



# Monge-Kantorovich Depth, Quantiles, Ranks, and Signs

Victor Chernozhukov, Alfred Galichon, Marc Hallin, Marc Henry

## ► To cite this version:

Victor Chernozhukov, Alfred Galichon, Marc Hallin, Marc Henry. Monge-Kantorovich Depth, Quantiles, Ranks, and Signs. *Annals of Statistics*, 2017, 45 (1), pp.223 - 256. 10.1214/16-AOS1450 . hal-03391975

**HAL Id: hal-03391975**

**<https://sciencespo.hal.science/hal-03391975>**

Submitted on 21 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## MONGE–KANTOROVICH DEPTH, QUANTILES, RANKS AND SIGNS

BY VICTOR CHERNOZHUKOV<sup>\*,1</sup>, ALFRED GALICHON<sup>†,2</sup>,  
 MARC HALLIN<sup>‡,3</sup> AND MARC HENRY<sup>§,4</sup>

*Massachusetts Institute of Technology*<sup>\*</sup>, *New York University*<sup>†</sup>,  
*Université libre de Bruxelles*<sup>‡</sup> and *Pennsylvania State University*<sup>§</sup>

We propose new concepts of statistical depth, multivariate quantiles, vector quantiles and ranks, 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 for  $d = 1$  and in the case of spherical 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, signs and vector quantiles and ranks, and show their consistency by establishing a uniform convergence property for empirical (forward and reverse) transport maps, which is the main theoretical result of this paper.

**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 [54]. Since then, many other concepts have been considered: simplicial depth [37], majority depth ([52] and [40]), projection depth ([38], building on [53] and [14, 61]), Mahalanobis depth [38, 40, 41], Oja depth [45], zonoid depth ([35] and [36]), spatial depth [6, 34, 44, 57],  $L^p$  depth [62], among many others (see, for instance, [39]). An axiomatic approach, aiming at unifying all those concepts, was initiated by Liu [37] and Zuo and Serfling [62], who list four properties that are generally considered desirable for any statistical depth function, namely

---

Received January 2015; revised February 2016.

<sup>1</sup>Supported by the NSF Grant SES 1061841.

<sup>2</sup>Supported by the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013), ERC Grant Agreement No. 313699.

<sup>3</sup>Supported by the IAP research network Grant P7/06 of the Belgian government (Belgian Science Policy) and a Crédit aux Chercheurs of the Fonds National de la Recherche Scientifique.

<sup>4</sup>Supported by the SSHRC Grant 435-2013-0292 and the NSERC Grant 356491-2013.

*MSC2010 subject classifications.* 62M15, 62G35.

*Key words and phrases.* Statistical depth, vector quantiles, vector ranks, multivariate signs, empirical transport maps, uniform convergence of empirical transport.

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 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 multi-modal 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 [1, 7, 31] and [47] 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 [51]: 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 univariate distributions with non-vanishing Lebesgue densities (either over  $\mathbb{R}$ , or over some interval of  $\mathbb{R}$ ). Here, the concepts of quantile and distribution functions, ranks and signs are the “classical” univariate ones. The second case is that of the family  $\mathcal{P}_{\text{ell}}^d$  of all full-rank elliptical distributions over  $\mathbb{R}^d$  ( $d > 1$ ) with radial densities over elliptical support sets. Recall that the family  $\mathcal{P}_{\text{ell};g}^d = \{P_{\mu,\Sigma,g}\}$  of elliptical distributions with given radial density  $g$  (a density over  $\mathbb{R}^+$ ) and radial distribution function  $G$  is a parametric family indexed by a location parameter  $\mu$  and a scatter parameter  $\Sigma$  (a symmetric positive definite real matrix) such that a random vector  $X$  has distribution  $P_{\mu,\Sigma,g}$  iff the *residual*  $Y := \Sigma^{-1/2}(X - \mu)$ , which results from transforming  $X$  into isotropic position, has spherical distribution  $P_{0,I,g}$ . Further, this is equivalent to  $R_P(Y) := (Y/\|Y\|)G(\|Y\|)$  having the spherical uniform distribution  $U_d$  on the unit ball  $\mathbb{S}^d$  in  $\mathbb{R}^d$ . By *spherical uniform*, we mean the distribution of a random vector  $r\varphi$ , where  $r$  is uniform on  $[0, 1]$ ,  $\varphi$  is uniform on the unit sphere  $\mathcal{S}^{d-1}$ , and  $r$  and  $\varphi$  are mutually independent. There, spherical contours with  $P_{\mu,I,g}$ -probability contents  $\tau$  coincide with the halfspace depth contours, and provide a natural definition of  $\tau$ -quantile contours for  $Y$ , while  $R_P(Y)$ ,  $R_P(Y)/\|R_P(Y)\|$  and  $\|R_P(Y)\|$  play the roles of vector ranks, signs and ranks, respectively [23–27]: we call them *spherical vector ranks*, *signs* and *ranks*. On the other hand, we call the inverse map  $u \mapsto Q_P(u)$  of the vector rank map  $y \mapsto R_P(y) = (y/\|y\|)G(\|y\|)$  the *vector quantile* map. In both cases, the relevance of ranks and signs, whether traditional or spherical, is related to their role as maximal invariants under groups of transformations minimally generating  $\mathcal{P}^1$  or the family  $\mathcal{P}_{\text{sph}}^d := \{P_{0,I,f}\}$  of spherical distributions, of which distribution-freeness of  $R_P$  is just a by-product, as

explained in [29]. 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  $\mathcal{P}^1$ , and spherical ones for  $\mathcal{P}_{\text{sph}}^d$ —hence should coincide with halfspace depth.

Now, a closer look at those two particular cases reveals that halfspace depth contours, in  $\mathcal{P}^1$  and  $\mathcal{P}_{\text{sph}}^d$ , are the images, by the vector quantile map  $Q_P$ , of the hyperspheres  $\mathcal{S}(\tau)$  with radii  $\tau \in [0, 1)$  centered at the origin. The map  $Q_P$  is the gradient of a convex function and it transports the spherical uniform distribution  $U_d$  on the unit ball  $\mathbb{S}^d$  of  $\mathbb{R}^d$  into the univariate distribution  $P \in \mathcal{P}^1$  or into the spherical distribution  $P = P_{0,I,f}$  of interest.

For the case of general distributions  $P$  over  $\mathbb{R}^d$ , we proceed similarly, and define the map  $Q_P$  as a gradient of a convex function that transform the spherical uniform distribution  $U_d$  into the target distribution, namely, such that if  $U \sim U_d$  then  $Y = Q_P(U) \sim P$ . It follows from McCann’s [42] extension of Brenier’s celebrated polar factorization theorem [4] that, for any distribution  $P$  on  $\mathbb{R}^d$ , such a gradient  $Q_P$  exists, and is essentially unique. Moreover, when  $P$  has finite moments of order two, that mapping  $Q_P$  is the Monge–Kantorovich optimal transport map that transfers the spherical uniform distribution  $U_d$  to  $P$ , where optimality is in the sense of minimizing the expected quadratic cost  $\mathbb{E}_U(Q(U) - U)^2$  subject to  $U \sim U_d$  and  $Q(U) \sim P$ .

This suggests a new concept of statistical depth, which we call the *Monge–Kantorovich* (or MK) *depth*  $D^{\text{MK}}$ , the contours of which are obtained as the images by  $Q_P$  of the hyperspheres with radius  $\tau \in [0, 1]$ . When restricted to  $\mathcal{P}^1$  or  $\mathcal{P}_{\text{sph}}^d$ , Monge–Kantorovich and halfspace depths coincide. Under suitable regularity conditions due to Caffarelli (see [58], Section 4.2.2),  $Q_P$  is a homeomorphism, and its inverse  $R_P := Q_P^{-1}$  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,

$$(1) \quad y_2 \succeq_{D^{\text{MK}}} y_1 \quad \text{if and only if} \quad \|R_P(y_2)\| \leq \|R_P(y_1)\|.$$

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

(a) an MK *vector quantile* map  $Q_P$ , and the associated MK *quantile* correspondence, which maps  $\tau \in [0, 1]$  to  $Q_P(\mathcal{S}(\tau))$ ,

(b) an MK *vector rank* (or MK *signed rank*) function  $R_P$ , which can be decomposed into an MK *rank* function  $r_P$  from  $\mathbb{R}^d$  to  $[0, 1]$ , with  $r_P(x) := \|R_P(x)\|$ , and an MK *sign* function  $u_P$ , mapping  $x \in \mathbb{R}^d$  to  $u_P(x) := R_P(x)/\|R_P(x)\| \in \mathcal{S}^{d-1}$ .

To the best of our knowledge, this is the first proposal of a depth concept based on the Monge–Kantorovich theory of measure transportation—hence, the first attempt to provide a measure-driven ordering of  $\mathbb{R}^d$  based on measure transportation theory. Previous proposals have been made, however, of measure transportation-based

vector quantile functions in Ekeland, Galichon and Henry [18] and Galichon and Henry [19] (with moment conditions) and Carlier, Chernozhukov and Galichon [5] (dropping moment conditions) who also extended the notion to vector quantile regression, creating a vector analogue of Koenker and Bassett's [33] scalar quantile regression. More recently, Decurninge [10] proposed a new concept of multivariate  $L^p$  moments based upon a similar notion. In these contributions, however, the focus is not statistical depth and the associated ranks and quantiles, 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 previously mentioned.

We then proceed to define the empirical notions corresponding to the concepts given above. We define the empirical MK vector quantiles and ranks as the essentially unique gradients  $\hat{Q}_n$  and  $\hat{R}_n$  of a pair of convex functions solving the Kantorovich dual problem for a Monge optimal transport with quadratic costs involving the empirical distribution of the observations. Using the plug-in principle, we then define the empirical rank and sign maps as  $\|\hat{R}_n\|$  and  $\hat{R}_n/\|\hat{R}_n\|$ , and the empirical  $\tau$ -quantile sets and contours as  $\hat{Q}_n(\mathbb{S}(\tau))$  and  $\hat{Q}_n(\mathcal{S}(\tau))$ . We establish the uniform convergence of these quantities to their theoretical counterparts. We derive these results as a consequence of the uniform convergence of empirical transport (vector quantile and rank) maps  $\hat{Q}_n$  and  $\hat{R}_n$  to their theoretical counterparts  $Q_P$  and  $R_P$  on compact subsets of the domain's interior. Although the definitions, in Section 2, address very general densities, consistency results are limited to compactly supported ones, due to the lack of more general convergence results in the measure transportation literature. This is the main theoretical result of the paper, presented in Theorem 3.1. This result is derived through an application of the extended continuous mapping theorem and a set of new theorems on stability of transports under deterministic perturbations of the source and target measures, given as Theorems A.1 and A.2 in the Appendix, which are new results of independent interest. Application of the extended continuous mapping theorem allows to us then to replace the deterministic perturbations by stochastic perturbations of measures and obtain the stochastic uniform convergence of the empirical transport maps.

*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, 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  $\tau$ . Let  $P_X$  stand for the distribution of the random vector  $X$ . The symbol  $\partial$  will denote either the boundary of a set or the subdifferential, as will be clear from the context. Following

Villani [58], we denote by  $g\#\mu$  the *image measure* (or *push-forward*) of a measure  $\mu \in \mathcal{P}$  by a measurable map  $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$ . Explicitly, for any Borel set  $A$ ,  $g\#\mu(A) := \mu(g^{-1}(A))$ . For a Borel subset  $\mathbb{D}$  of a vector space equipped with the norm  $\|\cdot\|$  and  $f : \mathbb{D} \rightarrow \mathbb{R}$ , let

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

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

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

which metrizes the topology of weak convergence. Throughout the paper, we let  $\mathcal{U} \subseteq \mathbb{R}^d$  and  $\mathcal{Y} \subseteq \mathbb{R}^d$  denote convex subsets of  $\mathbb{R}^d$  with non-empty interiors. A *convex* function on  $\mathcal{U}$  is a function  $\psi : \mathcal{U} \rightarrow \mathbb{R} \cup \{+\infty\}$  for which

$$\psi((1-t)x + tx') \leq (1-t)\psi(x) + t\psi(x')$$

for any  $t \in (0, 1)$  and any  $(x, x')$  such that  $\psi(x)$  and  $\psi(x')$  are finite. Such a function is continuous on the interior of the convex set  $\text{dom } \psi := \{x \in \mathcal{U} : \psi(x) < \infty\}$ , and differentiable Lebesgue-almost everywhere in  $\text{dom } \psi$ , by Rademacher's theorem. Write  $\nabla \psi$  for the gradient of  $\psi$ . For any function  $\psi : \mathcal{U} \mapsto \mathbb{R} \cup \{+\infty\}$ , the *conjugate*  $\psi^* : \mathcal{Y} \mapsto \mathbb{R} \cup \{+\infty\}$  of  $\psi$  is defined for each  $y \in \mathcal{Y}$  by

$$(2) \quad \psi^*(y) := \sup_{z \in \mathcal{U}} [y^\top z - \psi(z)].$$

The conjugate  $\psi^*$  of  $\psi$  is a convex lower-semi-continuous function on  $\mathcal{Y}$ . Call *conjugate pair of potentials* over  $(\mathcal{U}, \mathcal{Y})$  any pair of lower-semi-continuous convex functions  $(\psi, \psi^*)$  that are conjugates of each other. The transpose of a matrix  $A$  is denoted  $A^\top$ . We call *weak order* a complete reflexive and transitive binary relation. Finally, recall the definition of Hausdorff distance between two non-empty sets  $A$  and  $B$  in  $\mathbb{R}^d$ :

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

*Outline of the paper.* Section 2 introduces and motivates the concepts 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. Section 4 describes computational characterizations. The Appendix presents additional theoretical results and proofs.

## 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 (Depth and ordering).** A depth function is an upper semi-continuous mapping  $D : \mathbb{R}^d \mapsto \mathbb{R}$ . In our context these functions will be indexed by a distribution  $P$ . The quantity  $D_P(x)$  is called the *depth of  $x$  relative to  $P$* . For each  $P \in \mathcal{P}$ , the *depth ordering*  $\geq_{D_P}$  associated with  $D_P$  is the weak order on  $\mathbb{R}^d$  defined, for  $(y_1, y_2) \in \mathbb{R}^{2d}$ , by

$$y_1 \geq_{D_P} y_2 \quad \text{if and only if} \quad D_P(y_1) \geq D_P(y_2),$$

in which case  $y_1$  is said to be *deeper* than  $y_2$  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  $\mathbb{R}^d$ . The *region of depth  $d$*  is the upper contour set of level  $d$  of  $D_P$ , namely  $\mathbb{C}_P(d) := \{x \in \mathbb{R}^d : D_P(x) \geq d\}$ ; the *contour of depth  $d$*  is the boundary of that region:  $\mathcal{C}_P(d) := \partial \mathbb{C}_P(d)$ .

By construction, the depth regions are nested:

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

Hence, the depth ordering qualifies as a *center-outward ordering* of points in  $\mathbb{R}^d$  relative to the center given by the set of the deepest points,  $\arg \sup_{x \in \mathbb{R}^d} D_P(x)$ .

It is often convenient to work with depth regions indexed by their probability content.

**DEFINITION (Depth regions with probability content  $\tau$ ).** For  $\tau \in [0, 1]$ , the depth region with probability content at least  $\tau$  is

$$\mathbb{K}_P(\tau) := \mathbb{C}_P(d(\tau)), \quad d(\tau) := \inf\{d \in \mathbb{R} : P(\mathbb{C}(d)) \geq \tau\};$$

the corresponding contour region is the boundary  $\mathcal{K}_P(\tau) := \partial \mathbb{K}_P(d)$ .

**2.2. Liu–Zuo–Serfling axioms and Tukey’s halfspace depth.** The four axioms proposed by Liu [37] and Zuo and Serfling [62] to unify the diverse depth functions proposed in the literature are the following:

(A1) (Affine invariance)  $D_{P_{AX+b}}(Ax + b) = D_{P_X}(x)$  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,  $D_P(x_0) = \max_{x \in \mathbb{R}^d} D_P(x)$ .



(A3) (Linear monotonicity relative to the deepest points) If  $x_0$  is deepest, then  $D_P(x) \leq D_P((1 - \alpha)x_0 + \alpha x)$  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)  $\lim_{\|x\| \rightarrow \infty} D_P(x) = 0$ .

The earliest and most popular depth function is *halfspace depth* proposed by Tukey [54]:

DEFINITION (Tukey’s halfspace depth). The halfspace depth  $D_P^{\text{Tukey}}(x)$  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

$$D_{P_X}^{\text{Tukey}}(x) := \min_{\varphi \in \mathcal{S}^{d-1}} P_X[(X - x)^\top \varphi \geq 0].$$

Halfspace depth relative to any distribution with non-vanishing density on  $\mathbb{R}^d$  satisfies (A1)–(A4). The appealing properties of halfspace depth are well known and well documented: see Donoho and Gasko [15], Mosler [43], Koshevoy [36], Ghosh and Chaudhuri [20], Cuestas-Albertos and Nieto-Reyes [9], Hassairi and Regaieg [30], 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. [16]. 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.

Below, we propose 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 does account for the possible non-convex features of the underlying distribution. As a preview of the concept, without going through any definition, we illustrate in Figure 2 (using the same banana-shaped distribution 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 the univariate and multivariate spherical cases.

**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 an adequate mapping. The mapping proposed here is the gradient  $\nabla\psi$  of a convex function  $\psi$  such that if  $U$  has distribution  $F$ , then  $Y := \nabla\psi(U)$  has distribution  $P$ . The gradient  $\nabla\psi$  is said to *push  $F$  forward*



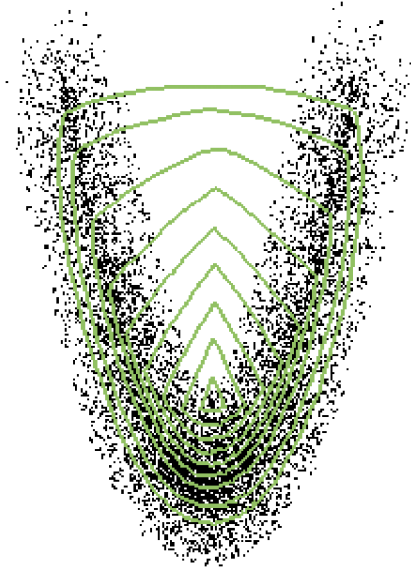


FIG. 1. Tukey halfspace depth contours for a banana-shaped distribution, produced with the algorithm of [28] and [46] from a sample of 9999 observations. The banana-like geometry of the data cloud is not picked by the convex contours, and the deepest point is close to the boundary of the support.

to  $P$ , which is conventionally denoted by the push-forward notation  $\nabla\psi\#F = P$  defined in the notation section.

The gradient of a convex function property is a generalization of the one-dimensional monotonicity property. When  $F$  and  $P$  have finite second-order moments, these maps are the optimal Monge–Kantorovich transport maps from  $F$  to  $P$  for the quadratic cost, as explained below. In the unidimensional case, when  $F$  is the standard uniform, the gradient/optimal transport map  $\nabla\psi$  coincides with the classical quantile function.

The following theorem, due to Brenier [4] and McCann [42], establishes the existence of gradients of convex functions with the required properties.

**THEOREM 2.1 (Brenier–McCann’s existence result).** *Let  $P$  and  $F$  be two distributions on  $\mathbb{R}^d$ . (1) If  $F$  is absolutely continuous with respect to the Lebesgue measure on  $\mathbb{R}^d$ , the following holds: for any convex set  $\mathcal{U} \subseteq \mathbb{R}^d$  containing the support of  $F$ , there exists a convex function  $\psi : \mathcal{U} \rightarrow \mathbb{R} \cup \{+\infty\}$  such that  $\nabla\psi\#F = P$ . The gradient  $\nabla\psi$  of that function exists and is unique,  $F$ -almost everywhere. (2) If, in addition,  $P$  is absolutely continuous on  $\mathbb{R}^d$ , the following holds: for any convex set  $\mathcal{Y} \subseteq \mathbb{R}^d$  containing the support of  $P$ , there exists a convex function  $\psi^* : \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$  such that  $\nabla\psi^*\#P = F$ . The gradient  $\nabla\psi^*$  of  $\psi^*$  exists, is unique, and  $\nabla\psi^* = \nabla\psi^{-1}$ ,  $P$ -almost everywhere.*

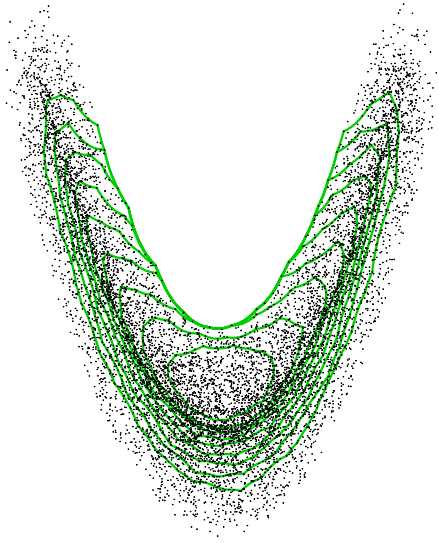


FIG. 2. The Monge–Kantorovich depth contours for the same banana-shaped distribution from a sample of 9999 observations, as in Figure 1. The banana-like geometry of the data cloud is correctly picked up by the non-convex contours.

REMARK 2.1 (Interpretation as a Monge–Brenier optimal transport). If  $P$  and  $F$  have finite second moments,  $\nabla\psi$  is an *optimal transport plan* from  $F$  to  $P$  for quadratic cost: more precisely, any transport coinciding, on  $\mathcal{U}$  (hence a.e.), with  $\nabla\psi$  solves the problem

$$\inf_Q \int (u - Q(u))^2 dF(u) : \quad Q\#F = P,$$

or, equivalently,

$$(3) \quad \sup_Q \int u^\top Q(u) dF(u) : \quad Q\#F = P.$$

This definition has a classical counterpart in the univariate case. When  $d = 1$  and  $F$  is uniform on  $[0, 1]$ , the optimal transport  $u \mapsto \nabla\psi(u)$  is the classical quantile function for distribution  $P$ .

We now state a fundamental duality result due to Kantorovich and Brenier, on which we explicitly rely in Section 3.

THEOREM 2.2 (Kantorovich–Brenier, see [58]). *Suppose hypothesis (1) of Theorem 2.1 holds and  $P$  and  $F$  have finite second moments: then the conjugate*

pair of potentials  $(\psi, \psi^*)$  solves the optimization problem

$$(4) \quad \int \psi dF + \int \psi^* dP = \inf_{(\varphi, \varphi^*)} \left( \int \varphi dF + \int \varphi^* dP \right),$$

where the infimum is taken over the class of conjugate pairs of potentials  $(\varphi, \varphi^*)$  over  $(\mathcal{U}, \mathcal{Y})$ .

Such pair  $(\psi, \psi^*)$  of potentials is called *optimal*.

REMARK 2.2. This problem is dual to the optimal transport problem (3). Moreover, under the hypotheses of Theorem 2.2,  $\nabla \psi$  is the unique optimal transport map from  $F$  to  $P$  for quadratic cost, in the sense that any other optimal transport coincides with  $\nabla \psi$  on a set of  $F$ -measure one (see [58]). Under the hypotheses of Theorem 2.2 and hypothesis (2) of Theorem 2.1,  $\nabla \psi^*$  is the unique optimal (reverse) transport map from  $P$  to  $F$  for quadratic cost, in the sense that any other optimal transport coincides with  $\nabla \psi^*$  on a set of  $P$ -measure one (see [58]).

Next, we use Theorem 2.1 to define a natural notion of *vector quantiles* and *vector ranks*.

DEFINITION 2.1 (Monge–Kantorovich vector quantiles and ranks). Let  $F$  be an absolutely continuous reference distribution, and let  $P$  be an arbitrary distribution over  $\mathbb{R}^d$ . Let  $\mathcal{U} \subseteq \mathbb{R}^d$  and  $\mathcal{Y} \subseteq \mathbb{R}^d$  be convex and contain the supports of  $F$  and  $P$ , respectively. Denote by  $\nabla \psi$  the  $F$ -almost surely unique gradient of a convex function  $\psi$  in part (1) of Theorem 2.1, and let  $\psi^*$  be the conjugate of  $\psi$  over  $(\mathcal{U}, \mathcal{Y})$  in the sense of (2). Vector quantiles  $Q_P$  and vector ranks  $R_P$  are defined as follows:

$$(5) \quad \begin{aligned} Q_P(u) &\in \arg \sup_{y \in \mathcal{Y}} [y^\top u - \psi^*(y)], & u \in \mathcal{U}; \\ R_P(y) &\in \arg \sup_{u \in \mathcal{U}} [y^\top u - \psi(u)], & y \in \mathbb{R}^d. \end{aligned}$$

REMARK 2.3. Thus, we define the MK vector quantiles  $Q_P$  and ranks  $R_P$  as any solutions of the optimization problems in (5). Our definition here does not impose any moment condition and ensures that the quantities are defined for every value of the argument in the appropriate domains. By the envelope theorem and Rademacher’s theorem [58], the maps  $Q_P$  and  $R_P$  essentially coincide with the gradients  $\nabla \psi$  and  $\nabla \psi^*$  of conjugate potentials  $\psi$  and  $\psi^*$ , namely

$$(6) \quad Q_P = \nabla \psi \quad \text{a.e. on } \mathcal{U}, \quad R_P = \nabla \psi^* \quad \text{a.e. on } \mathcal{Y},$$

where “a.e.” abbreviates “almost everywhere with respect to the Lebesgue measure.” In fact, the equality holds everywhere on certain domains under condition (C) stated below. Under the additional moment conditions of Theorem 2.2, the pair  $(\psi, \psi^*)$  has the variational characterization given in (4).

When requiring regularity of vector quantiles and ranks, we shall impose the following condition on the conjugate pair of potentials  $(\psi, \psi^*)$  of Definition 2.1:

(C) Let  $\psi : \mathcal{U} \mapsto \mathbb{R}$  and  $\psi^* : \mathcal{Y} \mapsto \mathbb{R}$  form a conjugate pair over  $(\mathcal{U}, \mathcal{Y})$ , where  $\mathcal{U} \subseteq \mathbb{R}^d$  and  $\mathcal{Y} \subseteq \mathbb{R}^d$  are closed convex subsets of  $\mathbb{R}^d$ . We say that  $(\psi, \psi^*)$  satisfies condition (C) if there exist open, non-empty subsets  $\mathcal{U}_0 \subset \mathcal{U}$  and  $\mathcal{Y}_0 \subset \mathcal{Y}$  such that (i)  $\psi$  and  $\psi^*$  possess gradients  $\nabla\psi(u)$  for all  $u \in \mathcal{U}_0$ , and  $\nabla\psi^*(y)$  for all  $y \in \mathcal{Y}_0$ , respectively, (ii) the restrictions  $\nabla\psi|_{\mathcal{U}_0} : \mathcal{U}_0 \mapsto \mathcal{Y}_0$  of  $\nabla\psi$  and  $\nabla\psi^*|_{\mathcal{Y}_0} : \mathcal{Y}_0 \mapsto \mathcal{U}_0$  of  $\nabla\psi^*$  are homeomorphisms, and (iii)  $\nabla\psi|_{\mathcal{U}_0} = (\nabla\psi^*|_{\mathcal{Y}_0})^{-1}$ .

Under condition (C), we have, for the vector quantiles and ranks defined in (5),

$$(7) \quad \begin{aligned} Q_P(u) &= \nabla\psi(u) && \text{for all } u \in \mathcal{U}_0, \\ R_P(y) &= \nabla\psi^*(y) = (\nabla\psi)^{-1}(y) && \text{for all } y \in \mathcal{Y}_0; \end{aligned}$$

that is, vector ranks and quantiles are defined as gradients of conjugate potentials for each (as opposed to almost every) value in the indicated sets, and inverse functions of each other.

Sufficient conditions for condition (C) in the context of Definition 2.1 are provided by Caffarelli’s regularity theory (Villani [58], Theorem 4.14). One set of sufficient conditions is as follows.

LEMMA 2.1 (Caffarelli’s regularity, [58], Theorem 4.14). *Suppose that  $P$  and  $F$  admit densities, which (i) are of smoothness class  $C^\beta$  for some  $\beta > 0$  on non-empty convex, compact support sets  $\text{cl}(\mathcal{Y}_0)$  and  $\text{cl}(\mathcal{U}_0)$ , and (ii) are bounded, and bounded away from zero, uniformly over  $\text{cl}(\mathcal{Y}_0)$  and  $\text{cl}(\mathcal{U}_0)$ . Then condition (C) is satisfied for the conjugate pairs  $(\psi, \psi^*)$  (the existence of which follows from Theorem 2.1) such that  $\nabla\psi\#F = P$  and  $\nabla\psi^*\#P = F$ .*

We now can give our *main* definition—that of multivariate notions of quantiles and ranks, through which a depth function will be inherited from the reference distribution  $F = U_d$ .

DEFINITION 2.2 (Monge–Kantorovich depth, quantiles, ranks and signs). Let  $F$  be the spherical uniform distribution  $U_d$  on the unit ball  $\mathcal{U} = \mathbb{S}^d$ , and let  $P$  be an arbitrary distribution with support in a convex region  $\mathcal{Y} \subseteq \mathbb{R}^d$ . Denote by  $Q_P$  and  $R_P$  the vector quantiles and ranks as in (6). MK quantiles, ranks, signs and depth are defined as follows:

1. The *MK rank* of  $y \in \mathbb{R}^d$  is  $\|R_P(y)\|$  and the *MK sign* is  $R_P(y)/\|R_P(y)\|$ .
2. The *MK  $\tau$ -quantile contour* is the set  $Q_P(\mathcal{S}(\tau))$ , and the *MK depth region* with probability content  $\tau$  is  $Q_P(\mathbb{S}(\tau))$ .

3. The *MK depth* of  $y \in \mathbb{R}^d$  with respect to  $P$  is the depth of  $R_P(y)$  under  $D_{U_d}^{\text{Tukey}}$ :

$$D_P^{\text{MK}}(y) := D_{U_d}^{\text{Tukey}}(R_P(y)).$$

The notion of depth proposed in Definition 2.2 is based on a transport map from the reference spherical uniform distribution  $F = U_d$  to the distribution of interest  $P$ . Under condition (C),  $Q_P$  and  $R_P$  are continuous and are mutually inverse maps, so that the MK  $\tau$ -quantile contours are continuously deformable into spheres and the MK depth regions with probability content  $\tau$  are nested.

EXAMPLE 2.1. When the distribution  $P$  is spherically symmetric uniform on the ball with radius 2, a simple calculation yields  $\psi(u) = \|u\|^2$  and  $\psi^*(y) = \|y\|^2/4$ . The vector quantile is then  $Q_P(u) = 2u$ , and the vector rank  $R_P(y) = y/2$ . Hence,

$$D_P^{\text{MK}}(y) = D_{U_d}^{\text{Tukey}}(R_P(y)) = D_P^{\text{Tukey}}(y);$$

thus, in this case, the Monge–Kantorovich and Tukey depths coincide—as shown in Section 2.4, this is a general property of spherical distributions.

By choosing other reference distributions  $F$ , such as the uniform distribution on a unit hypercube, or the standard Gaussian distribution, we can give a more general definition of MK ranks, quantiles and signs, which may be of interest.

DEFINITION 2.3 (Monge–Kantorovich depth, quantiles, ranks and signs for general  $F$ ). Let  $F$  be an absolutely continuous reference distribution with support contained in a convex region  $\mathcal{U} \subseteq \mathbb{R}^d$ , and let  $\|\cdot\|$  be a norm on  $\mathcal{U}$ . Let  $D_F : \mathbb{R}^d \rightarrow \mathbb{R}_+$  be an associated reference depth function,  $\mathcal{K}(\tau)$  the associated  $\tau$ -quantile contour and  $\mathbb{K}(\tau)$  the associated depth region with probability content  $\tau$ . Still denoting by  $Q_P$  and  $R_P$  the vector quantiles and ranks as in (6), the MK quantiles, ranks, signs and depth are defined as follows:

1. The *MK rank* of  $y \in \mathbb{R}^d$  is  $\|R_P(y)\|$  and the *MK sign* is  $R_P(y)/\|R_P(y)\|$ .
2. The *MK  $\tau$ -quantile* is the set  $Q_P(\mathcal{K}(\tau))$  and the *MK depth region* with probability mass  $\tau$  is  $Q_P(\mathbb{K}(\tau))$ .
3. The *MK depth* of  $y \in \mathbb{R}^d$  is the depth of  $R_P(y)$  under  $D_F$ :

$$D_P^{\text{MK}}(y) := D_F(R_P(y)).$$

EXAMPLE 2.2. When the reference distribution  $F$  is  $\mathcal{N}(0, I)$  and the distribution  $P$  is  $\mathcal{N}(\mu, \Sigma)$ , one sees that  $\psi(u) = (u^\top \Sigma^{1/2} u + 2u^\top \mu + \mu^\top \Sigma^{1/2} \mu)/2$ , and  $\psi^*(x) = (x - \mu)^\top \Sigma^{-1/2} (x - \mu)/2$ . Thus,  $R_P(x) = \Sigma^{-1/2} (x - \mu)$  and the MK depth is  $D_P^{\text{MK}}(y) = 1 - \Phi(\|\Sigma^{-1/2} (y - \mu)\|)$ , where  $\Phi$  as usual stands for the univariate standard normal distribution function.

Of course, all the quantities thus defined depend on the choice of the reference distribution  $F$  and the depth function  $D_F$ .

REMARK 2.4. When the reference distribution  $F$  is spherical, it is natural to use Tukey's depth function  $D_F = D_F^{\text{Tukey}}$  to define the MK depth of  $y \in \mathbb{R}^d$  relative to  $P$  as the halfspace depth of  $R_P(y)$  relative to the reference distribution  $F$ , namely

$$D_P^{\text{MK}}(y) := D_F^{\text{Tukey}}(R_P(y)).$$

The choice of halfspace depth may be less natural for non-spherical reference distributions. One example is where  $F$  is the standard uniform distribution  $U[0, 1]^d$  on the unit cube  $[0, 1]^d$ . On  $[0, 1]^d$ , it seems natural to use the sup norm  $\|\cdot\|_\infty$  as the norm  $\|\cdot\|$  and the depth function

$$D_{U[0,1]^d}(y) = 1/2 - \|y - \mathbf{1}/2\|_\infty,$$

where  $\mathbf{1} = (1, \dots, 1)'$ , in which case  $\mathbb{K}(\tau)$  is a cube of diameter  $\tau^{1/d}$  centered at  $\mathbf{1}/2$ , yielding MK depth

$$D_P^{\text{MK}}(y) := D_{U[0,1]^d}(R_P(y)).$$

2.4. *Monge–Kantorovich depth with spherical uniform reference distribution.* Here, we consider in more detail the Monge–Kantorovich depth defined from a baseline spherical uniform distribution  $U_d$  supported on the unit ball  $\mathbb{S}^d$  of  $\mathbb{R}^d$ . Recall that this distribution is that of a random vector  $r\varphi$ , where  $r$  is uniform on  $[0, 1]$ ,  $\varphi$  is uniform on the unit sphere  $\mathbb{S}^{d-1}$ , and  $r$  and  $\varphi$  are mutually independent.

The spherical symmetry of distribution  $U_d$  produces halfspace depth contours that are concentric spheres centered at the origin, the deepest point being the origin itself. The radius  $\tau$  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

$$(8) \quad D_U(\tau u) = \begin{cases} \pi^{-1}[\theta - \cos \theta \log |\sec \theta + \tan \theta|], & d \geq 2, \\ (1 - \tau)/2, & d = 1. \end{cases}$$

Note that for  $d = 1$ ,  $u$  takes values  $\pm 1$  and, in agreement with rotational symmetry of  $U_d$ , the depth  $D_U(\tau u)$  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 optimal transport plan from  $U_d$  to  $P$ . Under the sufficient conditions for (C) provided in Lemma 2.1 (note that the conditions

on  $F$  are automatically satisfied in case  $F = U_d$ ),  $Q_P$  and  $R_P$  are continuous, and are inverse maps of each other, so that the MK depth contours are continuously deformable into spheres, the MK depth regions are nested, and regions and contours, when indexed by probability content, take the respective forms

$$Q_P(\mathbb{S}(\tau)) \quad \text{and} \quad Q_P(\mathcal{S}(\tau)), \quad \text{for } \tau \in (0, 1].$$

*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  $x \mapsto R_P(x) = 2P(x) - 1$  is the derivative of a convex function, and transports distribution  $P$  to  $U_1$ , which is uniform on  $[-1, 1]$ , that is,  $R_P \# P = U_1$ . Hence,  $R_P$  coincides with the MK vector rank of Definition 2.1. Therefore, for each  $x \in \mathbb{R}$ ,

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

and MK depth coincides with Tukey depth for all distributions  $P$  with non-vanishing densities on the real line.

More generally (still in the univariate case), denoting by  $F_1$  and  $F_2$  the distribution functions associated with two absolutely continuous distributions  $P_1$  and  $P_2$ , the mapping  $F_2^{-1} \circ F_1$ , being monotone increasing, is also the optimal transport from  $P_1$  to  $P_2$ . The same transformation has been studied, in a different context, by Doksum [12] and Doksum and Sievers [13], and plays a central role in the asymptotic representation of linear rank statistics [22]; see also the concept of convex ordering proposed by van Zwet [55].

*MK depth is halfspace depth for elliptical families.* As explained in the [Introduction](#), a  $d$ -dimensional random vector  $X$  has elliptical distribution  $P_{\mu, \Sigma, g}$  with location  $\mu \in \mathbb{R}^d$ , positive definite symmetric  $d \times d$  scatter matrix  $\Sigma$  and radial density function  $g$  (radial distribution function  $G$ ) if and only if, denoting by  $\Sigma^{1/2}$  the symmetric root of  $\Sigma$ ,  $Y := \Sigma^{-1/2}(X - \mu)$  has spherical distribution  $P = P_{0, I, g}$  (hence,  $\|Y\|$  has density  $f$ ), which holds if and only if

$$(9) \quad R_P(Y) := \frac{Y}{\|Y\|} G(\|Y\|) \text{ is distributed according to } U_d.$$

Let  $\Psi(t) = \int_{-\infty}^t G(r) dr$ , and note that the map  $z \mapsto R_P(z)$  is the gradient of  $\psi^*(z) := \Psi(\|z\|)$  so that, from (9),  $\nabla \psi^* \# P = U_d$  as Definition 2.1 requires. That  $\psi^*$  is convex follows from Theorem 5.1 of [49] by noting that  $\psi^*$  is a composition of  $\Psi : \mathbb{R} \rightarrow \mathbb{R}$ , a convex, non-decreasing map, and  $\|\cdot\| : \mathbb{R}^d \rightarrow \mathbb{R}$ , a convex function by definition of the norm. As a consequence, the mapping  $R_P$  in (9) is the MK vector rank function associated with  $P = P_{0, I, f}$ ; and, the MK depth contours



(with probability content  $\tau$ ) of  $P$  are spheres with radii  $G^{-1}(\tau)$  centered at the origin:

$$D_P(x) = \{y \in \mathbb{R}^d : \|y\| \leq G^{-1}(\tau)\}.$$

These spheres are halfspace depth contours for  $P$ . This is the precise sense in which MK depth reduces to halfspace depth for elliptical families.

It should be noted above that we treat location and scatter parameters as known, and transform  $X$  to a vector  $Y$  in isotropic position. This transformation ensures basic invariance properties of the resulting depth, ranks and quantiles with respect to affine transformations. When those parameters are unknown, they will have to be replaced with by affine-equivariant estimators, as in the usual definition of elliptical ranks and signs (see, e.g., [23]) in order to insure similar invariance properties for the empirical analogs. Without the aforementioned transformation, however, the invariance properties are not guaranteed, owing to the fact that composition of two gradients of convex functions is not necessarily the gradient of a convex function, unless the composition has a specific structure, as is the case above.

**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 assume that condition (C) holds. Then the MK vector quantiles and ranks of Definition 2.1 are

$$(10) \quad Q_P(u) := \nabla \psi(u), \quad R_P(y) := \nabla \psi^*(y) = (\nabla \psi)^{-1}(y),$$

for each  $u \in \mathcal{U}_0$  and  $y \in \mathcal{Y}_0$ , respectively. We define  $\Phi_0(\mathcal{U}, \mathcal{Y})$  as a collection of conjugate potentials  $(\varphi, \varphi^*)$  on  $(\mathcal{U}, \mathcal{Y})$  such that  $\varphi(u_0) = 0$  for some fixed point  $u_0 \in \mathcal{U}_0$ . Under the conditions of Theorem 2.2, the potentials  $(\psi, \psi^*)$  solve the dual problem

$$(11) \quad \int \psi dF + \int \psi^* dP = \inf_{(\varphi, \varphi^*) \in \Phi_0(\mathcal{U}, \mathcal{Y})} \left( \int \varphi dF + \int \varphi^* dP \right).$$

Constraining the conjugate pair to lie in  $\Phi_0(\mathcal{U}, \mathcal{Y})$  is a normalization that (without any loss of generality) pins down the constant, so that  $(\psi, \psi^*)$  are uniquely determined, as argued in the proof.

We propose empirical versions of MK quantiles and ranks based on estimators  $\hat{P}$  of  $P$ . The typical case is when the reference measure  $F$  is known. However, our theory allows us to handle the case where  $F$  is itself unknown, and estimated by some  $\hat{F}$ . This is indeed useful for at least two reasons. First, we may be interested in a classical problem of comparing one distribution  $P$  to a reference distribution  $F$ , both of which are known only via a random sample available from each of them. Second, we may be interested in discretizing  $F$  for computational reasons, as we discuss in Section 4, in which case the discretized  $F$  is the estimator of  $F$ .

3.1. *Conditions on the estimators of  $P$  and  $F$ .* Suppose that  $\{\hat{P}_n\}_{n=1}^\infty$  and  $\{\hat{F}_n\}_{n=1}^\infty$  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 that

$$(12) \quad d_{\text{BL}}(\hat{P}_n, P) \xrightarrow{\mathbb{P}^*} 0, \quad d_{\text{BL}}(\hat{F}_n, F) \xrightarrow{\mathbb{P}^*} 0,$$

where  $\xrightarrow{\mathbb{P}^*}$  throughout denotes convergence in (outer) probability under the probability measure  $\mathbb{P}$  driving  $\hat{P}_n$  and  $\hat{F}_n$ ; see van der Vaart and Wellner [56]. A basic example is where  $\hat{P}_n$  is the empirical distribution of a random sample  $(Y_i)_{i=1}^n$  drawn from  $P$  and  $\hat{F}_n$  is the empirical distribution of a random sample  $(U_i)_{i=1}^n$  drawn from  $F$ . Other, much more complicated examples, including smoothed empirical measures and data originating from dependent processes, satisfy sufficient conditions for (12) that we now give. In order to develop some examples, we introduce an 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  $W_{t,n} \in \mathcal{W} \subseteq \mathbb{R}^d$  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  $\|g\|_{\text{BL}(\mathcal{W})} < \infty$ , the law of large numbers holds:

$$(13) \quad \frac{1}{n} \sum_{t=1}^n g(W_{t,n}) \xrightarrow{\mathbb{P}} \int g(w) dP_W(w).$$

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

(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 an irreducible and aperiodic Markov chain with invariant distribution  $P_W$ ;

(E.4)  $W_{t,n} = w_{t,n}$ , where  $(w_{t,n})_{t=1}^n$  is a deterministic sequence of points such that (13) holds deterministically.

For a detailed motivation and discussion of the use of deterministic sequences such as the so-called *low-discrepancy sequences*; see, for example, Chapter 9 of [32] and, more particularly, page 314.

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

$$\begin{aligned} \hat{P}_W(A) &= \frac{1}{n} \sum_{t=1}^n 1\{W_{t,n} \in A\}, \\ \tilde{P}_W(A) &= \frac{1}{n} \sum_{t=1}^n \int_{\mathbb{R}^d} 1\{W_{t,n} + h_n \varepsilon \in A \cap \mathcal{W}\} d\Phi(\varepsilon), \end{aligned}$$

where  $\Phi$  is the probability law of the standard  $d$ -dimensional Gaussian vector  $N(0, I_d)$ , and  $h_n \geq 0$  a semi-positive-definite matrix of bandwidths such that  $\|h_n\| \rightarrow 0$  as  $n \rightarrow \infty$ . 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 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*

$$d_{\text{BL}}(\hat{P}_W, P_W) \xrightarrow{\mathbb{P}^*} 0, \quad d_{\text{BL}}(\tilde{P}_W, P_W) \xrightarrow{\mathbb{P}^*} 0.$$

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  $\tilde{P}_U$  satisfy condition (12).

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

3.2. *Empirical vector quantiles and ranks.* We base empirical versions of MK quantiles, ranks and depth on estimators  $\hat{P}_n$  for  $P$  and  $\hat{F}_n$  for  $F$  satisfying (12). This includes cases where the reference measure  $F$  is known, that is,  $\hat{F}_n = F$ . Recall assumption (C) is maintained throughout this section.

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

$$(14) \quad \hat{Q}_n(u) \in \arg \sup_{y \in \mathcal{Y}} [y^\top u - \hat{\psi}_n^*(y)], \quad \hat{R}_n(y) \in \arg \sup_{u \in \mathcal{U}} [y^\top u - \hat{\psi}_n(u)],$$

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

$$(15) \quad \int \hat{\psi}_n d\hat{F}_n + \int \hat{\psi}_n^* d\hat{P}_n = \inf_{(\varphi, \varphi^*) \in \Phi_0(\mathcal{U}, \mathcal{Y})} \left( \int \varphi d\hat{F}_n + \int \varphi^* d\hat{P}_n \right).$$

We now state the main result of Section 3.

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 the probability measures  $P$  and  $F$  are absolutely continuous with respect to the Lebesgue measure, with  $\text{support}(P) \subseteq \mathcal{Y}$  and  $\text{support}(F) \subseteq \mathcal{U}$ . 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 (12). Suppose that condition (C) holds for the solution*

of (11) with  $\mathcal{Y}_0 := \text{int}(\text{support}(P))$  and  $\mathcal{U}_0 := \text{int}(\text{support}(F))$ . Then, as  $n \rightarrow \infty$ , for any closed set  $K \subset \mathcal{U}_0$  and any closed set  $K' \subset \mathcal{Y}_0$ ,

$$\sup_{u \in K} \|\hat{Q}_n(u) - Q_P(u)\|_{\mathbb{P}^*} \rightarrow 0, \quad \sup_{y \in K'} \|\hat{R}_n(y) - R_P(y)\|_{\mathbb{P}^*} \rightarrow 0,$$

and

$$\sup_{A \subseteq K} d_H(\hat{Q}_n(A), Q_P(A))_{\mathbb{P}^*} \rightarrow 0, \quad \sup_{A' \subseteq K'} d_H(\hat{R}_n(A'), R_P(A'))_{\mathbb{P}^*} \rightarrow 0,$$

where the suprema are taken over non-empty subsets.

The first result establishes the uniform consistency of empirical vector quantile and rank maps, hence also of empirical ranks and signs. The set  $Q_P(K)$  with  $K = \mathbb{K}(\tau)$  is the statistical depth contour with probability content  $\tau$ . The second result, therefore, establishes consistency of the approximation  $\hat{Q}_n(K)$  to the theoretical depth region  $Q_P(K)$ .

**3.3. Empirical MK quantiles, ranks and signs, and their convergence.** We work with the conditions of the previous theorem, but here, for the sake of simplicity, we first consider the lead case where  $F$  is known, that is,  $\hat{F}_n = F$ .

**DEFINITION 3.2** (Empirical MK depth, quantiles, ranks and signs for known  $F$ ). Let  $F$  be an absolutely continuous reference distribution with support contained in a convex region  $\mathcal{U} \subseteq \mathbb{R}^d$ , and let  $\|\cdot\|$  be a norm on  $\mathcal{U}$ . The MK empirical quantiles, ranks, signs and depth are defined as follows:

1. The MK empirical rank and sign of  $y \in \mathbb{R}^d$  are  $\|\hat{R}_n(y)\|$  and  $\hat{R}_n(y)/\|\hat{R}_n(y)\|$ .
2. The MK empirical  $\tau$ -quantile contour is the set  $\hat{Q}_n(\mathbb{K}(\tau))$  and the MK empirical depth region with probability mass  $\tau$  is  $\hat{Q}_n(\mathbb{K}(\tau))$ .
3. The MK empirical depth of  $y \in \mathbb{R}^d$  is the depth of  $\hat{R}_n(y)$  under  $D_F$ :

$$\hat{D}_{P,n}^{\text{MK}}(y) := D_F(\hat{R}_n(y)).$$

Uniform convergence of empirical MK rank, signs and depth to their theoretical counterparts follows by an application of the extended continuous mapping theorem.

**COROLLARY 3.1.** Under the assumptions of Theorem 3.1, and assuming that  $D_F$  is continuous on  $\mathcal{U}_0$ , for any closed set  $K' \subset \mathcal{Y}_0$ , as  $n \rightarrow \infty$ ,

$$\sup_{y \in K'} \|\|\hat{R}_n(y)\| - \|R_P(y)\|\|_{\mathbb{P}^*} \rightarrow 0,$$

$$\sup_{y \in K'} \|\hat{R}_n(y)/\|\hat{R}_n(y)\| - R_P(y)/\|R_P(y)\|\|_{\mathbb{P}^*} \rightarrow 0,$$

$$\sup_{y \in K'} |\hat{D}_{P,n}^{\text{MK}}(y) - D_P^{\text{MK}}(y)|_{\mathbb{P}^*} \rightarrow 0.$$

Uniform convergence of MK empirical  $\tau$ -quantile contours and MK empirical depth regions with probability content  $\tau$  follows also through an application of the extended continuous mapping theorem.

**COROLLARY 3.2.** *Under the assumptions of Theorem 3.1, for any  $\mathcal{T} \subset (0, 1)$  such that  $\text{cl}(\bigcup_{\tau \in \mathcal{T}} \mathbb{K}(\tau)) \subset \mathcal{U}_0$ , as  $n \rightarrow \infty$ ,*

$$\sup_{\tau \in \mathcal{T}} d_H(\hat{Q}_n(\mathbb{K}(\tau)), Q_P(\mathbb{K}(\tau))) \xrightarrow{\mathbb{P}^*} 0, \quad \sup_{\tau \in \mathcal{T}} d_H(\hat{Q}_n(\mathcal{K}(\tau)), Q_P(\mathcal{K}(\tau))) \xrightarrow{\mathbb{P}^*} 0.$$

The main results are derived for a fully specified reference distribution  $F$  and the associated depth function  $D_F$ , depth regions  $\mathbb{K}(\tau)$  and quantile contours  $\mathcal{K}(\tau)$ . There are cases where these will be approximated numerically or using data-driven substitutes. The same definitions and results extend naturally when  $D_F$ ,  $\mathbb{K}(\tau)$  and  $\mathcal{K}(\tau)$  are replaced by uniformly consistent estimators  $\hat{D}_{F,n}$ ,  $\hat{\mathbb{K}}_n(\tau)$  and  $\hat{\mathcal{K}}_n(\tau)$  such that

$$(16) \quad \begin{aligned} \sup_{u \in K} |\hat{D}_{F,n}(u) - D_F(u)| &\xrightarrow{\mathbb{P}^*} 0, \\ \sup_{\tau \in \mathcal{T}} d_H(\hat{\mathbb{K}}_n(\tau), \mathbb{K}(\tau)) &\xrightarrow{\mathbb{P}^*} 0, \\ \sup_{\tau \in \mathcal{T}} d_H(\hat{\mathcal{K}}_n(\tau), \mathcal{K}(\tau)) &\xrightarrow{\mathbb{P}^*} 0, \end{aligned}$$

where  $K$  is any closed subset of  $\mathcal{U}_0$ . These high-level conditions hold trivially for the numerical approximations we use in Section 4. They also hold, for example, for Tukey's halfspace depth under regularity conditions. We will not discuss these conditions here.

**DEFINITION 3.3** (Empirical MK depth, quantiles, ranks and signs with estimated  $F$ ). Let  $F$  be an absolutely continuous reference distribution with support contained in a convex and compact region  $\mathcal{U} \subset \mathbb{R}^d$ , and let  $\|\cdot\|$  be a norm on  $\mathcal{U}$ . Given estimators  $\hat{D}_{F,n}$ ,  $\hat{\mathbb{K}}_n(\tau)$  and  $\hat{\mathcal{K}}_n(\tau)$  satisfying (16), the MK empirical quantiles, ranks, signs and depth are defined as follows:

1. The MK empirical rank and sign of  $y \in \mathbb{R}^d$  are  $\|\hat{\mathbf{R}}_n(y)\|$  and  $\hat{\mathbf{R}}_n(y)/\|\hat{\mathbf{R}}_n(y)\|$ .
2. The MK empirical  $\tau$ -quantile contour is the set  $\hat{Q}_n(\hat{\mathcal{K}}_n(\tau))$ , and the MK empirical depth region with probability mass  $\tau$  is  $\hat{Q}_n(\hat{\mathbb{K}}_n(\tau))$ .
3. The MK empirical depth of  $y \in \mathbb{R}^d$  is the depth of  $\hat{\mathbf{R}}_n(y)$  under  $\hat{D}_{F,n}$ :

$$\hat{D}_{P,n}^{\text{MK}}(y) := \hat{D}_{F,n}(\hat{\mathbf{R}}_n(y)).$$

**COROLLARY 3.3.** *Under the assumptions of the previous corollary and conditions (16), the conclusions of Corollary 3.1 hold, and the conclusions of Corollary 3.2 hold in the following form:*

$$\sup_{\tau \in \mathcal{T}} d_H(\hat{Q}_n(\hat{\mathbb{K}}_n(\tau)), Q_P(\mathbb{K}(\tau))) \xrightarrow{\mathbb{P}^*} 0, \quad \sup_{\tau \in \mathcal{T}} d_H(\hat{Q}_n(\hat{\mathcal{K}}_n(\tau)), Q_P(\mathcal{K}(\tau))) \xrightarrow{\mathbb{P}^*} 0.$$

**4. Computing empirical quantiles and depth regions.** Here, we provide computational characterizations of the empirical quantiles, ranks and depth regions for various cases of interest.

*Smooth  $\hat{P}_n$  and  $\hat{F}_n$ .* Suppose  $\hat{P}_n$  and  $\hat{F}_n$  satisfy Caffarelli regularity conditions, so that  $\hat{Q}_n = \nabla \hat{\psi}_n$  and  $\hat{R}_n = \nabla \hat{\psi}_n^*$ , with  $(\hat{\psi}_n, \hat{\psi}_n^*)$  satisfying (C). The MK empirical vector quantile maps  $\hat{Q}_n$  and  $\hat{R}_n$  can then be computed with the algorithm of Benamou and Brenier [3].

*Discrete  $\hat{P}_n$  and smooth  $\hat{F}_n$ .* Suppose now  $\hat{P}_n$  is a discrete estimator of  $P$  and  $\hat{F}_n$  an absolutely continuous distribution with convex compact support  $\mathcal{U} \subset \mathbb{R}^d$ . Let  $\hat{P}_n$  be of the form  $\hat{P}_n = \sum_{k=1}^{K_n} p_{k,n} \delta_{y_{k,n}}$  for some integer  $K_n$ , some non-negative weights  $p_{1,n}, \dots, p_{K_n,n}$  such that  $\sum_{k=1}^{K_n} p_{k,n} = 1$ , and  $y_{1,n}, \dots, y_{K_n,n} \in \mathbb{R}^d$ . 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 MK empirical vector quantile map  $\hat{Q}_n$  is then equal (almost everywhere) to the gradient of a convex map  $\hat{\psi}_n$  such that  $\nabla \hat{\psi}_n \# \hat{F}_n = \hat{P}_n$ , that is, the  $\hat{F}_n$ -almost surely unique map  $\hat{Q}_n = \nabla \hat{\psi}_n$  satisfying the following:

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

The following characterization of  $\hat{\psi}_n$  specializes Kantorovich duality to this discrete-continuous case (see, e.g., [18]).

**LEMMA 4.1.** *There exist unique (up to an additive constant)  $n$ -tuples of weights  $\{v_1^*, \dots, v_n^*\}$  such that  $\hat{\psi}_n(u) = \max_{1 \leq k \leq K_n} \{u^\top y_{k,n} - v_k^*\}$  satisfies conditions (1), (2) and (3). The function  $v = \{v_1, \dots, v_n\} \mapsto \int \hat{\psi}_n d\hat{F}_n + \sum_{k=1}^{K_n} p_{k,n} v_k$  is convex and minimized at  $v^* = \{v_1^*, \dots, v_n^*\}$ .*

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

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

maps  $\{y_{1,n}, \dots, y_{K_n,n}\}$  into  $K_n$  regions of a partition of  $\mathcal{U}$ , called a *power diagram*. The estimator  $\hat{R}_n$  of the MK vector rank can be computed according to formula (14) after computing the conjugate  $\hat{\psi}_n^*$  of  $\hat{\psi}_n$  via

$$\hat{\psi}_n^*(y) = \sup_{u \in \mathcal{U}} \{u^\top y - \hat{\psi}_n(u)\}.$$

The empirical depth, depth regions and quantiles can be computed using the depth function according to their theoretical definitions.

*Discrete  $\hat{P}_n$  and  $\hat{F}_n$ .* Particularly amenable to computation is the case when 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 some  $n$ -tuple  $\mathcal{Y}_n = \{y_1, \dots, y_n\}$  of points in  $\mathbb{R}^d$  and  $\hat{F}_n = \sum_{j=1}^n \delta_{u_j}/n$ , for some other  $n$ -tuple  $\mathcal{U}_n = \{u_1, \dots, u_n\}$  of points in  $\mathbb{R}^d$ . The restriction of the quantile map  $\hat{Q}_n$  to  $\mathcal{U}_n$  is the bijection  $u \mapsto y = \hat{Q}_n|_{\mathcal{U}_n}(u)$  from  $\mathcal{U}_n$  onto  $\mathcal{Y}_n$  and  $\hat{R}_n|_{\mathcal{Y}_n}$  is its inverse. The solutions  $\hat{Q}_n$  and  $\hat{R}_n$  can be computed with any optimal assignment algorithm. More generally, in the case of any two discrete estimators  $\hat{P}_n$  and  $\hat{F}_n$ , the problem of computing  $\hat{Q}_n$  or  $\hat{R}_n$  is a linear programming problem.

*Visualization of empirical MK depth and quantile contours.* Whenever  $\hat{P}_n$  is finitely discrete, then the MK empirical depth regions and quantile contours are finite sets of points. For visualization purposes it may be helpful to transform them into nicer looking objects which are close to the original objects in terms of Hausdorff distance. In the example below, we used  $\alpha$ -hulls to create approximations to the depth regions and took the boundaries of the set as a numerical approximation to the quantile contours. It may also be possible to use polygonization methods such as those in [11] for  $d = 2$  and [21] for  $d = 3$ .

**EXAMPLE 4.1** (Computing MK depth regions). In the example illustrated in Figure 2, we use a discrete approximation  $\hat{F}_n$  to the spherical uniform reference distribution. Figure 2 shows the MK empirical depth contours for the same banana-shaped distribution as in Figure 1. The specific construction used for Figure 2 is the following:  $\hat{P}_n$  is the empirical distribution of a random sample  $\mathcal{Y}_n$  drawn from the banana-shaped distribution in  $\mathbb{R}^2$ , with  $n = 9999$ ;  $\hat{F}_n$  is a discrete approximation to  $F$  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. MK empirical depth regions are  $\alpha$ -hulls of  $\hat{Q}_n(\mathcal{U}_n \cap \mathbb{S}(\tau))$  for 11 values of  $\tau \in (0, 1)$  (see [17] for a definition of  $\alpha$ -hulls). The  $\alpha$ -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  $[-1, 1]$ ,  $\Phi$  is uniform on  $[0, 2\pi]$ ,  $Z$  is uniform on  $[0, 1]$ ,  $X$ ,  $Z$  and  $\Phi$  are independent, and  $R = 0.2Z(1 + (1 - |X|)/2)$ .

## 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  $(\psi, \psi^*)$  of convex potentials  $\psi : \mathcal{U} \mapsto \mathbb{R} \cup \{\infty\}$

and  $\psi^* : \mathcal{Y} \mapsto \mathbb{R} \cup \{\infty\}$  is a conjugate pair over  $(\mathcal{U}, \mathcal{Y})$  if, for each  $u \in \mathcal{U}$  and  $y \in \mathcal{Y}$ ,

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

In the sequel, we consider a fixed pair  $(\psi, \psi^*)$  satisfying condition (C) and a sequence  $(\psi_n, \psi_n^*)$ ,  $n \in \mathbb{N}$ , of conjugate potentials approaching  $(\psi, \psi^*)$  in the following sense:

(A)  $\psi_n(u) \rightarrow \psi(u)$  in  $\mathbb{R} \cup \{\infty\}$  pointwise in  $u$  in a dense subset of  $\mathcal{U}$ , and  $\psi_n^*(y) \rightarrow \psi^*(y)$  in  $\mathbb{R} \cup \{\infty\}$  pointwise in  $y$  in a dense subset of  $\mathcal{Y}$ , as  $n \rightarrow \infty$ .

Condition (A) is equivalent to requiring that either  $\psi_n$  or  $\psi_n^*$  converge pointwise over dense subsets. There is no loss of generality in assuming that both converge.

Define the maps

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

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

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

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

$$(17) \quad Q_n(u) \in \arg \sup_{y \in \mathcal{Y}} [y^\top u - \psi_n^*(y)], \quad R_n(y) \in \arg \sup_{u \in \mathcal{U}} [y^\top u - \psi_n(u)].$$

It is useful to note that

$$R_n(y) \in \partial \psi_n^*(y) \quad \text{for } y \in \mathcal{Y}; \quad Q_n(u) \in \partial \psi_n(u) \quad \text{for } u \in \mathcal{U},$$

where  $\partial$  denotes the subdifferential 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 (17) (by Proposition 2.4 in Villani [58]). 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., [58]), so the solutions to (17) are unique almost everywhere on  $u \in \mathcal{U}$  and  $y \in \mathcal{Y}$ .

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

$$\sup_{u \in K} \|Q_n(u) - Q(u)\| \rightarrow 0, \quad \sup_{y \in K'} \|R_n(y) - R(y)\| \rightarrow 0.$$



REMARK A.1. This result appears to be new. It complements the result stated in Lemma 5.4 in Villani [59] 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,  $1 \leq p < \infty$ :

$$\int_{\mathcal{U}} \|Q_n(u) - Q(u)\|^p dF(u) \rightarrow 0,$$

$$\int_{\mathcal{Y}} \|R_n(y) - R(y)\|^p dP(y) \rightarrow 0,$$

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

$$\sup_{n \in \mathbb{N}} \int_{\mathcal{U}} \|Q_n(u)\|^{\bar{p}} + \|Q(u)\|^p dF(u) < \infty,$$

$$\sup_{n \in \mathbb{N}} \int_{\mathcal{Y}} \|R_n(y)\|^{\bar{p}} + \|R(y)\|^p dP(y) < \infty.$$

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 [58, 60]).

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

$$\psi_n(u) = \int_0^u Q_n(t) dt, \quad Q_n(t) = t \cdot 1(t \leq 1 - 1/n) + 10 \cdot 1(t > 1 - 1/n).$$

Then

$$\psi_n(u) = \frac{1}{2}u^2 1\left(u \leq 1 - \frac{1}{n}\right) + \left\{10\left(u - \left(1 - \frac{1}{n}\right)\right) + \frac{1}{2}\left(1 - \frac{1}{n}\right)^2\right\} 1\left(u > 1 - \frac{1}{n}\right)$$

converges uniformly on  $[0, 1]$  to  $\varphi(u) = 2^{-1}u^2$ . The latter potential has the gradient map  $Q : [0, 1] \mapsto \mathcal{Y}_0 = [0, 1]$  defined by  $Q(t) = t$ . We clearly have that  $\sup_{t \in K} |Q_n(t) - Q(t)| \rightarrow 0$  for any compact subset  $K$  of  $(0, 1)$ . However, the uniform convergence over the entire region  $[0, 1]$  fails, since, for all  $n$ ,  $\sup_{t \in [0, 1]} |Q_n(t) - Q(t)| \geq 9$ . Therefore, the theorem cannot be strengthened in general.

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

$$Q_n(A) := \{Q_n(u) : u \in A\}, \quad Q(A) := \{Q(u) : u \in A\}, \quad A \subseteq K,$$

$$R_n(A') := \{R_n(y) : y \in A'\}, \quad R(A') := \{R(y) : y \in A'\}, \quad A' \subseteq K',$$

where  $K \subset \mathcal{U}_0$  and  $K' \subset \mathcal{Y}_0$  are compact sets, and the subsets  $A$  and  $A'$  are understood to be non-empty.

**COROLLARY A.1** (Convergence of sets of subdifferentials). *Under the conditions of Theorem A.1, we have that*

$$\sup_{A \subseteq K} d_H(Q_n(A), Q(A)) \rightarrow 0, \quad \sup_{A' \subseteq K'} d_H(R_n(A'), R(A')) \rightarrow 0.$$

**COROLLARY A.2** (Convergence of sets of subdifferentials). *Assume the conditions of Theorem A.1 hold. For any sequence of sets  $\{A_n\} \subseteq K$  and  $\{A'_n\} \subseteq K'$  such that  $d_H(A_n, A) \rightarrow 0$  and  $d_H(A'_n, A') \rightarrow 0$  for some sets  $A$  and  $A'$ , we have*

$$d_H(Q_n(A_n), Q(A)) \rightarrow 0, \quad d_H(R_n(A'_n), R(A')) \rightarrow 0.$$

**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.

Equip  $\mathcal{Y}$  and  $\mathcal{U}$  with absolutely continuous probability measures  $P$  and  $F$ , respectively, and let

$$\mathcal{Y}_0 := \text{int}(\text{support}(P)), \quad \mathcal{U}_0 := \text{int}(\text{support}(F)).$$

We consider sequences of measures  $P_n$  and  $F_n$  approximating  $P$  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:

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

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

$$(18) \quad \int \psi_n dF_n + \int \psi_n^* dP_n = \inf_{(\varphi, \varphi^*) \in \Phi_0(\mathcal{U}, \mathcal{Y})} \left( \int \varphi dF_n + \int \varphi^* dP_n \right).$$

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

$$(19) \quad \int \psi dF + \int \psi^* dP = \inf_{(\varphi, \varphi^*) \in \Phi_0(\mathcal{U}, \mathcal{Y})} \left( \int \varphi dF + \int \varphi^* dP \right).$$

It is known that solutions to these problems exist; see, for example, Villani [58]. 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 the probability measures  $P$  and  $F$  are absolutely continuous with respect to the Lebesgue measure,*

with  $\text{support}(P) \subseteq \mathcal{Y}$  and  $\text{support}(F) \subseteq \mathcal{U}$ . Suppose that condition (W) holds, and that condition (C) holds for a solution  $(\psi, \psi^*)$  of (19) for the sets  $\mathcal{U}_0$  and  $\mathcal{Y}_0$  defined as above. Then the conclusions of Theorem A.1 and Corollaries A.1 and A.2 hold.

## APPENDIX B: PROOFS

**B.1. Proof of Theorem A.1.** The proof of Theorem A.1 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  $x_n \rightarrow x$  in  $\mathbb{D}$ , we have that  $f_n(x_n) \rightarrow f(x)$ .*

For the proof, see, for example, Rockafellar and Wets [50]. The proof also relies on the following convergence result, which is a consequence of Theorem 7.17 in Rockafellar and Wets [50]. 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  $\mathcal{X}_0 = \arg \inf_{x \in \mathbb{R}^d} g(x) \subset \text{int}(\mathcal{D}_0)$ , where  $\mathcal{D}_0 = \{x \in \mathbb{R}^d : g(x) < \infty\}$ , and  $\text{int}(\mathcal{D}_0)$  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  $g_n(x) \rightarrow g(x)$  pointwise in  $x \in \mathbb{R}_0^d$ , where  $\mathbb{R}_0^d$  is a countable dense subset of  $\mathbb{R}^d$ . Then any  $x_n \in \arg \inf_{x \in \mathbb{R}^d} g_n(x)$  satisfies*

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

*In particular, if  $\mathcal{X}_0$  is a singleton  $\{x_0\}$ ,  $x_n \rightarrow x_0$ .*

The proof of this lemma is given below, immediately after the conclusion of the proof of Theorem A.1.

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

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

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

By the convexity of  $\psi_n$  and  $\psi_n^*$  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  $n$  in the definitions of  $g_{n,u}$  and  $\bar{g}_{n,y}$  above.

Condition (A) assumes pointwise convergence of  $\psi_n^*$  to  $\psi^*$  on a dense subset of  $\mathcal{Y}$ . By Theorem 7.17 in Rockafellar and Wets [50], this implies the uniform convergence of  $\psi_n^*$  to  $\psi^*$  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  $u_n \rightarrow u \in K$ , a compact subset of  $\mathcal{U}_0$ , and any  $y \in (\text{int } \mathcal{Y}) \setminus \partial \mathcal{D}_1$ ,

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

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

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

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

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

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

$$Q_n(u) \rightarrow Q(u) \quad \text{uniformly in } u \in K.$$

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

**B.2. Proof of Lemma B.2.** By assumption,  $\mathcal{X}_0 = \arg \min g \subset \text{int}(\mathcal{D}_0)$ , 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  $\varepsilon_0$  such that  $B_{\varepsilon_0}(\mathcal{X}_0) \subset \text{int}(\mathcal{D}_0)$ ,

$$(20) \quad g(x_0) < \inf_{x \in \partial B_\varepsilon(\mathcal{X}_0)} g(x),$$

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 [50], 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  $\partial \mathcal{D}_0$ , that is,  $K \cap \partial \mathcal{D}_0 = \emptyset$ . Hence,  $g_n \rightarrow g$  uniformly on  $B_{\varepsilon_0}(\mathcal{X}_0)$ . This and (20) imply that eventually, that is, 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  $\varepsilon > 0$  can be set as small as desired, it follows that any  $x_n \in \arg \inf g_n$  is such that  $d(x_n, \mathcal{X}_0) \rightarrow 0$ .

**B.3. Proof of Corollary A.1.** By Theorem A.1 and the definition of Hausdorff distance, for any non-empty subset  $A$ ,

$$\begin{aligned} & \sup_{A \subseteq K} d_H(Q_n(A), Q(A)) \\ &= \sup_{A \subseteq K} \left( \sup_{u \in A} \inf_{\bar{u} \in A} \|Q_n(\bar{u}) - Q(u)\| \vee \sup_{\bar{u} \in A} \inf_{u \in A} \|Q_n(\bar{u}) - Q(u)\| \right) \\ &\leq \sup_{A \subseteq K} \left( \sup_{u \in A} \|Q_n(u) - Q(u)\| \vee \sup_{\bar{u} \in A} \|Q_n(\bar{u}) - Q(\bar{u})\| \right) \\ &= \sup_{u \in K} \|Q_n(u) - Q(u)\| \rightarrow 0, \end{aligned}$$

as was to be shown. The proof of the second claim is identical.

**B.4. Proof of Corollary A.2.** We have that

$$\begin{aligned} & d_H(Q_n(A_n), Q(A)) \\ &\leq d_H(Q_n(A_n), Q(A_n)) + d_H(Q(A_n), Q(A)) \\ &\leq \sup_{A \subseteq K} d_H(Q_n(A), Q(A)) \\ &\quad + \sup_{\bar{u}, u \in K} \{Q(\bar{u}) - Q(u) : \|\bar{u} - u\| \leq d_H(A_n, A)\} \rightarrow 0, \end{aligned}$$

where the first inequality holds by the triangle inequality, the second inequality holds by definition and by  $A_n, A \subseteq K$ , and the last conclusion follows by Corollary A.1 and continuity of the map  $u \mapsto Q(u)$  on  $u \in K$ .

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 uniform convergence. First, we notice that, for any pair  $(\varphi, \varphi^*) \in \Phi_0(\mathcal{U}, \mathcal{Y})$ ,

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

with  $\|A\| := \sup_{a \in A} \|a\|$  for  $A \subseteq \mathbb{R}^d$ , 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 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  $\mathbb{N}' \subseteq \mathbb{N}$ ) have their limit points in this set:

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

This is true, since uniform limits of convex functions are necessarily convex (see [50]), and since

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

and

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

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

*Step 2.* The claim here is that

$$(21) \quad I_n := \int \psi_n dF_n + \int \psi_n^* dP_n \xrightarrow{n \in \mathbb{N}} \int \psi dF + \int \psi^* dP =: I_0.$$

Indeed,

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

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

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

where  $x \lesssim y$  means  $x \leq Ay$ , for some constant  $A$  that does not depend on  $n$ . Moreover, by definition,

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

but

$$\begin{aligned} |I_n - II_n| &\leq \left| \int \psi_n d(F_n - F) \right| + \left| \int \psi_n^* d(P_n - P) \right| \\ &\lesssim d_{\text{BL}}(F_n, F) + d_{\text{BL}}(P_n, P) \rightarrow 0. \end{aligned}$$

*Step 3.* Here we conclude. First, we observe that the solution pair  $(\psi, \psi^*)$  to the limit Kantorovich problem is unique on  $\mathcal{U}_0 \times \mathcal{Y}_0$  in the sense that any other solution  $(\varphi, \varphi^*)$  agrees with  $(\psi, \psi^*)$  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 imply 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

$$\begin{aligned} \psi(u) - \psi(u_0) &= \int_0^1 \nabla\psi(u_0 + v^\top(u - u_0))^\top(u - u_0) dv \\ &= \int_0^1 \nabla\varphi(u_0 + v^\top(u - u_0))^\top(u - u_0) dv \\ &= \varphi(u) - \varphi(u_0) \end{aligned}$$

for almost all  $u \in B_\varepsilon(u_1)$ , which is a contradiction]. However, the statement that  $\nabla\psi \neq \nabla\varphi$  on a set  $K \subset \mathcal{U}_0$  of positive Lebesgue measure would contradict the fact that any solution  $\psi$  or  $\varphi$  of the Kantorovich problem must obey

$$\int h \circ \nabla\varphi dF = \int h \circ \nabla\psi dF = \int h dP$$

for each bounded continuous  $h$ , that is,  $\nabla\varphi\#F = \nabla\psi\#F = P$ , see page 72 in Villani [58]. Analogous arguments apply to establish uniqueness of  $\psi^*$  on the set  $\mathcal{Y}_0$ .

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

$$(22) \quad (\psi_n, \psi_n^*) \xrightarrow{n \in \mathbb{N}_j} (\varphi_j, \varphi_j^*) \in \Phi_0(\mathcal{U}, \mathcal{Y}), \quad \text{uniformly on } \mathcal{U} \times \mathcal{Y}.$$

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  $(\varphi_j, \varphi_j^*)$  is the solution to the limit Kantorovich problem, and by the uniqueness established above we have that

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

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

$$\begin{aligned} Q(u) &= \nabla\psi(u) = \nabla\varphi_j(u), \\ R(u) &= \nabla\psi^*(u) = \nabla\varphi_j^*(u). \end{aligned}$$

By (22) and condition (C), we can invoke Theorem A.1 to conclude that  $Q_n \rightarrow Q$  uniformly on compact subsets of  $\mathcal{U}_0$ , and  $R_n \rightarrow R$  uniformly on compact subsets of  $\mathcal{Y}_0$ .

**B.6. Proof of Lemma 3.1.** The proof is a variant of standard arguments, for example, those given in [48], proof of Theorem 2.1, so is relegated to the supplementary material, Appendix [8].

**B.7. 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 [56], Theorem A.1 and Corollary A.1.

That theorem, specialized to our context, reads as follows: Let  $\mathbb{D}$  and  $\mathbb{E}$  be normed spaces and let  $x \in \mathbb{D}$ . Let  $\mathbb{D}_n \subseteq \mathbb{D}$  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  $x_n \rightarrow x$  along a subsequence, we have that  $g_n(x_n) \rightarrow g_0(x)$ , along the same subsequence. Then, for arbitrary (i.e., possibly non-measurable) maps  $X_n : \Omega \mapsto \mathbb{D}_n$  such that  $X_n \rightarrow_{\mathbb{P}^*} x$ , we have that  $g_n(X_n) \rightarrow_{\mathbb{P}^*} g_0(x)$ .

In our case,  $X_n = (\hat{P}_n, \hat{F}_n)$  is a stochastic element of  $\mathbb{D}$ , viewed as an arbitrary map from  $\Omega$  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 (see notation section)  $\|\cdot\|_{\mathbb{D}}$ , where

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

Moreover,  $X_n \rightarrow_{\mathbb{P}^*} x$  with respect to this norm, that is,

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

Then  $g_n(X_n) := (\hat{Q}_n, \hat{R}_n)$  and  $g(x) := (Q, R)$  are viewed as elements of the space  $\mathbb{E} = \ell^\infty(K \times K', \mathbb{R}^d \times \mathbb{R}^d)$  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  $\|x_n - x\|_{\mathbb{D}} \rightarrow 0$  along a subsequence, then  $\|g_n(x_n) - g(x)\|_{\mathbb{E}} \rightarrow 0$  along the same subsequence, as established by Theorem A.1. Hence, conclude that  $g_n(X_n) \rightarrow_{\mathbb{P}^*} g(x)$ .

The second claim follows by the extended continuous mapping theorem and Corollary A.1.

**B.8. Proof of Corollaries 3.1, 3.2 and 3.3.** Corollaries 3.1 and 3.2 follow by Theorem 3.1 and the extended continuous mapping theorem; Corollary 3.3 follows by Theorem 3.1, the extended continuous mapping theorem and Corollary A.2.

**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 inspiring remarks and discussions. We also thank Denis Chetverikov, Yaroslav Mukhin and Robert Serfling for insightful comments, and express our gratitude to Miroslav Šiman for sharing his codes, and to an Associate Editor and two anonymous referees for helpful reports that helped improving the paper.

## SUPPLEMENTARY MATERIAL

**Supplement to “Monge–Kantorovich depth, quantiles, ranks and signs”** (DOI: [10.1214/16-AOS1450SUPP](https://doi.org/10.1214/16-AOS1450SUPP); .zip). In the online supplement [8], we provide a proof of Lemma 3.1.

## REFERENCES

- [1] AGOSTINELLI, C. and ROMANAZZI, M. (2011). Local depth. *J. Statist. Plann. Inference* **141** 817–830. [MR2732952](#)
- [2] AURENHAMMER, F., HOFFMANN, F. and ARONOV, B. (1998). Minkowski-type theorems and mean-square clustering. *Algorithmica* **20** 61–76.
- [3] BENAMOU, J.-D. and BRENIER, Y. (2000). A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numer. Math.* **84** 375–393. [MR1738163](#)
- [4] BRENIER, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44** 375–417. [MR1100809](#)
- [5] CARLIER, G., CHERNOZHUKOV, V. and GALICHON, A. (2016). Vector quantile regression: An optimal transport approach. *Ann. Statist.* **44** 1165–1192. [MR3485957](#)
- [6] CHAUDHURI, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* **91** 862–872. [MR1395753](#)
- [7] CHEN, Y., DANG, X., PENG, H. and BART, H. L. J. (2009). Outlier detection with the kernelized spatial depth function. *IEEE Trans. Pattern Anal. Mach. Intell.* **31** 288–305.
- [8] CHERNOZHUKOV, V., GALICHON, A., HALLIN, M. and HENRY, M. (2016). Supplement to “Monge–Kantorovich depth, quantiles, ranks and signs.” DOI:[10.1214/16-AOS1450SUPP](https://doi.org/10.1214/16-AOS1450SUPP).
- [9] CUESTA-ALBERTOS, J. A. and NIETO-REYES, A. (2008). The random Tukey depth. *Comput. Statist. Data Anal.* **52** 4979–4988. [MR2526207](#)
- [10] DECURNINGE, A. (2014). Multivariate quantiles and multivariate  $L$ -moments. Preprint. Available at [arXiv:1409.6013](https://arxiv.org/abs/1409.6013).
- [11] DENEEN, L. and SHUTE, G. (1988). Polygonizations of point sets in the plane. *Discrete Comput. Geom.* **3** 77–87. [MR0918181](#)
- [12] DOKSUM, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Statist.* **2** 267–277. [MR0356350](#)
- [13] DOKSUM, K. A. and SIEVERS, G. L. (1976). Plotting with confidence: Graphical comparisons of two populations. *Biometrika* **63** 421–434. [MR0443210](#)
- [14] DONOHO, D. L. (1982). Breakdown properties of multivariate location estimators. Qualifying paper, Harvard Univ.
- [15] DONOHO, D. L. and GASKO, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827. [MR1193313](#)
- [16] DUTTA, S., GHOSH, A. K. and CHAUDHURI, P. (2011). Some intriguing properties of Tukey’s halfspace depth. *Bernoulli* **17** 1420–1434. [MR2854779](#)

- [17] EDELSBRUNNER, H., KIRKPATRICK, D. G. and SEIDEL, R. (1983). On the shape of a set of points in the plane. *IEEE Trans. Inform. Theory* **29** 551–559. [MR0713690](#)
- [18] EKELAND, I., GALICHON, A. and HENRY, M. (2012). Comonotonic measures of multivariate risks. *Math. Finance* **22** 109–132. [MR2881882](#)
- [19] GALICHON, A. and HENRY, M. (2012). Dual theory of choice with multivariate risks. *J. Econom. Theory* **147** 1501–1516. [MR2942620](#)
- [20] GHOSH, A. K. and CHAUDHURI, P. (2005). On maximum depth and related classifiers. *Scand. J. Statist.* **32** 327–350. [MR2188677](#)
- [21] GRÜNBAUM, B. (1994). Hamiltonian polygons and polyhedra. *Geombinatorics* **3** 83–89. [MR1326479](#)
- [22] HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic Press, New York. [MR0229351](#)
- [23] HALLIN, M. and PAINDAVEINE, D. (2002). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.* **30** 1103–1133. [MR1926170](#)
- [24] HALLIN, M. and PAINDAVEINE, D. (2004). Rank-based optimal tests of the adequacy of an elliptic VARMA model. *Ann. Statist.* **32** 2642–2678. [MR2153998](#)
- [25] HALLIN, M. and PAINDAVEINE, D. (2005). Affine-invariant aligned rank tests for the multivariate general linear model with VARMA errors. *J. Multivariate Anal.* **93** 122–163. [MR2119768](#)
- [26] HALLIN, M. and PAINDAVEINE, D. (2006). Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity. *Ann. Statist.* **34** 2707–2756. [MR2329465](#)
- [27] HALLIN, M. and PAINDAVEINE, D. (2008). Optimal rank-based tests for homogeneity of scatter. *Ann. Statist.* **36** 1261–1298. [MR2418657](#)
- [28] 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](#)
- [29] HALLIN, M. and WERKER, B. J. M. (2003). Semi-parametric efficiency, distribution-freeness and invariance. *Bernoulli* **9** 137–165. [MR1963675](#)
- [30] HASSAIRI, A. and REGAIEG, O. (2008). On the Tukey depth of a continuous probability distribution. *Statist. Probab. Lett.* **78** 2308–2313. [MR2462666](#)
- [31] HLUBINKA, D., KOTÍK, L. and VENCÁLEK, O. (2010). Weighted halfspace depth. *Kybernetika (Prague)* **46** 125–148. [MR2666899](#)
- [32] JUDD, K. L. (1998). *Numerical Methods in Economics*. MIT Press, Cambridge, MA. [MR1646331](#)
- [33] KOENKER, R. and BASSETT, G. JR. (1978). Regression quantiles. *Econometrica* **46** 33–50. [MR0474644](#)
- [34] KOLTCHINSKII, V. and DUDLEY, R. (1992). On spatial quantiles. Unpublished manuscript.
- [35] KOSHEVOY, G. and MOSLER, K. (1997). Zonoid trimming for multivariate distributions. *Ann. Statist.* **25** 1998–2017. [MR1474078](#)
- [36] KOSHEVOY, G. A. (2002). The Tukey depth characterizes the atomic measure. *J. Multivariate Anal.* **83** 360–364. [MR1945958](#)
- [37] LIU, R. Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405–414. [MR1041400](#)
- [38] LIU, R. Y. (1992). Data depth and multivariate rank tests. In  *$L_1$ -Statistical Analysis and Related Methods (Neuchâtel, 1992)* (Y. Dodge, ed.) 279–294. North-Holland, Amsterdam. [MR1214839](#)
- [39] LIU, R. Y., PARELIUS, J. M. and SINGH, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Ann. Statist.* **27** 783–858. [MR1724033](#)

- [40] LIU, R. Y. and SINGH, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.* **88** 252–260. [MR1212489](#)
- [41] MAHALANOBIS, P. C. (1936). On the generalized distance in statistics. *Proc. Natl. Acad. Sci., India* **12** 49–55.
- [42] MCCANN, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.* **80** 309–323. [MR1369395](#)
- [43] MOSLER, K. (2002). *Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach. Lecture Notes in Statistics* **165**. Springer, Berlin. [MR1913862](#)
- [44] MÖTTÖNEN, J. and OJA, H. (1995). Multivariate sign and rank methods. *J. Nonparametr. Stat.* **5** 201–213.
- [45] OJA, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.* **1** 327–332. [MR0721446](#)
- [46] PAINDAVEINE, D. and ŠIMAN, M. (2012). Computing multiple-output regression quantile regions. *Comput. Statist. Data Anal.* **56** 840–853. [MR2888729](#)
- [47] PAINDAVEINE, D. and VAN BEVER, G. (2013). From depth to local depth. *J. Amer. Statist. Assoc.* **108** 1105–1119.
- [48] POLITIS, D., ROMANO, J. P. and WOLF, M. (1999). Weak convergence of dependent empirical measures with application to subsampling in function spaces. *J. Statist. Plann. Inference* **79** 179–190. [MR1700758](#)
- [49] ROCKAFELLAR, R. T. (1997). *Convex Analysis. Princeton Landmarks in Mathematics*. Princeton Univ. Press, Princeton, NJ. [MR1451876](#)
- [50] ROCKAFELLAR, R. T. and WETS, R. J.-B. (1998). *Variational Analysis. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **317**. Springer, Berlin. [MR1491362](#)
- [51] SERFLING, R. (2002). Quantile functions for multivariate analysis: Approaches and applications. *Stat. Neerl.* **56** 214–232. [MR1916321](#)
- [52] SINGH, K. (1991). Majority depth. Unpublished manuscript.
- [53] STAHEL, W. (1981). Robuste Schätzungen: Infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen. Ph.D. thesis, Univ. Zürich.
- [54] TUKEY, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974)*, Vol. 2 523–531. Canad. Math. Congress, Montreal, Que. [MR0426989](#)
- [55] VAN ZWET, W. R. (1964). *Convex Transformations of Random Variables*. Mathematisch Centrum, Amsterdam.
- [56] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence*. Springer, New York.
- [57] VARDI, Y. and ZHANG, C.-H. (2000). The multivariate  $L_1$ -median and associated data depth. *Proc. Natl. Acad. Sci. USA* **97** 1423–1426 (electronic). [MR1740461](#)
- [58] VILLANI, C. (2003). *Topics in Optimal Transportation. Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. [MR1964483](#)
- [59] VILLANI, C. (2008). Stability of a 4th-order curvature condition arising in optimal transport theory. *J. Funct. Anal.* **255** 2683–2708. [MR2473273](#)
- [60] VILLANI, C. (2009). *Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **338**. Springer, Berlin. [MR2459454](#)
- [61] ZUO, Y. (2003). Projection-based depth functions and associated medians. *Ann. Statist.* **31** 1460–1490. [MR2012822](#)
- [62] ZUO, Y. and SERFLING, R. (2000). General notions of statistical depth function. *Ann. Statist.* **28** 461–482. [MR1790005](#)

V. CHERNOZHUKOV  
DEPARTMENT OF ECONOMICS  
AND CENTER FOR STATISTICS  
MASSACHUSETTS INSTITUTE OF TECHNOLOGY  
CAMBRIDGE, MASSACHUSETTS 02139  
USA  
E-MAIL: [vchern@mit.edu](mailto:vchern@mit.edu)

M. HALLIN  
ECARES  
UNIVERSITÉ LIBRE DE BRUXELLES  
CP 114 BRUSSELS  
BELGIUM  
E-MAIL: [mhallin@ulb.ac.be](mailto:mhallin@ulb.ac.be)

A. GALICHON  
ECONOMICS DEPARTMENT  
AND COURANT INSTITUTE  
NEW YORK UNIVERSITY  
NEW YORK, NEW YORK 10012  
USA  
AND  
SCIENCES PO, ECONOMICS DEPARTMENT  
75007 PARIS  
FRANCE  
E-MAIL: [ag133@nyu.edu](mailto:ag133@nyu.edu)

M. HENRY  
DEPARTMENT OF ECONOMICS  
PENNSYLVANIA STATE UNIVERSITY  
UNIVERSITY PARK, PENNSYLVANIA 16802  
USA  
E-MAIL: [marc.henry@psu.edu](mailto:marc.henry@psu.edu)