d$

that maximizes

$$\int u^\top \mathrm{Q}(u)\, dF(u) \text{ subject to } \mathrm{Q}\#F = P. \tag{2.1}$$

This definition has a classical counterpart in case of univariate distributions.

Proposition. *When $d = 1$ and F is uniform on $[0,1]$, $u \mapsto \mathrm{Q}(u)$ is the classical quantile function for distribution P.*

In order to base our notion of depth and quantiles for a distribution P on the optimal transport map from F to P, we need to ensure existence and uniqueness of the latter. We also need to extend this notion to define depth relative to distributions without finite second order moments. The following theorem, due to Brenier [4] and McCann [38] achieves both.

Theorem 2.1. *Let P and F be two distributions on $\mathbb{R}^d$. If F is absolutely continuous with respect to Lebesgue measure, the following hold.*

(1) There exists a convex function $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ such that $\nabla\psi\#F = P$. The function $\nabla\psi$ exists and is unique, F-almost everywhere.
(2) In addition, if P and F have finite second moments, $\nabla\psi$ is the unique optimal transport map from F to P for quadratic cost.

By the Kantorovich Duality Theorem (see Villani [51], Theorem 1.3), the function ψ, called *transportation potential* (hereafter simply *potential*), also solves the dual optimization problem

$$\int \psi dF + \int \psi^* dP = \inf_{\varphi} \int \varphi dF + \int \varphi^* dP, \tag{2.2}$$

where the infimum is over lower-semi-continuous convex functions φ. The pair (ψ, ψ^*) will be called *conjugate pair of potentials.*

On the basis of Theorem 2.1, we can define multivariate notions of quantiles and ranks, through which a depth function will be inherited from the reference distribution F.

Definition 2.1 (Monge-Kantorovich depth, quantiles, ranks and signs)**.** Let F be an absolutely continuous reference distribution on $\mathbb{R}^d$. Vector quantiles, ranks, signs and depth are defined as follows.

(1) The Monge-Kantorovich (hereafter MK) vector quantile function relative to distribution P is defined for each $u \in \mathbb{R}^d$ as the F-almost surely unique gradient of a convex function $\mathrm{Q}_P(u) := \nabla\psi(u)$ such that $\nabla\psi\#F = P$.

(2) The MK vector rank of $x \in \mathbb{R}^d$ is $\mathrm{R}_P(x) := \nabla\psi^*(x)$, where ψ^* is the conjugate of ψ. The MK rank is $\|\mathrm{R}_P(x)\|$ and the MK sign is $\mathrm{R}_P(x)/\|\mathrm{R}_P(x)\|$.
(3) The MK depth of $x \in \mathbb{R}^d$ relative to P is consequently defined as the halfspace depth of $\mathrm{R}_P(x)$ relative to the reference distribution F.
$$\mathrm{D}_P^{\mathrm{MK}}(x) := \mathrm{D}_F^{\mathrm{Tukey}}(\mathrm{R}_P(x)).$$

The notion of depth proposed in Definition 2.1 is based on an optimal transport map from the baseline distribution F to the distribution of interest. Each reference distribution will therefore generate, through the optimal transport map, a depth weak order on $\mathbb{R}^d$, relative to a distribution of interest P. This order is defined for each $(x_1, x_2) \in \mathbb{R}^{2d}$ by

$$x_1 \geq_{\mathrm{D}_{P;F}} x_2 \text{ if and only if } \mathrm{D}_F^{\mathrm{Tukey}}(\mathrm{R}_{P;F}(x_1)) \geq \mathrm{D}_F^{\mathrm{Tukey}}(\mathrm{R}_{P;F}(x_2)),$$

where the dependence of the rank function $\mathrm{R}_{P;F}$ and hence $D_{P;F}$ on the reference distribution is emphasized here, although, as in Definition 2.1, it will be omitted in the notation when there is no ambiguity. When requiring regularity of vector quantiles and ranks and of depth contours, we shall work within the following environment for the conjugate pair of potentials (ψ, ψ^*).

(C) Let $\mathcal{U}$ and $\mathcal{Y}$ be closed, convex subsets of $\mathbb{R}^d$, and $\mathcal{U}_0 \subset \mathcal{U}$ and $\mathcal{Y}_0 \subset \mathcal{Y}$ are some open, non-empty sets in $\mathbb{R}^d$. Let $\psi : \mathcal{U} \mapsto \mathbb{R}$ and $\psi^* : \mathcal{Y} \mapsto \mathbb{R}$ be a conjugate pair over $(\mathcal{U}, \mathcal{Y})$ that possess gradients $\nabla\psi(u)$ for all $u \in \mathcal{U}_0$, and $\nabla\psi^*(y)$ for all $y \in \mathcal{Y}_0$. The gradients $\nabla\psi|_{\mathcal{U}_0} : \mathcal{U}_0 \mapsto \mathcal{Y}_0$ and $\nabla\psi^*|_{\mathcal{Y}_0} : \mathcal{Y}_0 \mapsto \mathcal{U}_0$ are homeomorphisms and $\nabla\psi|_{\mathcal{U}_0} = (\nabla\psi^*|_{\mathcal{Y}_0})^{-1}$.

Sufficient conditions for condition (C) in the context of Definition 2.1 are provided by Caffarelli's regularity theory (Villani [51], Theorem 4.14). One set of sufficient conditions is as follows.

Proposition (Caffarelli). *Suppose that P and F admit densities, which are of smoothness class C^β for $\beta > 0$ on convex, compact support sets $\mathrm{cl}(\mathcal{Y}_0)$ and $\mathrm{cl}(\mathcal{U}_0)$, and the densities are bounded away from zero and above uniformly on the support sets. Then Condition (C) is satisfied for the conjugate pair (ψ, ψ^*) such that $\nabla\psi\#F = P$ and $\nabla\psi^*\#P = F$.*

Under sufficient conditions for (C) to be satisfied for MK vector quantiles Q_P and vector ranks R_P relative to distribution P, Q_P and R_P are continuous and inverse of each other, so that the MK depth contours are continuous, MK depth regions are nested and regions and contours take the following respective forms:

$$\mathbb{C}_P^{MK}(d) := \mathrm{Q}_P\left(\mathbb{C}_F^{\mathrm{Tukey}}(d)\right) \text{ and } \mathcal{C}_P^{MK}(d) := \mathrm{Q}_P\left(\mathcal{C}_F^{\mathrm{Tukey}}(d)\right), \text{ for } d \in (0, 1/2].$$

2.4. **Monge-Kantorovich depth with spherical uniform reference distribution.** Consider now Monge-Kantorovich depth defined from a baseline distribution with spherical uniform symmetry. We define the spherical uniform distribution supported on the unit ball $\mathbb{S}^d$ of $\mathbb{R}^d$ as follows.

Definition (Spherical uniform distribution)**.** The spherical uniform distribution U_d is the distribution of a random vector $r\varphi$, where r is uniform on $[0,1]$, φ is uniform on the unit sphere $\mathcal{S}^{d-1}$, and r and φ are mutually independent.

The spherical symmetry of distribution U_d produces halfspace depth contours that are concentric spheres, the deepest point being the origin. The radius τ of the ball $\mathbb{S}(\tau) = \{x \in \mathbb{R}^d : \|x\| \leq \tau\}$ is also its U_d-probability contents, that is, $\tau = U_d(\mathbb{S}(\tau))$. Letting $\theta := \arccos\tau$, the halfspace depth with respect to U_d of a point $\tau u \in \mathcal{S}(\tau) := \{x \in \mathbb{R}^d : \|x\| = \tau\}$, where $\tau \in (0,1]$ and $u \in \mathbb{S}^d$, is

$$\mathrm{D}_U(\tau u) = \begin{cases} \pi^{-1}[\theta - \cos\theta \log|\sec\theta + \tan\theta|] & d \geq 2 \\ (1-\tau)/2 & d = 1. \end{cases} \tag{2.3}$$

Note that for $d = 1$, u takes values ± 1 and that, in agreement with rotational symmetry of U_d, depth does not depend on u.

The principle behind the notion of depth we investigate further here is to map the depth regions and contours relative to the spherical uniform distribution U_d, namely, the concentric spheres, into depth contours and regions relative to a distribution of interest P on $\mathbb{R}^d$ using the chosen transport plan from U_d to P. Under sufficient conditions for (C) to be satisfied for MK vector quantiles Q_P and ranks R_P relative to distribution P (note that the conditions on F are automatically satisfied in case $F = U_d$), Q_P and R_P are continuous and inverse of each other, so that the MK depth contours are continuous, MK depth regions are nested and regions and contours take the following respective forms, when indexed by probability content.

$$\mathbb{C}_P^{MK}(\tau) := \mathrm{Q}_P\left(\mathbb{S}(\tau)\right) \text{ and } \mathcal{C}_P^{MK}(\tau) := \mathrm{Q}_P\left(\mathcal{S}(\tau)\right), \text{ for } \tau \in (0,1].$$

By construction, depth and depth contours coincide with Tukey depth and depth contours for the baseline distribution U_d. We now show that MK depth of Definition 2.1 still coincides with Tukey depth in case of univariate distributions as well as in case of elliptical distributions.

MK depth is halfspace depth in dimension 1. The halfspace depth of a point $x \in \mathbb{R}$ relative to a distribution P over $\mathbb{R}$ takes the very simple form

$$D_P^{\text{Tukey}}(x) = \min(P(x), 1 - P(x)),$$

where, by abuse of notation, P stands for both distribution and distribution function. The non decreasing map defined for each $x \in \mathbb{R}$ by $x \mapsto \mathrm{R}_P(x) = 2P(x) - 1$ is the derivative of a convex function and it transports distribution P to U_1, which is uniform on $[-1, 1]$, i.e., $\mathrm{R}_P \# P = U_1$. Hence R_P coincides with the MK vector rank of Definition 2.1. Therefore, for each $x \in \mathbb{R}$,

$$D_P(x) = D^{\text{Tukey}}_{U_d}(\mathrm{R}_P(x)) = \min(P(x), 1 - P(x))$$

and MK depth coincides with Tukey depth in case of all distributions with nonvanishing densities on the real line.

MK depth is halfspace depth for elliptical distributions. A d-dimensional random vector X has elliptical distribution $P_{\mu,\Sigma,f}$ with location $\mu \in \mathbb{R}^d$, positive definite symmetric $d \times d$ scatter matrix Σ and radial distribution function f if and only if

$$\mathrm{R}(X) := \frac{\Sigma^{-1/2}(X-\mu)}{\|\Sigma^{-1/2}(X-\mu)\|_2} F\Big(\|\Sigma^{-1/2}(X-\mu)\|_2\Big) \sim U_d, \tag{2.4}$$

where F, with density f, is the distribution function of $\|\Sigma^{-1/2}(X-\mu)\|_2$. The halfspace depth contours of $P_{\mu,\Sigma;F}$ coincide with its ellipsoidal density contours, hence only depend on μ and Σ. Their indexation, however, depends on F. The location parameter μ, with depth $1/2$, is the deepest point. In Proposition 2.1, we show that the mapping R is the rank function associated to $P_{\mu,\Sigma;f}$ according to our Definition 2.1.

Proposition 2.1. *The mapping defined for each $x \in \mathbb{R}^d$ by (2.4) is the gradient of a convex function ψ^* such that $\nabla\psi^* \# P_{\mu,\Sigma;f} = U_d$.*

The mapping R is therefore the MK vector rank function associated with $P_{\mu,\Sigma;f}$, and MK depth relative to the elliptical distribution $P_{\mu,\Sigma;f}$ is equal to halfspace depth. MK ranks, quantiles and depth therefore share invariance and equivariance properties of halfspace depth within the class of elliptical families, see [19], [20], [21] and [22].

3. Empirical depth, ranks and quantiles

Having defined Monge-Kantorovich vector quantiles, ranks and depth relative to a distribution P based on reference distribution F on $\mathbb{R}^d$, we now turn to the estimation of these quantities. Hereafter, we shall work within the environment defined by (C). We define $\Phi_0(\mathcal{U}, \mathcal{Y})$ as a collection of conjugate potentials (φ, φ^*) on $(\mathcal{U}, \mathcal{Y})$ such that $\varphi(u_0) = 0$ for some fixed point $u_0 \in \mathcal{U}_0$. Then, the MK vector quantiles and ranks of Definition 2.1 are

$$\mathrm{Q}_P(u) := \nabla\psi(u), \quad \mathrm{R}_P(y) := \nabla\psi^*(y) = (\nabla\psi)^{-1}(y), \tag{3.5}$$

for each $u \in \mathcal{U}_0$ and $y \in \mathcal{Y}_0$, respectively, where the potentials $(\psi, \psi^*) \in \Phi_0(\mathcal{U}, \mathcal{Y})$ are such that:

$$\int \psi dF + \int \psi^* dP = \inf_{(\varphi,\varphi^*)\in\Phi_0(\mathcal{U},\mathcal{Y})} \int \varphi dF + \int \varphi^* dP. \tag{3.6}$$

Constraining the conjugate pair to lie in Φ_0 is a normalization that pins down the constant, so that (ψ, ψ^*) are uniquely determined. We propose empirical versions of MK quantiles and ranks based on estimators of P, and possibly F, if necessary for computational reasons.

3.1. **Data generating processes.** Suppose that $\{\hat{P}_n\}$ and $\{\hat{F}_n\}$ are sequences of random measures on $\mathcal{Y}$ and $\mathcal{U}$, with finite total mass, that are consistent for P and F:

$$d_{\mathrm{BL}}(\hat{P}_n, P) \to_{\mathbb{P}^*} 0, \quad d_{\mathrm{BL}}(\hat{F}_n, F) \to_{\mathbb{P}^*} 0, \tag{3.7}$$

where $\to_{\mathbb{P}^*}$ denotes convergence in (outer) probability under probability measure $\mathbb{P}$, see van der Vaart and Wellner [49]. A basic example is where $\hat{P}_n$ is the empirical distribution of the random sample $(Y_i)_{i=1}^n$ drawn from P and $\hat{F}_n$ is the empirical distribution of the random sample $(U_i)_{i=1}^n$ drawn from F. Other, much more complicated examples, including smoothed empirical measures and data coming from dependent processes, satisfy sufficient conditions for (3.7) that we now give. In order to develop some examples, we introduce the ergodicity condition:

(E) Let $\mathcal{W}$ be a measurable subset of $\mathbb{R}^d$. A data stream $\{(W_{t,n})_{t=1}^n\}_{n=1}^\infty$, with $W_{t,n} \in \mathcal{W} \subset \mathbb{R}^d$ for each t and n, is ergodic for the probability law P_W on $\mathcal{W}$ if for each $g : \mathcal{W} \mapsto \mathbb{R}$ such that $\|g\|_{\mathrm{BL}(\mathcal{W})} < \infty$,

$$\frac{1}{n}\sum_{t=1}^n g(W_{t,n}) \to_{\mathbb{P}} \int g(w) dP_W(w). \tag{3.8}$$

The class of ergodic processes is extremely rich, including in particular the following.

- (E.1) $W_{t,n} = W_t$, where $(W_t)_{t=1}^\infty$ are independent, identically distributed random vectors with distribution P_W;
- (E.2) $W_{t,n} = W_t$, where $(W_t)_{t=1}^\infty$ is stationary strongly mixing process with marginal distribution P_W;
- (E.3) $W_{t,n} = W_t$, where $(W_t)_{t=1}^\infty$ is a non-stationary irreducible and aperiodic Markov chain with stationary distribution P_W;
- (E.4) $W_{t,n} = w_{t,n}$, where $(w_{t,n})_{t=1}^n$ is a deterministic allocations of points such that (3.8) holds deterministically.

Thus, if we observe the data sequence $\{(W_{t,n})_{t=1}^n\}_{n=1}^\infty$ that is ergodic for P_W, we can estimate P_W by the empirical and smoothed empirical measures

$$\hat{P}_W(A) = \frac{1}{n}\sum_{t=1}^n 1\{W_{t,n} \in A\}, \qquad \tilde{P}_W(A) = \frac{1}{n}\sum_{t=1}^n \int 1\{W_{t,n} + h_n\varepsilon \in A \cap \mathcal{W}\} d\Phi(\varepsilon),$$

where Φ is the probability law of the standard d-dimensional Gaussian vector, $N(0, I_d)$, and $h_n \geq 0$ is a semi-positive-definite matrix of bandwidths such that $\|h_n\| \to 0$ as $n \to \infty$. Note that $\tilde{P}_W$ may not integrate to 1, since we are forcing it to have support in $\mathcal{W}$.

Lemma 3.1. *Suppose that P_W is absolutely continuous with support contained in the compact set $\mathcal{W} \subset \mathbb{R}^d$. If $\{(W_{t,n})_{t=1}^n\}_{n=1}^\infty$ is ergodic for P_W on $\mathcal{W}$, then*

$$d_{\mathrm{BL}}(\hat{P}_W, P_W) \to_{\mathbf{P}^*} 0, \quad d_{\mathrm{BL}}(\tilde{P}_W, P_W) \to_{\mathbf{P}^*} 0.$$

Thus, if $P_Y := P$ and $P_U := F$ are absolutely continuous with support sets contained in compact sets $\mathcal{Y}$ and $\mathcal{U}$, and if $\{(Y_{t,n})_{t=1}^n\}_{n=1}^\infty$ is ergodic for P_Y on $\mathcal{Y}$ and $\{(U_{t,n})_{t=1}^n\}_{n=1}^\infty$ is ergodic for P_U on $\mathcal{U}$, then $\hat{P}_n = \hat{P}_W$ or $\tilde{P}_W$ and $\hat{F}_n = \hat{P}_U$ or $\tilde{P}_U$ obey condition (3.7).

Comment 3.1. Absolute continuity of P_W in Lemma 3.1 is only used to show that the smoothed estimator $\tilde{P}_W$ is asymptotically non-defective.

3.2. **Empirical quantiles, ranks and depth.** We base empirical versions of MK quantiles, ranks and depth on estimators $\hat{P}_n$ for P and $\hat{F}_n$ for F satisfying (3.7). We define empirical versions in the general case, before discussing their construction in some special cases for $\hat{P}_n$ and $\hat{F}_n$ below. Recall Assumption (C) is maintained throughout this section.

Definition 3.1 (Empirical quantiles and ranks)**.** Empirical vector quantile $\hat{\mathrm{Q}}_n$ and vector rank $\hat{\mathrm{R}}_n$ are any pair of functions satisfying, for each $u \in \mathcal{U}$ and $y \in \mathcal{Y}$,

$$\hat{\mathrm{Q}}_n(u) \in \arg\sup_{y\in\mathcal{Y}} y^\top u - \hat{\psi}_n^*(y), \quad \hat{\mathrm{R}}_n(y) \in \arg\sup_{u\in\mathcal{U}} y^\top u - \hat{\psi}_n(u), \tag{3.9}$$

where $(\hat{\psi}_n, \hat{\psi}_n^*) \in \Phi_0(\mathcal{U}, \mathcal{Y})$ is such that

$$\int \hat{\psi}_n d\hat{F}_n + \int \hat{\psi}_n^* d\hat{P}_n = \inf_{(\varphi,\varphi^*)\in\Phi_0(\mathcal{U},\mathcal{Y})} \int \varphi d\hat{F}_n + \int \varphi^* d\hat{P}_n. \tag{3.10}$$

Depth, depth contours and depth regions relative to P are then estimated with empirical versions inherited from $\hat{\mathrm{R}}_n$. For any $x \in \mathbb{R}^d$, the depth of x relative to P is estimated with

$$\hat{\mathrm{D}}_n(x) = \mathrm{D}_F^{\mathrm{Tukey}}(\hat{\mathrm{R}}_n(x)),$$

and for any $d \in (0, 1/2]$, the d-depth region relative to P and the corresponding contour are estimated with the following:

$$\hat{\mathbb{C}}_n(d) := \{x \in \mathbb{R}^d : \hat{\mathrm{D}}_n(x) \geq d\} \text{ and } \hat{\mathcal{C}}_n(d) := \{x \in \mathbb{R}^d : \hat{\mathrm{D}}_n(x) = d\}. \tag{3.11}$$

A more direct approach to estimating the regions and contours may be computationally more appealing. Even though $\hat{\mathrm{Q}}_n(\mathbb{C}_F^{\mathrm{Tukey}}(d))$ and $\hat{\mathrm{Q}}_n(\mathcal{C}_F^{\mathrm{Tukey}}(d))$ may now be finite sets of points, in case $\hat{P}_n$ is discrete, they are shown to converge to the population d-depth region and corresponding contour and can therefore be used for the construction of empirical counterparts. In case of discrete $\hat{P}_n$, the latter can be constructed from a *polyhedron supported by* $\hat{\mathrm{Q}}_n(\mathbb{C}_F^{\mathrm{Tukey}}(d))$ or $\hat{\mathrm{Q}}_n(\mathcal{C}_F^{\mathrm{Tukey}}(d))$. For precise definitions, existence and uniqueness of such polyhedra, see [10] for $d = 2$ and [18] for $d = 3$.

3.2.1. *Smooth* $\hat{P}_n$ *and* $\hat{F}_n$. Suppose $\hat{P}_n$ and $\hat{F}_n$ satisfy Caffarelli regularity conditions, so that $\hat{\mathrm{Q}}_n = \nabla\hat{\psi}_n$ and $\hat{\mathrm{R}}_n = \nabla\hat{\psi}_n^*$, with $(\hat{\psi}_n, \hat{\psi}_n^*)$ satisfying (C). Empirical versions are then defined identically to their theoretical counterparts. Depth, depth contours and depth regions relative to P are then estimated with empirical versions inherited from $\hat{\mathrm{Q}}_n$ and $\hat{\mathrm{R}}_n$. In particular, for any $x \in \mathbb{R}^d$, the depth of x relative to P is estimated with

$$\hat{\mathrm{D}}_n(x) = \mathrm{D}_F^{\mathrm{Tukey}}(\hat{\mathrm{R}}_n(x)).$$

Since $\hat{\mathrm{R}}_n = \hat{\mathrm{Q}}_n^{-1}$, as for the theoretical counterparts, for any $\tau \in (0, 1]$, the estimated depth region relative to P with probability content τ and the corresponding contour can be computed as

$$\hat{\mathbb{C}}_n(\tau) := \hat{\mathrm{Q}}_n\left(\mathbb{C}_F^{\mathrm{Tukey}}(d)\right) \text{ and } \hat{\mathcal{C}}_n(\tau) := \hat{\mathrm{Q}}_n\left(\mathcal{C}_F^{\mathrm{Tukey}}(d)\right).$$

Empirical depth regions are nested, and empirical depth contours are continuous, as are their theoretical counterparts. The estimators $\hat{\mathrm{Q}}_n$ and $\hat{\mathrm{R}}_n$ can be computed with the algorithm of Benamou and Brenier [3][1]. In the case where the reference distribution is the spherical uniform distribution, i.e., $F = U_d$, the estimated depth region relative to P with probability content τ and the corresponding contour can be computed as

$$\hat{\mathbb{C}}_n(\tau) := \hat{\mathrm{Q}}_n\left(\mathbb{S}(\tau)\right) \text{ and } \hat{\mathcal{C}}_n(\tau) := \hat{\mathrm{Q}}_n\left(\mathcal{S}(\tau)\right),$$

where $\mathbb{S}(\tau)$ and $\mathcal{S}(\tau)$ are the ball and the sphere of radius τ, respectively.

[1]A guide to implementation is given at http://www.numerical-tours.com/matlab/optimaltransp_2_benamou_brenier/).

3.2.2. *Discrete $\hat{P}_n$ and smooth $\hat{F}_n$.* Suppose now $\hat{P}_n$ is a discrete estimator of P and $\hat{F}_n$ is an absolutely continuous distribution with convex compact support $\mathbb{B} \subseteq \mathbb{R}^d$. Let $\hat{P}_n = \sum_{k=1}^{K_n} p_{k,n}\delta_{y_{k,n}}$, for some integer K_n, some non negative weights $p_{1,n}, \ldots, p_{K_n,n}$ such that $\sum_{k=1}^{K_n} p_{k,n} = 1$, and $y_{1,n}, \ldots, y_{K_n,n} \in \mathbb{R}^d$. The leading example is when $\hat{P}_n$ is the empirical distribution of a random sample $(Y_i)_{i=1}^n$ drawn from P.

The empirical quantile $\hat{\mathrm{Q}}_n$ is then equal to the $\hat{F}_n$-almost surely unique gradient of a convex map $\nabla\hat{\psi}_n$ such that $\nabla\hat{\psi}_n \# \hat{F}_n = \hat{P}_n$, i.e., the $\hat{F}_n$-almost surely unique map $\hat{\mathrm{Q}}_n = \nabla\hat{\psi}_n$ satisfying the following:

(1) $\nabla\hat{\psi}_n(u) \in \{y_{1,n}, \ldots, y_{K_n,n}\}$, for Lebesgue-almost all $u \in \mathbb{B}$,
(2) $\hat{F}_n\Big(\{u \in \mathbb{B} : \ \nabla\hat{\psi}_n(u) = y_{k,n}\}\Big) = p_{k,n}$, for each $k \in \{1, \ldots, K_n\}$,
(3) $\hat{\psi}_n$ is a convex function.

The following characterization of $\hat{\psi}_n$ specializes Kantorovich duality to this discrete-continuous case (for a direct proof, see for instance [15]).

Lemma. *There exist unique (up to an additive constant) weights $\{v_1^*, \ldots, v_n^*\}$ such that*

$$\hat{\psi}_n(u) = \max_{1\leq k\leq K_n} \{u^\top y_{k,n} - v_k^*\}$$

satisfies (1), (2) and (3). The function

$$v^* \mapsto \int \hat{\psi}_n d\hat{F}_n + \sum_{k=1}^{K_n} p_{k,n} v_k^*$$

is convex and minimized at $v^* = \{v_1^*, \ldots, v_n^*\}$.

The lemma allows efficient computation of $\hat{\mathrm{Q}}_n$ using a gradient algorithm proposed in [2]. $\hat{\psi}_n$ is piecewise affine and $\hat{\mathrm{Q}}_n$ is piecewise constant. The correspondence $\hat{\mathrm{Q}}_n^{-1}$ defined for each $k \leq K_n$ by

$$y_{k,n} \mapsto \hat{\mathrm{Q}}_n^{-1}(y_{k,n}) := \{u \in \mathbb{B} : \ \nabla\hat{\psi}_n(u) = y_{k,n}\}$$

maps $\{y_{1,n}, \ldots, y_{K_n,n}\}$ into K_n regions of a partition of $\mathbb{B}$, called a *power diagram.*

The estimator $\hat{\mathrm{R}}_n$ of the rank function can be any measurable selection from the correspondence $\hat{\mathrm{Q}}_n^{-1}$. Empirical depth is then $\hat{\mathrm{D}}_n(x) = \mathrm{D}_F^{\mathrm{Tukey}}(\hat{\mathrm{R}}_n(x))$, and depth regions and contours can be computed using the depth function, according to their definition as in (3.11), or from a polyhedron supported by $\hat{\mathrm{Q}}_n(\mathbb{C}_F^{\mathrm{Tukey}}(d))$ or $\hat{\mathrm{Q}}_n(\mathcal{C}_F^{\mathrm{Tukey}}(d))$ as before.

3.2.3. *Discrete $\hat{P}_n$ and $\hat{F}_n$.* Particularly amenable to computation is the case, where both distribution estimators $\hat{P}_n$ and $\hat{F}_n$ are discrete with uniformly distributed mass on sets of points of the same cardinality. Let $\hat{P}_n = \sum_{j=1}^n \delta_{y_j}/n$ for a set $\mathcal{Y}_n = \{y_1, \dots, y_n\}$ of points in $\mathbb{R}^d$ and $\hat{F}_n = \sum_{j=1}^n \delta_{u_j}/n$, for a set $\mathcal{U}_n = \{u_1, \dots, u_n\}$ of points in $\mathbb{R}^d$. The restriction of the quantile map $\hat{\mathrm{Q}}_n$ to $\mathcal{U}_n$ is the bijection

$$\begin{array}{rccc} \hat{\mathrm{Q}}_n|_{\mathcal{U}_n}: & \mathcal{U}_n & \longrightarrow & \mathcal{Y}_n \\ & u & \longmapsto & y = \hat{\mathrm{Q}}_n|_{\mathcal{U}_n}(u) \end{array}$$

that minimizes

$$\sum_{j=1}^n u_j^\top \hat{\mathrm{Q}}_n|_{\mathcal{U}_n}(u_j),$$

and $\hat{\mathrm{R}}_n|_{\mathcal{Y}_n}$ is its inverse. The solutions $\hat{\mathrm{Q}}_n$ and $\hat{\mathrm{R}}_n$ can be computed with any assignment algorithm. More generally, in the case of any two discrete estimators $\hat{P}_n$ and $\hat{F}_n$, the problem of finding $\hat{\mathrm{Q}}_n$ or $\hat{\mathrm{R}}_n$ is a linear programming problem.

In the case of the spherical uniform reference distribution $F = U_d$, empirical depth contours $\hat{\mathcal{C}}_n(\tau)$ and regions $\hat{\mathbb{C}}_n(\tau)$ can be computed from a polyhedron supported by $\hat{\mathrm{Q}}_n(\mathcal{U}_n(\tau))$, where $\mathcal{U}_n(\tau) = \{u \in \mathcal{U}_n : \ \|u\| \leq \tau\}$, $\tau \in (0,1]$. Estimated depth contours are illustrated in Figure 2 for the same banana-shaped distribution as in Figure 1. The specific construction to produce Figure 2 is the following: $\hat{P}_n$ is the empirical distribution of a random sample $\mathcal{Y}_n$ drawn from the banana distribution in $\mathbb{R}^2$, with $n = 9999$; $\hat{F}_n$ is the discrete distribution with mass $1/n$ on each of the points in $\mathcal{U}_n$. The latter is a collection of 99 evenly spaced points on each of 101 circles, of evenly spaced radii in $(0,1]$. The sets $\mathcal{Y}_n$ and $\mathcal{U}_n$ are matched optimally with the assignment algorithm of the *adagio* package in R. Empirical depth contours $\hat{\mathcal{C}}_n(\tau)$ are α-hulls of $\hat{\mathrm{Q}}_n(\mathcal{U}_n(\tau))$ for 11 values of $\tau \in (0,1)$ (see [14] for a definition of α-hulls). The α-hulls are computed using the *alphahull* package in R, with $\alpha = 0.3$. The banana-shaped distribution considered is the distribution of the vector $(X + R\cos\Phi, X^2 + R\sin\Phi)$, where X is uniform on $[-1,1]$, Φ is uniform on $[0, 2\pi]$, Z is uniform on $[0,1]$, X, Z and Φ are independent, and $R = 0.2Z(1 + (1 - |X|)/2)$.

3.3. **Convergence of empirical quantiles, ranks and depth contours.** Empirical quantiles, ranks and depth contours are now shown to converge uniformly to their theoretical counterparts.

Theorem 3.1 (Uniform Convergence of Empirical Transport Maps)**.** *Suppose that the sets $\mathcal{U}$ and $\mathcal{Y}$ are compact subsets of $\mathbb{R}^d$, and that probability measures P and F are absolutely continuous with respect to the Lebesgue measure with* support$(P) \subset \mathcal{Y}$ *and* support$(F) \subset \mathcal{U}$*. Suppose that $\{\hat{P}_n\}$ and $\{\hat{F}_n\}$ are sequences of random* *measures on $\mathcal{Y}$ and $\mathcal{U}$, with finite total mass, that are consistent for P and F in the sense of (3.7). Suppose that condition (C) holds for the solution of (3.6) for $\mathcal{Y}_0 := \text{int}(\text{support}(P))$ and $\mathcal{U}_0 := \text{int}(\text{support}(F))$. Then, as $n \to \infty$, for any compact set $K \subset \mathcal{U}_0$ and any compact set $K' \subset \mathcal{Y}_0$,*

$$\sup_{u \in K} \|\hat{\mathrm{Q}}_n(u) - \mathrm{Q}_P(u)\| \to_{\mathbf{P}^*} 0, \quad \sup_{y \in K'} \|\hat{\mathrm{R}}_n(y) - \mathrm{R}_P(y)\| \to_{\mathbf{P}^*} 0,$$

$$d_H(\hat{\mathrm{Q}}_n(K), \mathrm{Q}_P(K)) \to_{\mathbf{P}^*} 0, \quad d_H(\hat{\mathrm{R}}_n(K'), \mathrm{R}_P(K')) \to_{\mathbf{P}^*} 0.$$

The first result establishes the uniform consistency of empirical vector quantile and rank maps, hence also of empirical ranks and signs. The set $\mathrm{Q}(K)$ such that $\mathbb{P}_{U_d}(U \in K) = \tau$ is the statistical depth contour with probability content τ. The second result, therefore, establishes consistency of the approximation $\hat{\mathrm{Q}}_n(K)$ to the theoretical depth contour $\mathrm{Q}(K)$. The proof is given in the appendix.

Uniform convergence of empirical Monge-Kantorovitch quantiles $\tau \mapsto \hat{\mathrm{Q}}_n(\mathcal{S}(\tau))$, ranks $\hat{r}_n := \|\hat{\mathrm{R}}_n\|$ and signs $\hat{u}_n := \hat{\mathrm{R}}_n / \|\hat{\mathrm{R}}_n\|$ to their theoretical counterparts r_P and u_P follows by an application of the Continuous Mapping Theorem.

Corollary 3.1. *Under the assumptions of Theorem 3.1, as $n \to \infty$, for any compact set $K \subset \mathcal{U}_0$ and any compact set $K' \subset \mathcal{Y}_0$,*

$$\sup_{y \in K} \|\hat{r}_n(y) - r_P(y)\| \to_{\mathbf{P}^*} 0, \quad \sup_{y \in K'} \|\hat{u}_n(y) - u_P(y)\| \to_{\mathbf{P}^*} 0,$$

$$\sup_{\tau \in (0,1)} d_H(\hat{\mathrm{Q}}_n(\mathcal{S}(\tau)), \mathrm{Q}_P(\mathcal{S}(\tau))) \to_{\mathbf{P}^*} 0, \quad \sup_{\tau \in (0,1)} d_H(\hat{\mathrm{Q}}_n(\mathbb{S}(\tau)), \mathrm{Q}_P(\mathbb{S}(\tau))) \to_{\mathbf{P}^*} 0.$$

Appendix A. Uniform Convergence of Subdifferentials and Transport Maps

A.1. **Uniform Convergence of Subdifferentials.** Let $\mathcal{U}$ and $\mathcal{Y}$ be convex, closed subsets of $\mathbb{R}^d$. A pair of convex potentials $\psi : \mathcal{U} \mapsto \mathbb{R} \cup \{\infty\}$ and $\psi^* : \mathcal{Y} \mapsto \mathbb{R} \cup \{\infty\}$ is conjugate over $(\mathcal{U}, \mathcal{Y})$ if, for each $u \in \mathcal{U}$ and $y \in \mathcal{Y}$,

$$\psi(u) = \sup_{y \in \mathcal{Y}} y^\top u - \psi^*(y), \quad \psi^*(y) = \sup_{u \in \mathcal{U}} y^\top u - \psi(u).$$

Recall that we work within the following environment.

(C) Let $\mathcal{U}$ and $\mathcal{Y}$ be closed, convex subsets of $\mathbb{R}^d$, and $\mathcal{U}_0 \subset \mathcal{U}$ and $\mathcal{Y}_0 \subset \mathcal{Y}$ are some open, non-empty sets in $\mathbb{R}^d$. Let $\psi : \mathcal{U} \mapsto \mathbb{R}$ and $\psi^* : \mathcal{Y} \mapsto \mathbb{R}$ be a conjugate pair over $(\mathcal{U}, \mathcal{Y})$ that possess gradients $\nabla\psi(u)$ for all $u \in \mathcal{U}_0$, and $\nabla\psi^*(y)$ for all $y \in \mathcal{Y}_0$. The gradients $\nabla\psi|_{\mathcal{U}_0} : \mathcal{U}_0 \mapsto \mathcal{Y}_0$ and $\nabla\psi^*|_{\mathcal{Y}_0} : \mathcal{Y}_0 \mapsto \mathcal{U}_0$ are homeomorphisms and $\nabla\psi|_{\mathcal{U}_0} = (\nabla\psi^*|_{\mathcal{Y}_0})^{-1}$.

We also consider a sequence of conjugate potentials approaching (ψ, ψ^*).

(A) A sequence of conjugate potentials (ψ_n, ψ_n^*) over $(\mathcal{U}, \mathcal{Y})$, with $n \in \mathbb{N}$, is such that: $\psi_n(u) \to \psi(u)$ in $\mathbb{R} \cup \{\infty\}$ pointwise in u in a dense subset of $\mathcal{U}$ and $\psi_n^*(y) \to \psi^*(y)$ in $\mathbb{R} \cup \{\infty\}$ pointwise in y in a dense subset of $\mathcal{Y}$, as $n \to \infty$.

The condition (A) is equivalent to requiring that either ψ_n or ψ_n^* converge pointwise over dense subsets. There is no loss of generality in stating that both converge.

Define the maps

$$\mathrm{Q}(u) := \arg\sup_{y \in \mathcal{Y}} y^\top u - \psi^*(y), \quad \mathrm{R}(y) := \arg\sup_{u \in \mathcal{U}} y^\top u - \psi(u),$$

for each $u \in \mathcal{U}_0$ and $y \in \mathcal{Y}_0$. By the envelope theorem,

$$\mathrm{R}(y) = \nabla \psi^*(y), \text{ for } y \in \mathcal{Y}_0; \quad \mathrm{Q}(u) = \nabla \psi(u), \text{ for } u \in \mathcal{U}_0.$$

Hence, Q is the vector quantile function and R is its inverse, the vector rank function, from Definition 2.1.

Let us define, for each $u \in \mathcal{U}$ and $y \in \mathcal{Y}$,

$$\mathrm{Q}_n(u) \in \arg\sup_{y \in \mathcal{Y}} y^\top u - \psi_n^*(y), \quad \mathrm{R}_n(y) \in \arg\sup_{u \in \mathcal{U}} y^\top u - \psi_n(u). \tag{A.12}$$

It is useful to note that

$$\mathrm{R}_n(y) \in \partial \psi_n^*(y) \text{ for } y \in \mathcal{Y}; \quad \mathrm{Q}_n(u) \in \partial \psi_n(u) \text{ for } u \in \mathcal{U},$$

where ∂ denotes the sub-differential of a convex function; conversely, any pair of elements of $\partial \psi_n^*(y)$ and $\partial \psi_n(u)$, respectively, could be taken as solutions to the problem (A.12) (by Proposition 2.4 in Villani [51]). Hence, the problem of convergence of Q_n and R_n to Q and R is equivalent to the problem of convergence of subdifferentials. Moreover, by Rademacher's theorem, $\partial \psi_n^*(y) = \nabla \psi_n^*(y)$ and $\partial \psi_n(u) = \nabla \psi_n(u)$ almost everywhere with respect to the Lebesgue measure (see, e.g., [51]), so the solutions to (A.12) are unique almost everywhere on $u \in \mathcal{U}$ and $y \in \mathcal{Y}$.

Theorem A.1 (Local uniform convergence of subdifferentials)**.** *Suppose conditions (A) and (C) hold. Then, as $n \to \infty$, for any compact set $K \subset \mathcal{U}_0$ and any compact set $K' \subset \mathcal{Y}_0$,*

$$\sup_{u \in K} \|\mathrm{Q}_n(u) - \mathrm{Q}(u)\| \to 0, \quad \sup_{y \in K'} \|\mathrm{R}_n(y) - \mathrm{R}(y)\| \to 0.$$

Comment A.1. This result appears to be new. It complements the result stated in Lemma 5.4 in Villani [53] for the case $\mathcal{U}_0 = \mathcal{U} = \mathcal{Y}_0 = \mathcal{Y} = \mathbb{R}^d$. This result also trivially implies convergence in L^p norms, $1 \leq p < \infty$:

$$\int_{\mathcal{U}} \|\mathrm{Q}_n(u) - \mathrm{Q}(u)\|^p dF(u) \to 0, \quad \int_{\mathcal{Y}_0} \|\mathrm{R}_n(y) - \mathrm{R}(y)\|^p dP(y) \to 0,$$

for probability laws F on $\mathcal{U}$ and P on $\mathcal{Y}$, whenever for some $\bar{p} > p$

$$\sup_{n\in\mathbb{N}} \int_{\mathcal{U}} \|\mathrm{Q}_n(u)\|^{\bar{p}} + \|\mathrm{Q}(u)\|^p dF(u) < \infty, \quad \sup_{n\in\mathbb{N}} \int_{\mathcal{Y}_0} \|\mathrm{R}_n(y)\|^{\bar{p}} + \|\mathrm{R}(y)\|^p dP(y) < \infty.$$

Hence, the new result is stronger than available results on convergence in measure (including L^p convergence results) in the optimal transport literature (see, e.g., Villani [51, 52]). □

Comment A.2. The following example also shows that, in general, our result can not be strengthened to the uniform convergence over entire sets $\mathcal{U}$ and $\mathcal{Y}$. Consider the sequence of potential maps $\psi_n : \mathcal{U} = [0,1] \mapsto \mathbb{R}$:

$$\psi_n(u) = \int_0^u \mathrm{Q}_n(t)dt, \quad \mathrm{Q}_n(t) = t \cdot 1(t \leq 1 - 1/n) + 10 \cdot 1(t > 1 - 1/n).$$

Then $\psi_n(u) = 2^{-1}u^2 1(u \leq 1 - 1/n) + \left\{10(u - (1 - 1/n)) + 2^{-1}(1 - 1/n)^2\right\} 1(u > 1 - 1/n)$ converges uniformly on $[0,1]$ to $\varphi(u) = 2^{-1}u^2$. The latter potential has the gradient map $\mathrm{Q} : [0,1] \mapsto \mathcal{Y}_0 = [0,1]$ defined by $\mathrm{Q}(t) = t$. We have that $\sup_{t\in K} |\mathrm{Q}_n(t) - \mathrm{Q}(t)| \to 0$ for any compact subset K of $(0,1)$. However, the uniform convergence over the entire region $[0,1]$ fails, since $\sup_{t\in[0,1]} |\mathrm{Q}_n(t) - \mathrm{Q}(t)| \geq 9$ for all n. Therefore, the theorem can not be strengthened in general. □

We next consider the behavior of image sets of gradients defined as follows:

$$\mathrm{Q}_n(K) := \{\mathrm{Q}_n(u) : u \in K\}, \quad \mathrm{Q}(K) := \{\mathrm{Q}(u) : u \in K\},$$

$$\mathrm{R}_n(K') := \{\mathrm{R}_n(y) : y \in K'\}, \quad \mathrm{R}(K') := \{\mathrm{R}(y) : y \in K'\},$$

where $K \subset \mathcal{U}_0$ and $K' \subset \mathcal{Y}_0$ are compact sets. Also recall the definition of the Hausdorff distance between two non-empty sets A and B in $\mathbb{R}^d$:

$$d_H(A,B) := \sup_{b\in B} \inf_{a\in A} \|a - b\| \vee \sup_{a\in A} \inf_{b\in B} \|a - b\|.$$

Corollary A.1 (Convergence of sets of subdifferentials)**.** *Under the conditions of the previous theorem, we have that*

$$d_H(\mathrm{Q}_n(K), \mathrm{Q}(K)) \to 0, \quad d_H(\mathrm{R}_n(K'), \mathrm{R}(K')) \to 0.$$

A.2. **Uniform Convergence of Transport Maps.** We next consider the problem of convergence for potentials and transport (vector quantile and rank) maps arising from the Kantorovich dual optimal transport problem.

We equip $\mathcal{Y}$ with an absolutely continuous probability measure P and let

$$\mathcal{Y}_0 := \mathrm{int}(\mathrm{support}(P)).$$

We equip $\mathcal{U}$ with an absolutely continuous probability measure F and let

$$\mathcal{U}_0 := \mathrm{int}(\mathrm{support}(F)).$$

We consider a sequence of measures P_n and F_n that approximate P and F:

(W) There are sequences of measures $\{P_n\}_{n\in\mathbb{N}}$ on $\mathcal{Y}$ and $\{F_n\}_{n\in\mathbb{N}}$ on $\mathcal{U}$, with finite total mass, that converge to P and F, respectively, in the topology of weak convergence:

$$d_{\mathrm{BL}}(P_n, P) \to 0, \quad d_{\mathrm{BL}}(F_n, F) \to 0.$$

Recall that we defined $\Phi_0(\mathcal{U}, \mathcal{Y})$ as a collection of conjugate potentials (φ, φ^*) on $(\mathcal{U}, \mathcal{Y})$ such that $\varphi(u_0) = 0$ for some fixed point $u_0 \in \mathcal{U}_0$. Let $(\psi_n, \psi_n^*) \in \Phi_0(\mathcal{U}, \mathcal{Y})$ solve the Kantorovich problem for the pair (P_n, F_n)

$$\int \psi_n dF_n + \int \psi_n^* dP_n = \inf_{(\varphi,\varphi^*)\in\Phi_0(\mathcal{U},\mathcal{Y})} \int \varphi dF_n + \int \varphi^* dP_n. \tag{A.13}$$

Also, let $(\psi, \psi^*) \in \Phi_0(\mathcal{U}, \mathcal{Y})$ solve the Kantorovich problem for the pair (P, F):

$$\int \psi dF + \int \psi^* dP = \inf_{(\varphi,\varphi^*)\in\Phi_0(\mathcal{U},\mathcal{Y})} \int \varphi dF + \int \varphi^* dP. \tag{A.14}$$

It is known that solutions to these problems exist; see, e.g., Villani [51]. Recall also that we imposed the normalization condition in the definition of $\Phi_0(\mathcal{U}, \mathcal{Y})$ to pin down the constants.

Theorem A.2 (Local uniform convergence of transport maps)**.** *Suppose that the sets $\mathcal{U}$ and $\mathcal{Y}$ are compact subsets of $\mathbb{R}^d$, and that probability measures P and F are absolutely continuous with respect to the Lebesgue measure with* $\mathrm{support}(P) \subset \mathcal{Y}$ *and* $\mathrm{support}(F) \subset \mathcal{U}$*. Suppose that Condition (W) holds, and that Condition (C) holds for a solution (ψ, ψ^*) of (A.14) for the sets $\mathcal{U}_0$ and $\mathcal{Y}_0$ defined as above. Then conclusions of Theorem A.1 and Corollary A.1 hold.*

Appendix B. Proofs

B.1. **Proof of Proposition 2.1.** Denote by Ψ a primitive of F. It is easily checked that $x \mapsto R(x)$ is the gradient of $\Psi(x) := \Sigma^{1/2}\Psi(\|\Sigma^{-1/2}(x-\mu)\|)$. In order to show that Ψ is convex, it is sufficient to check that its Hessian, that is, the Jacobian of R, is positive definite. The Jacobian of R is

$$\frac{F(\|\Sigma^{-1/2}(x-\mu)\|_2)}{\|\Sigma^{-1/2}(x-\mu)\|_2}\Sigma^{-1/2} \quad - \quad \frac{F(\|\Sigma^{-1/2}(x-\mu)\|_2)}{2\|\Sigma^{-1/2}(x-\mu)\|_2^3}\Sigma^{-1/2}(x-\mu)(x-\mu)^\top\Sigma^{-1}$$
$$+ \quad \frac{f(\|\Sigma^{-1/2}(x-\mu)\|_2)}{2\|\Sigma^{-1/2}(x-\mu)\|_2^2}\Sigma^{-1/2}(x-\mu)(x-\mu)^\top\Sigma^{-1},$$

which is positive semidefinite if $I-\frac{1}{2}UU^\top$ is, where $U := \Sigma^{-1/2}(x-\mu)/\|\Sigma^{-1/2}(x-\mu)\|_2$. Denoting by $U, U_2, \ldots, U_d$ an orthonormal basis of $\mathbb{R}^d$, we obtain

$$I-\frac{1}{2}UU^\top = \frac{1}{2}UU^\top + U_2U_2^\top + \ldots + U_dU_d^\top,$$

which is clearly positive definite.

B.2. **Proof of Theorem A.1.** The proof relies on the equivalence of the uniform and continuous convergence.

Lemma B.1 (Uniform convergence via continuous convergence)**.** *Let $\mathbb{D}$ and $\mathbb{E}$ be complete separable metric spaces, with $\mathbb{D}$ compact. Suppose $f : \mathbb{D} \mapsto \mathbb{E}$ is continuous. Then a sequence of functions $f_n : \mathbb{D} \mapsto \mathbb{E}$ converges to f uniformly on $\mathbb{D}$ if and only if, for any convergent sequence $x_n \to x$ in $\mathbb{D}$, we have that $f_n(x_n) \to f(x)$.*

For the proof, see, e.g., Rockafellar and Wets [44]. The proof also relies on the following convergence result, which is a consequence of Theorem 7.17 in Rockafellar and Wets [44]. For a point a and a non-empty set A in $\mathbb{R}^d$, define $d(a, A) := \inf_{a' \in A} \|a - a'\|$.

Lemma B.2 (Argmin convergence for convex problems)**.** *Suppose that g is a lower-semi-continuous convex function mapping $\mathbb{R}^d$ to $\mathbb{R} \cup \{+\infty\}$ that attains a minimum on the set $\mathcal{X}_0 = \arg\inf_{x\in\mathbb{R}^d} g(x) \subset \mathcal{D}_0$, where $\mathcal{D}_0 = \{x \in \mathbb{R}^d : g(x) < \infty\}$ is a non-empty, open set in $\mathbb{R}^d$. Let $\{g_n\}$ be a sequence of convex, lower-semi-continuous functions mapping $\mathbb{R}^d$ to $\mathbb{R} \cup \{+\infty\}$ and such that $g_n(x) \to g(x)$ pointwise in $x \in \mathbb{R}^d_0$, where $\mathbb{R}^d_0$ is a countable dense subset of $\mathbb{R}^d$. Then any $x_n \in \arg\inf_{x\in\mathbb{R}^d} g_n(x)$ obeys*

$$d(x_n, \mathcal{X}_0) \to 0,$$

and, in particular, if $\mathcal{X}_0$ is a singleton $\{x_0\}$, $x_n \to x_0$.

The proof of this lemma is given below, immediately after the conclusion of the proof of this theorem.

We define the extension maps $y \mapsto g_{n,u}(y)$ and $u \mapsto \bar{g}_{n,y}(u)$ mapping $\mathbb{R}^d$ to $\mathbb{R} \cup \{-\infty\}$

$$g_{n,u}(y) = \begin{cases} y^\top u - \psi_n^*(y) & \text{if } y \in \mathcal{Y} \\ -\infty & \text{if } y \notin \mathcal{Y} \end{cases}, \quad \bar{g}_{n,y}(u) = \begin{cases} y^\top u - \psi_n(u) & \text{if } u \in \mathcal{U} \\ -\infty & \text{if } u \notin \mathcal{U}. \end{cases}$$

By the convexity of ψ_n and ψ_n^* over convex, closed sets $\mathcal{Y}$ and $\mathcal{U}$, we have that the functions are proper upper-semi-continuous concave functions. Define the extension maps $y \mapsto g_u(y)$ and $u \mapsto \bar{g}_y(u)$ mapping $\mathbb{R}^d$ to $\mathbb{R} \cup \{-\infty\}$ analogously, by removing the index n above.

Condition (A) assumes pointwise convergence of ψ_n^* to ψ^* on a dense subset of $\mathcal{Y}$. By Theorem 7.17 in Rockafellar and Wets [44], this implies the uniform convergence of ψ_n^* to ψ^* on any compact set $K \subset \text{ int } \mathcal{Y}$ that does not overlap with the boundary of the set $\mathcal{D}_1 = \{y \in \mathcal{Y} : \psi^*(y) < +\infty\}$. Hence, for any sequence $\{u_n\}$ such that $u_n \to u \in K$, a compact subset of $\mathcal{U}_0$, and any $y \in (\text{int } \mathcal{Y}) \setminus \partial\mathcal{D}_1$,

$$g_{n,u_n}(y) = y^\top u_n - \psi_n^*(y) \to g_u(y) = y^\top u - \psi^*(y).$$

Next, consider any $y \notin \mathcal{Y}$, in which case, $g_{n,u_n}(y) = -\infty \to g_u(y) = -\infty$. Hence,

$$g_{n,u_n}(y) \to g_u(y) \text{ in } \mathbb{R} \cup \{-\infty\}, \text{ for all } y \in \mathbb{R}_1^d = \mathbb{R}^d \setminus (\partial\mathcal{Y} \cup \partial\mathcal{D}_1),$$

where $\mathbb{R}_1^d$ is a dense subset of $\mathbb{R}^d$. We apply Lemma B.2 to conclude that

$$\arg\sup_{y \in \mathbb{R}^d} g_{n,u_n}(y) \ni \mathrm{Q}_n(u_n) \to \mathrm{Q}(u) = \arg\sup_{y \in \mathbb{R}^d} g_u(y) = \nabla\psi(u).$$

Take K as any compact subset of $\mathcal{U}_0$. The above argument applies for every point $u \in K$ and every convergent sequence $u_n \to u$. Therefore, since by Assumption (C) $u \mapsto Q(u) = \nabla\psi(u)$ is continuous in $u \in K$, we conclude by the equivalence of the continuous and uniform convergence, Lemma B.1, that

$$\mathrm{Q}_n(u) \to \mathrm{Q}(u) \text{ uniformly in } u \in K.$$

By symmetry, the proof of the second claim is identical to the proof of the first claim. □

B.3. **Proof of Lemma B.2.** By assumption, $\mathcal{X}_0 = \arg\min g \in \mathcal{D}_0$, and $\mathcal{X}_0$ is convex and closed. Let x_0 be an element of $\mathcal{X}_0$. We have that, for all $0 < \varepsilon \leq \varepsilon_0$ with ε_0 such that $B_{\varepsilon_0}(\mathcal{X}) \subset \mathcal{D}_0$,

$$g(x_0) < \inf_{x \in \partial B_\varepsilon(\mathcal{X}_0)} g(x), \tag{B.15}$$

where $B_\varepsilon(\mathcal{X}_0) := \{x \in \mathbb{R}^d : d(x, \mathcal{X}_0) \leq \varepsilon\}$ is convex and closed.

Fix an $\varepsilon \in (0, \varepsilon_0]$. By convexity of g and g_n and by Theorem 7.17 in Rockafellar and Wets [44], the pointwise convergence of g_n to g on a dense subset of $\mathbb{R}^d$ is equivalent to the uniform convergence of g_n to g on any compact set K that does not overlap with $\partial\mathcal{D}_0$, i.e. $K \cap \partial\mathcal{D}_0 = \emptyset$. Hence, $g_n \to g$ uniformly on $B_{\varepsilon_0}(\mathcal{X}_0)$. This and (B.15) imply that eventually, i.e. for all $n \geq n_\varepsilon$,

$$g_n(x_0) < \inf_{x \in \partial B_\varepsilon(\mathcal{X}_0)} g_n(x).$$

By convexity of g_n, this implies that $g_n(x_0) < \inf_{x \notin B_\varepsilon(\mathcal{X}_0)} g_n(x)$ for all $n \geq n_\varepsilon$, which is to say that, for all $n \geq n_\varepsilon$,

$$\arg\inf g_n = \arg\min g_n \subset B_\varepsilon(\mathcal{X}_0).$$

Since ε can be set as small as desired, it follows that any $x_n \in \arg\inf g_n$ obeys $d(x_n, \mathcal{X}_0) \to 0$. □

B.4. **Proof of Corollary A.1.** By Lemma A.1 and the definition of the Hausdorff distance,

$$\begin{aligned} d_H(\mathrm{Q}_n(K), \mathrm{Q}(K)) \quad & \leq \sup_{u \in K} \inf_{u' \in K} \|\mathrm{Q}_n(u) - \mathrm{Q}(u')\| \vee \sup_{u' \in K} \inf_{u \in K} \|\mathrm{Q}_n(u') - \mathrm{Q}(u)\| \\ & \leq \sup_{u \in K} \|\mathrm{Q}_n(u) - \mathrm{Q}(u)\| \vee \sup_{u' \in \mathcal{U}} \|\mathrm{Q}_n(u') - \mathrm{Q}(u')\| \\ & \leq \sup_{u \in K} \|\mathrm{Q}_n(u) - \mathrm{Q}(u)\| \to 0. \end{aligned}$$

The proof of the second claim is identical. □.

B.5. **Proof of Theorem A.2.** Step 1. Here we show that the set of conjugate pairs is compact in the topology of the uniform convergence. First we notice that, for any $(\varphi, \varphi^*) \in \Phi_0(\mathcal{U}, \mathcal{Y})$,

$$\|\varphi\|_{\mathrm{BL}(\mathcal{U})} \;\leq\; (\|\mathcal{Y}\|\|\mathcal{U}\|) \vee \|\mathcal{Y}\| < \infty, \quad \|\varphi^*\|_{\mathrm{BL}(\mathcal{Y})} \leq (2\|\mathcal{Y}\|\|\mathcal{U}\|) \vee \|\mathcal{U}\| < \infty,$$

where $\|A\| := \sup_{a \in A} \|a\|$ for $A \subset \mathbb{R}^d$ and where we have used the fact that $\varphi(u_0) = 0$ for some $u_0 \in \mathcal{U}$ as well as compactness of $\mathcal{Y}$ and $\mathcal{U}$.

The Arzela-Ascoli theorem implies that $\Phi_0(\mathcal{U}, \mathcal{Y})$ is relatively compact in the topology of the uniform convergence. We want to show compactness, namely that this set is also closed. For this we need to show that all uniformly convergent subsequences $(\varphi_n, \varphi_n^*)_{n \in \mathbb{N}'}$ (where $\mathbb{N}' \subset \mathbb{N}$) have the limit point

$$(\varphi, \varphi^*) = \lim_{n \in \mathbb{N}'} (\varphi_n, \varphi_n^*) \in \Phi_0(\mathcal{U}, \mathcal{Y}).$$

This is true, since uniform limits of convex functions are necessarily convex ([44]) and since

$$\begin{aligned}
\varphi(u) &= \lim_{n\in\mathbb{N}'} \left[\sup_{y\in\mathcal{Y}} u^\top y - \varphi_n^*(y)\right] \\
&\leq \limsup_{n\in\mathbb{N}'} \left[\sup_{y\in\mathcal{Y}} (u^\top y - \varphi^*(y)) + \sup_{y\in\mathcal{Y}} |\varphi_n^*(y) - \varphi^*(y)|\right] = \sup_{y\in\mathcal{Y}} u^\top y - \varphi^*(y); \\
\varphi(u) &= \lim_{n\in\mathbb{N}'} \left[\sup_{y\in\mathcal{Y}} u^\top y - \varphi_n^*(y)\right] \\
&\geq \liminf_{n\in\mathbb{N}'} \left[\sup_{y\in\mathcal{Y}} (u^\top y - \varphi^*(y)) - \sup_{y\in\mathcal{Y}} |\varphi_n^*(y) - \varphi^*(y)|\right] = \sup_{y\in\mathcal{Y}} u^\top y - \varphi^*(y);
\end{aligned}$$

Analogously, $\varphi^*(y) = \sup_{u\in\mathcal{U}} u^\top y - \varphi(y)$.

Step 2. The claim here is that

$$I_n := \int \psi_n dF_n + \int \psi_n^* dP_n \to_{n\in\mathbb{N}} \int \psi dF + \int \psi^* dP =: I_0. \tag{B.16}$$

Indeed,

$$I_n \leq \int \psi dF_n + \int \psi^* dP_n \to_{n\in\mathbb{N}} I_0,$$

where the inequality holds by definition, and the convergence holds by

$$\left|\int \psi d(F_n - F)\right| + \left|\int \psi^* d(P_n - P)\right| \lesssim d_{\mathrm{BL}}(F_n, F) + d_{\mathrm{BL}}(P_n, P) \to 0.$$

Moreover, by definition

$$II_n := \int \psi_n dF + \int \psi_n^* dP \geq I_0,$$

but

$$|I_n - II_n| \leq \left|\int \psi_n d(F_n - F)\right| + \left|\int \psi_n^* d(P_n - P)\right| \lesssim d_{\mathrm{BL}}(F_n, F) + d_{\mathrm{BL}}(P_n, P) \to 0.$$

Step 3. Here we conclude.

First, we observe that the solution pair (ψ, ψ^*) to the limit Kantorovich problem is unique on $\mathcal{U}_0 \times \mathcal{Y}_0$ in the sense that any other solution (φ, φ^*) agrees with (ψ, ψ^*) on $\mathcal{U}_0 \times \mathcal{Y}_0$. Indeed, suppose that $\varphi(u_1) \neq \psi(u_1)$ for some $u_1 \in \mathcal{U}_0$. By the uniform continuity of elements of $\Phi_0(\mathcal{U}, \mathcal{Y})$ and openness of $\mathcal{U}_0$, there exists a ball $B_\varepsilon(u_1) \subset \mathcal{U}_0$ such that $\psi(u) \neq \varphi(u)$ for all $u \in B_\varepsilon(u_1)$. By the normalization assumption $\varphi(u_0) = \psi(u_0) = 0$, there does not exist a constant $c \neq 0$ such that $\psi(u) = \varphi(u) + c$ for all $u \in \mathcal{U}_0$, so this must mean that $\nabla\psi(u) \neq \nabla\varphi(u)$ on a set $K \subset \mathcal{U}_0$ of positive measure (otherwise, if they disagree only on a set of measure zero, we would have $\psi(u)-\psi(u_0) = \int_0^1 \nabla\psi(u_0+v^\top(u-u_0))^\top(u-u_0)dv = \int_0^1 \nabla\varphi(u_0 + v^\top(u - u_0))^\top(u - u_0)dv = \varphi(u) - \varphi(u_0)$ for almost all $u \in B_\varepsilon(u_1)$, which is a contradiction). However, the statement $\nabla\psi \neq \nabla\varphi$ on a set $K \subset \mathcal{U}_0$ of positive Lebesgue measure would contradict the fact that any solution ψ or φ of the Kantorovich problem must obey

$$\int h \circ \nabla\varphi dF = \int h \circ \nabla\psi dF = \int hdP,$$

for each bounded continuous h, i.e. that $\nabla\varphi\#F = \nabla\psi\#F = P$, established on p.72 in Villani [51]. Analogous argument applies to establish uniqueness of ψ^* on the set $\mathcal{Y}_0$.

Second, we can split $\mathbb{N}$ into subsequences $\mathbb{N} = \cup_{j=1}^\infty \mathbb{N}_j$ such that for each j:

$$(\psi_n, \psi_n^*) \to_{n\in\mathbb{N}_j} (\varphi_j, \varphi_j^*) \in \Phi_0(\mathcal{U}, \mathcal{Y}), \quad \text{uniformly on } \mathcal{U} \times \mathcal{Y}. \tag{B.17}$$

But by Step 2 this means that

$$\int \varphi_j dF + \int \varphi_j^* dP = \int \psi dF + \int \psi^* dP.$$

It must be that each pair (φ_j, φ_j^*) is the solution to the limit Kantorovich problem, and by the uniqueness established above we have that

$$(\varphi_j, \varphi_j^*) = (\psi, \psi^*) \text{ on } \mathcal{U}_0 \times \mathcal{Y}_0.$$

By Condition (C) we have that, for $u \in \mathcal{U}_0$ and $y \in \mathcal{Y}_0$:

$$\mathrm{Q}(u) = \nabla\psi(u) = \nabla\varphi_j(u), \quad \mathrm{R}(u) = \nabla\psi^*(u) = \nabla\varphi_j^*(u).$$

By (B.17) and Condition (C) we can invoke Theorem A.1 to conclude that $\mathrm{Q}_n \to \mathrm{Q}$ uniformly on compact subsets of $\mathcal{U}_0$ and $\mathrm{R}_n \to \mathrm{R}$ uniformly on compact subsets of $\mathcal{Y}_0$. $\square$

B.6. **Proof of Theorem 3.1.** The proof is an immediate consequence of the Extended Continuous Mapping Theorem, as given in van der Vaart and Wellner [49], Theorem A.1 and Corollary A.1.

The theorem, specialized to our context, reads: Let $\mathbb{D}$ and $\mathbb{E}$ be normed spaces and let $x \in \mathbb{D}$. Let $\mathbb{D}_n \subset \mathbb{D}$ be arbitrary subsets and $g_n : \mathbb{D}_n \mapsto \mathbb{E}$ be arbitrary maps ($n \geq 0$), such that for every sequence $x_n \in \mathbb{D}_n$ such that $x_n \to x$, along a subsequence, we have that $g_n(x_n) \to g_0(x)$, along the same subsequence. Then, for arbitrary (i.e. possibly non-measurable) maps $X_n : \Omega \mapsto \mathbb{D}_n$ such that $X_n \to_{\mathbf{P}^*} x$, we have that $g_n(X_n) \to_{\mathbf{P}^*} g_0(x)$.

In our case $X_n = (\hat{P}_n, \hat{F}_n)$ is a stochastic element of $\mathbb{D}$, viewed as an arbitrary map from Ω to $\mathbb{D}$, and $x = (P, F)$ is a non-stochastic element of $\mathbb{D}$, where $\mathbb{D}$ is the space of linear operators $\mathbb{D}$ acting on the space of bounded Lipschitz functions. This space can be equipped with the norm

$$\|\cdot\|_{\mathbb{D}} : \|(x_1, x_2)\|_{\mathbb{D}} = \|x_1\|_{\mathrm{BL}(\mathcal{Y})} \vee \|x_2\|_{\mathrm{BL}(\mathcal{U})}.$$

Moreover, $X_n \to_{\mathbf{P}^*} x$ with respect to this norm, i.e.

$$\|X_n - x\|_{\mathbb{D}} := \|\hat{P}_n - P\|_{\mathrm{BL}(\mathcal{Y})} \vee \|\hat{F}_n - F\|_{\mathrm{BL}(\mathcal{U})} \to_{\mathbf{P}^*} 0.$$

Then $g_n(X_n) := (\hat{\mathrm{Q}}_n, \hat{\mathrm{R}}_n)$ and $g(x) := (\mathrm{Q}, \mathrm{R})$ are viewed as elements of $\mathbb{E} = \ell^\infty(K \times K', \mathbb{R}^d \times \mathbb{R}^d)$, the space of bounded functions mapping $K \times K'$ to $\mathbb{R}^d \times \mathbb{R}^d$, equipped with the supremum norm. The maps have the continuity property: if $\|x_n - x\|_{\mathbb{D}} \to 0$ along a subsequence, then $\|g_n(x_n) - g(x)\|_{\mathbb{E}} \to 0$ along the same subsequence, as established by Theorem A.1 and Corollary A.1. Hence conclude that $g_n(X_n) \to_{\mathbf{P}^*} g(x)$. □

B.7. **Proof of Lemma 3.1.** Step 1. The set $\mathcal{G}_1 = \{g : \mathcal{W} \mapsto \mathbb{R} : \|g\|_{\mathrm{BL}(\mathcal{W})} \leq 1\}$ is compact in the topology of the uniform convergence by the Arzela-Ascoli theorem. Consider the sup norm $\|g\|_\infty = \sup_{w \in \mathcal{W}} |g(w)|$. By compactness, any cover of $\mathcal{G}_1$ by balls, with the diameter $\varepsilon > 0$ under the sup norm, has a finite subcover with the number of balls $N(\varepsilon)$. Let $(g_{j,\varepsilon})_{j=1}^{N(\varepsilon)}$ denote some points ("centers") in these balls. Thus, by the ergodicity condition (E) and $N(\varepsilon)$ being finite, we have that

$$\begin{aligned}
\sup_{g \in \mathcal{G}_1} \int g d(\hat{P}_W - P_W) &\leq \max_{j \in \{1,\ldots,N(\varepsilon)\}} \int g_{j,\varepsilon} d(\hat{P}_W - P_W) + \varepsilon \int |d\hat{P}_W| + |dP_W| \\
&= \max_{j \in \{1,\ldots,N(\varepsilon)\}} \int g_{j,\varepsilon} d(\hat{P}_W - P_W) + 2\varepsilon \to 2\varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, conclude $\sup_{g \in \mathcal{G}_1} \int g d(\hat{P}_W - P_W) \to_{\mathbf{P}^*} 0$.

Step 2. The same argument works for $\tilde{P}_W$ in place of P_W, since

$$\begin{aligned}
&\sup_{g \in \mathcal{G}_1} \int g d(\hat{P}_W - \tilde{P}_W) \\
&\leq \sup_{g \in \mathcal{G}_1} \int \int \{g(w) - g(w + \varepsilon h_n)\} 1(w + \varepsilon h_n \in \mathcal{W}) d\Phi(\varepsilon) d\hat{P}_W(w) \\
&+ \int \int 1\{w + h_n \varepsilon \in \mathcal{W}^c\} d\Phi(\varepsilon) d\hat{P}_W(w),
\end{aligned}$$

where both terms converge in probability to zero. The first term is bounded by

$$n^{-1} \sum_{t=1}^{n} \int \|\varepsilon h_n\| d\Phi(\varepsilon) \lesssim \|h_n\| \to 0.$$

As for the second term, we first approximate the indicator $x \mapsto 1(x \in \mathcal{W}^c)$ from above by a function $x \mapsto g_\delta(x) = (1 - d(x, \mathcal{W}^c)/\delta) \vee 0$, which is bounded above by 1 and obeys $\|g_\delta\|_{\mathrm{BL}(\mathbb{R}^d)} \leq 1 \vee \delta^{-1} < \infty$. Then the second term is bounded by $\int\int g_\delta\{w + h_n\varepsilon\} d\Phi(\varepsilon) d\hat{P}_W(w)$, which converges in probability to $\int g_\delta dP_W$ by Step 1. By absolute continuity of P_W and support$(P_W) \cap \mathcal{W}^c = \emptyset$, holding by assumption, and by the definition of g_δ, we can set $\int g_\delta dP_W$ arbitrarily small by setting δ arbitrarily small. □

Acknowledgments. We thank seminar participants at Oberwolfach 2012, LMPA, Paris 2013, the 2nd Conference of the International Society for Nonparametric Statistics, Cadiz 2014, the 3rd Institute of Mathematical Statistics Asia Pacific Rim Meeting, Taipei 2014, and the International Conference on Robust Statistics, Kolkata 2015, for useful discussions, and Denis Chetverikov and Yaroslav Mukhin for excellent comments. The authors also thank Mirek Šiman for sharing his code, including the data-generating process for the banana-shaped distribution.

References

[1] Agostinelli, C., and Romanazzi, M. (2011). Local depth, *Journal of Statistical Planning and Inference* **141**, 817-830.

[2] Aurenhammer, F., Hoffmann, F., and Aronov, B. (1998). Minkowski-type theorems and mean-square clustering, *Algorithmica* **20**, 61–76.

[3] Benamou, J.-D., Brenier, Y. (2000). A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik* **84**, 375–393.

[4] Brenier, Y (1991). "Polar factorization and monotone rearrangement of vector-valued functions, *Communications in Pure and Applied Mathematics* **44**, 375–417.

[5] Carlier, G., Chernozhukov, V., and Galichon, A. (2014). Vector quantile regression. ArXiv preprint arXiv:1406.4643.

[6] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data, *Journal of the American Statistical Association*, **91**, 862-872.

[7] Chen, Y., Dang, X., Peng, H., and Bart, H. L. J. (2009). Outlier detection with the kernelized spatial depth function, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31**, 288-305.

[8] Cuesta-Albertos, J., and Nieto-Reyes, A. (2008). The random Tukey depth, *Computational Statistics and Data Analysis* **52**, 4979-4988.

[9] Deurninge, A. (2014). Multivariate quantiles and multivariate L-moments, ArXiv preprint arXiv:1409.6013

[10] Deneen, L., and Shute, G. (1988). "Polygonization of point sets in the plane," *Discrete and Computational Geometry* **3**, 77-87.

[11] Donoho, D. L. (1982). Breakdown properties of multivariate location estimators, Qualifying Paper, Harvard University.

[12] Donoho, D. L., and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness, *The Annals of Statistics* **20**, 1803–1827.

[13] Dutta, S., Ghosh, A. K., and Chaudhuri, P. (2011). Some intriguing properties of Tukey's halfspace depth, *Bernoulli* **17**, 1420-1434.

[14] Edelsbrunner, H., Kirkpatrick, D., and Seidel, R. (1983). On the shape of a set of points in the plane, *IEEE Transactions on Information Theory* **29**, 551-559.

[15] Ekeland, I., Galichon, A., and Henry, M. (2012). Comonotonic measures of multivariate risks, *Mathematical Finance* **22**, 109–132.
[16] Galichon, A., and Henry, M. (2012). Dual theory of choice under multivariate risk, *Journal of Economic Theory* **147**, 1501–1516.
[17] Ghosh, A. K., and Chaudhuri, P. (2005). On maximum depth and related classifiers, *Scandinavian Journal of Statistics* **32**, 327-350.
[18] Gruenbaum, B. (1994). "Hamiltonian polygons and polyhedra," *Geombinatorics* **3**, 83–89.
[19] Hallin, M., and Paindaveine, D. (2002). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks, *The Annals of Statistics* **30**, 1103–1133.
[20] Hallin, M., and Paindaveine, D. (2004). Rank-based optimal tests of the adequacy of an elliptic VARMA model, *The Annals of Statistics* **32**, 2642-2678.
[21] Hallin, M., and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity, *The Annals of Statistics* **34**, 2707-2756.
[22] Hallin, M., and Paindaveine, D. (2008). Optimal rank-based tests for homogeneity of scatter, *The Annals of Statistics* **36**, 1261-1298.
[23] Hallin, M., and Werker, B. J. M. (2003). Semiparametric efficiency, distribution-freeness, and invariance, *Bernoulli* **9**, 137-165.
[24] Hallin, M., Paindaveine, D., and Šiman, M. (2010). Multivariate quantiles and multiple-output regression quantiles: from L^1 optimization to halfspace depth (with discussion), *The Annals of Statistics* **38**, 635-669.
[25] Hardy, G., Littlewood, J., and Pólya, G. (1952). *Inequalities.* Cambridge: Cambridge University Press.
[26] Hassairi, A., and Regaieg, O. (2008). On the Tukey depth of a continuous probability distribution, *Statistics and Probability Letters* **78**, 2308–2313.
[27] Hlubinka, D., Kotík, L., and Vencálek, O. (2010). Weighted halfspace depth, *Kybernetika* **46**, 125-148.
[28] Koenker, R., and Bassett, G., Jr. (1987). Regression quantiles, *Econometrica* **46**, 33–50.
[29] Koshevoy, G. (2002). The Tukey depth characterizes the atomic measure, *Journal of Multivariate Analysis* **83**, 360-364.
[30] Koshevoy, G., and Mosler, K. (1997). Zonoid trimming for multivariate distributions, *The Annals of Statistics* **25**, 1998-2017.
[31] Koltchinskii, V. (1997). "M-estimation, convexity and quantiles". *Annals of Statistics* **25**, 435–477.
[32] Koltchinskii, V., and Dudley, R. (1992) On spatial quantiles, unpublished manuscript.
[33] Liu, R. Y. (1990). On a notion of data depth based on random simplices, *The Annals of Statistics* **18**, 405-414.
[34] Liu, R. Y. (1992). Data depth and multivariate rank tests, in L^1*-Statistics and Related Methods* (Y. Dodge, ed.) 279–294. North-Holland, Amsterdam.
[35] Liu, R. Y., Parelius, J. M., and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference, (with discussion), *The Annals of Statistics* **27**, 783-858.
[36] Liu, R., and Singh, K. (1993) . A quality index based on data depth and multivariate rank tests, *Journal of the American Statistical Association* **88**, 257–260.
[37] Mahalanobis, P. C. (1936). On the generalized distance in statistics, *Proceedings of the National Academy of Sciences of India* **12**, 49–55.
[38] McCann, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps," *Duke Mathematical Journal* **80**, 309–324
[39] Mosler, K. (2002). *Multivariate dispersion, central regions and depth: the lift zonoid approach,* New York: Springer.

[40] Möttönen, J., and Oja, H. (1995). Multivariate sign and rank methods, *Journal of Nonparametric Statistics* **5**, 201–213.
[41] Oja, H. (1983). Descriptive statistics for multivariate distributions, *Statistics and Probability Letters* **1**, 327-332.
[42] Paindaveine, D., and Van Bever, G. (2013). From depth to local depth, *Journal of the American Statistical Association* **108**, 1105–1119.
[43] Paindaveine, D., and Šiman, M. (2012). Computing multiple-output regression quantile regions, *Computational Statistics and Data Analysis* **56**, 841–853.
[44] Rockafellar, R. T., and Wets, R. J.-B. (1998). *Variational analysis,* Berlin: Springer.
[45] Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications, *Statistica Neerlandica* **56**, 214–232.
[46] Singh, K. (1991). Majority depth, unpublished manuscript.
[47] Stahel, W. (1981). Robuste Schätzungen : infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen, PhD Thesis, University of Zürich.
[48] Tukey, J. W. (1975), Mathematics and the Picturing of Data, in *Proceedings of the International Congress of Mathematicians* (Vancouver, B. C., 1974), Vol. 2, Montreal. QC: Canadian Mathematical Congress, pp. 523-531.
[49] van der Vaart, A. W., and Wellner, J. A. (1996). *Weak Convergence*. Springer New York.
[50] Vardi, Y., and Zhang, C.-H. (2000). The multivariate L^1-median and associated data depth, *Proceedings of the National Academy of Sciences* **97**, 1423–1426.
[51] Villani, C. (2003). *Topics in Optimal Transportation*. Providence: American Mathematical Society.
[52] Villani, C. (2009). *Optimal transport: Old and New*. Grundlehren der Mathematischen Wissenschaften. Springer-Verlag: Heidelberg.
[53] Villani, C. (2008). Stability of a 4th-order curvature condition arising in optimal transport theory," *Journal of Functional Analysis* **255**, 2683–2708.
[54] Zuo, Y. (2003). Projection-based depth functions and associated medians, *The Annals of Statistics* **31**, 1460-1490.
[55] Zuo, Y., and Serfling, R. (2000). General Notions of Statistical Depth Function, *The Annals of Statistics* **28**, 461-482.

MIT

Sciences-Po, Paris

ECARES, Université libre de Bruxelles and ORFE, Princeton University

The Pennsylvania State University