Distribution and Quantile Functions, Ranks, and Signs in dimension d: a measure transportation approach

Marc Hallin, Eustasio del Barrio, Juan Cuesta Albertos, and Carlos Matrán

e-mail: mhallin@ulb.ac.be
ECARES and Département de Mathématique
Université libre de Bruxelles, Brussels, Belgium

e-mail: tasio@eio.uva.es
Departamento de Estadística e Investigaciòn Operativa
Facultad de Ciencias, Universidad de Valladolid, Spain

e-mail: juan.cuesta@unican.es
Departamento de Matemáticas
Facultad de Ciencias, Universidad de Cantabria, Santander, Spain

e-mail: carlos.matran@uva.es
Departamento de Estadística e Investigación Operativa
Facultad de Ciencias, Universidad de Valladolid, Spain

Abstract: Unlike the real line, the real space $\mathbb{R}^d$, for $d \geq 2$, is not canonically ordered. As a consequence, such fundamental univariate concepts as quantile and distribution functions, and their empirical counterparts, involving ranks and signs, do not canonically extend to the multivariate context. Palliating that lack of a canonical ordering has been an open problem for more than half a century, generating an abundant literature and motivating, among others, the development of statistical depth and copula-based methods. We show that, unlike the many definitions proposed in the literature, the measure transportation-based ranks introduced in Chernozhukov et al. (2017) enjoy all the properties that make univariate ranks a successful tool for semiparametric inference. Related with those ranks, we propose a new *center-outward* definition of multivariate distribution and quantile functions, along with their empirical counterparts, for which we establish a Glivenko-Cantelli result. Our approach is based on McCann (1995) and our results, unlike those of Chernozhukov et al. (2017), do not require any moment assumptions. The resulting ranks and signs are shown to be strictly distribution-free and essentially maximal ancillary in the sense of Basu (1959) which, in semiparametric models involving noise with unspecified density, can be interpreted as a finite-sample form of semiparametric efficiency. Although constituting a sufficient summary of the sample, empirical center-outward distribution functions are defined at observed values only. A continuous extension to the entire d-dimensional space, yielding smooth empirical quantile contours and sign curves while preserving the essential monotonicity and Glivenko-Cantelli features, is provided. A numerical study of the resulting empirical quantile contours is conducted.

MSC 2010 subject classifications: Primary 62G30; secondary 62B05.
Keywords and phrases: Multivariate distribution function; multivariate quantiles, multivariate ranks; multivariate signs; Glivenko-Cantelli theorem; Basu theorem; distribution-freeness; cyclical monotonicity.

1

1. Introduction

Unlike the real line, the real space $\mathbb{R}^d$, for $d \geq 2$, is not canonically ordered. As a consequence, such fundamental concepts as quantile and distribution functions, which are strongly related to the ordering of the observation space, and their empirical counterparts—ranks and empirical quantiles—playing, in dimension $d = 1$, a fundamental role in statistical inference, do not canonically extend to dimension $d \geq 2$.

Of course, a classical concept of distribution function—the familiar one, based on marginal orderings—does exist. That concept, from a probabilistic point of view, does the job of characterizing the underlying distribution. However, the corresponding quantile function does not mean much (see, e.g., Genest and Rivest (2001)), and the corresponding empirical versions (related to their population counterparts via a Glivenko-Cantelli result) do not possess any of the properties that make them successful inferential tools in dimension $d = 1$.

That observation about traditional multivariate distribution functions is not new: palliating the lack of a "natural" ordering of $\mathbb{R}^d$—hence, defining statistically sound concepts of distribution and quantile functions—has been an open problem for more than half a century, generating an abundant literature that includes, among others, the theory of copulas and the theory of statistical depth.

A number of most ingenious solutions have been proposed, each of them extending some chosen features of the well-understood univariate concepts, with which they coincide for $d = 1$. Coinciding, for $d = 1$, with the univariate concepts obviously is important, but hardly sufficient for qualifying as a statistically pertinent multivariate extension. For statisticians, distribution and quantile functions are not just probabilistic notions: above all, their empirical versions (empirical quantiles and ranks) constitute fundamental tools for inference. A multivariate extension yielding quantiles and ranks that do not enjoy, in dimension $d \geq 2$, the properties that make traditional ranks natural and successful tools for inference for $d = 1$ is not a statistically sound extension.

Those inferential concerns are at the heart of the approach adopted here.

1.1. Ranks and rank-based inference

To facilitate the exposition, let us focus on ranks and their role in testing problems. Univariate rank-based methods naturally enter the picture in the context of semiparametric statistical models under which the distribution $\mathrm{P}_{\boldsymbol{\theta},f}^{(n)}$ of some real-valued observation $\mathbf{X} = (X_1, \ldots, X_n)'$, besides a finite-dimensional parameter of interest $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, also depends on the unspecified density $f \in \mathcal{F}_1$ ($\mathcal{F}_1$ the family of Lebesgue densities over $\mathbb{R}$) of some unobserved univariate noise $Z_i(\boldsymbol{\theta})$, say. More precisely, $\mathbf{X} \sim \mathrm{P}_{\boldsymbol{\theta},f}^{(n)}$ iff the $\boldsymbol{\theta}$-*residuals* $Z_1(\boldsymbol{\theta}), \ldots, Z_n(\boldsymbol{\theta}) =: \mathbf{Z}^{(n)}(\boldsymbol{\theta})$ are i.i.d.[1] with density f. In such models—call them i.i.d. noise models[2]—testing the

---

[1] Although i.i.d.-*ness* can be relaxed into *exchangeability*, we are sticking to the former.

[2] Typical examples are linear models, with $Z_i(\boldsymbol{\theta}) = X_i - \mathbf{c}_i'\boldsymbol{\theta}$ ($\mathbf{c}_i$ a q-vector of covariates and $\boldsymbol{\theta} \in \mathbb{R}^q$), or first-order autoregressive models, with $Z_i(\theta) = X_i - \theta X_{i-1}$ (where i denotes time and $\theta \in (-1, 1)$; see, e.g., Hallin and Werker (1998)), etc.

null hypothesis $H_0^{(n)}: \boldsymbol{\theta} = \boldsymbol{\theta}_0$ (that is, $\mathrm{P}_{\boldsymbol{\theta},f}^{(n)} \in \mathcal{P}_{\boldsymbol{\theta}_0}^{(n)} := \{\mathrm{P}_{\boldsymbol{\theta}_0,f}^{(n)} | f \in \mathcal{F}_1\}$) reduces to the problem of testing that $Z_1(\boldsymbol{\theta}_0), \ldots, Z_n(\boldsymbol{\theta}_0)$ are i.i.d. with unspecified density $f \in \mathcal{F}_1$. Invariance arguments suggest tests based on the ranks $\mathbf{R}^{(n)}(\boldsymbol{\theta}_0)$ of the residuals $\mathbf{Z}^{(n)}(\boldsymbol{\theta}_0)$[3]; such tests are *distribution-free* under $H_0^{(n)}$.

Distribution-freeness (DF) is often considered as the trademark and main virtue of (univariate) ranks; it guarantees the validity and similarity of rank-based tests of $H_0^{(n)}$. Distribution-freeness alone is not sufficient, though, for explaining the success of rank tests: other classes of distribution-free methods indeed can be constructed, such as sign or runs tests, that do not perform as well as the rank-based ones. The reason is that, unlike the ranks, they do not fully exploit the information available once the nuisance (the unknown f) has been controlled for via some minimal sufficient statistic. That feature of ranks originates in the fact that

(DF$^+$) (essential maximal ancillarity) *the sub-σ-field generated by the residual ranks $\mathbf{R}^{(n)}(\boldsymbol{\theta})$ is* essentially maximal ancillary *(hence distribution-free) for $\mathcal{P}_{\boldsymbol{\theta}}^{(n)}$ in the sense of Basu (1959) (see, e.g., Example 7 in Lehmann and Scholz (1992)).*

while the sub-σ-field generated by the residual order statistic $\mathbf{Z}_{(\cdot)}^{(n)}(\boldsymbol{\theta})$ is *minimal sufficient* and *complete* (still for $\mathcal{P}_{\boldsymbol{\theta}}^{(n)}$).

In families satisfying the condition (Koehn and Thomas 1975) of non-existence of a *splitting set*—which is the case here whenever f ranges over $\mathcal{F}_1$—Theorems 1 and 2 in Basu (1955) imply that essential maximal ancillarity is equivalent to "essential maximal independence with respect to the complete (hence minimal) sufficient statistic $\mathbf{Z}_{(\cdot)}^{(n)}(\boldsymbol{\theta})$."[4] Intuitively, thus, and leaving aside the required mathematical precautions, the order statistic $\mathbf{Z}_{(\cdot)}^{(n)}(\boldsymbol{\theta})$, being minimal sufficient for $\mathcal{P}_{\boldsymbol{\theta}}^{(n)}$, is carrying all the information about the nuisance f and nothing but that information, while the ranks, being (essentially) "maximal independent of $\mathbf{Z}_{(\cdot)}^{(n)}(\boldsymbol{\theta})$," are carrying whatever information is left for $\boldsymbol{\theta}$. This can be interpreted as a finite-sample form of semiparametric efficiency[5]

In the same vein, it also has been shown (Hallin and Werker 2003) that, under appropriate regularity conditions, univariate ranks preserve semiparametric efficiency in models where that concept makes sense:

(HW) (preservation of semiparametric efficiency) *the semiparametric efficiency bound at arbitrary $(\boldsymbol{\theta}, f)$ can be reached, under $\mathrm{P}_{\boldsymbol{\theta},f}^{(n)}$, via rank-based procedures (tests that are measurable functions of the ranks of $\boldsymbol{\theta}$-residuals $Z_i(\boldsymbol{\theta})$).*

The latter property, contrary to (DF) and (DF$^+$), is of a local and asymptotic nature; in Hallin and Werker (2003), it follows from the maximal invariance

---

[3]Those ranks indeed are maximal invariant under the group of continuous monotone increasing transformations of $Z_1(\boldsymbol{\theta}_0), \ldots, Z_n(\boldsymbol{\theta}_0)$; see, e.g., Example 7 in Lehmann and Scholz (1992).

[4]We refer to Appendix D.1 for precise definitions, a more general and stronger version of this property, and a proof.

[5]Semiparametric efficiency indeed is characterized as asymptotic orthogonality, with respect to the central sequences carrying information about parametric perturbations of the nuisance; asymptotic orthogonality here is replaced with finite-sample independence.

property of ranks under a group of order-preserving transformations of $\mathbb{R}^n$ *generating* the fixed-$\boldsymbol{\theta}$ submodel (that is, yielding a unique orbit in the family $\mathcal{P}_{\boldsymbol{\theta}}^{(n)}$ of fixed-$\boldsymbol{\theta}$ model distributions). Being intimately related to the concept of order-preserving transformation, this invariance approach is much more delicate in dimension $d > 1$. For lack of space, we do not investigate it any further here, leaving a formal multivariate extension of (HW) for further research.

Properties (DF$^+$) and (HW), which indicate, roughly, that the order statistic only carries information about the nuisance f while the ranks carry all the information available about $\boldsymbol{\theta}$, are those a statistician definitely would like to see satisfied by any sensible multivariate extension of the concept.

1.2. *Multivariate ranks and the ordering of* $\mathbb{R}^d$, $d \geq 2$

The problem of ordering $\mathbb{R}^d$ for $d \geq 2$, thus defining multivariate concepts of ranks, signs, empirical distribution functions and quantiles, is not new, and has a rather long history in statistics. Many concepts have been proposed in the literature, a complete list of which cannot be given here. Focusing again on ranks, four types of *multivariate ranks*, essentially, can be found:

(a) *Componentwise ranks.* The idea of componentwise ranks goes back as far as Hodges (1955), Bickel (1965) or Puri and Sen (1966, 1967, 1969). It culminates in the monograph by Puri and Sen (1971), where inference procedures based on componentwise ranks are proposed, basically, for all classical problems of multivariate analysis. Time-series testing methods based on the same ranks have been considered in Hallin, Ingenbleek, and Puri (1989). That strand of literature is still alive: see Chaudhuri and Sengupta (1993), Segers, van den Akker, and Werker (2015), ... to quote only a very few. Componentwise ranks actually are intimately related to copula transforms, of which they constitute the empirical version: rather than solving the tricky problem of ordering $\mathbb{R}^d$, they bypass it by considering d univariate marginal rankings. As a consequence, they crucially depend on the choice of a coordinate system. Unless the underlying distribution has independent components (Nordhausen et al. 2009, Ilmonen and Paindaveine 2011, Hallin and Mehta 2015) coinciding with the chosen coordinates, componentwise ranks in general are not even asymptotically distribution-free: neither (DF) nor (DF$^+$) hold.

(b) *Spatial ranks and signs.* This class of multivariate ranks (Möttönen and Oja 1995; Möttönen et al. 1997; Chaudhuri 1996; Koltchinskii 1997; Oja and Randles 2004, Oja 2010, and many others) includes several very ingenious, elegant, and appealing concepts. Similar ideas also have been developed by Choi and Marden (1997) and, more recently, in high dimension, by Biswas, Mukhopadhyay, and Ghosh (2014) and Chakraborthy and Chaudhuri (1996, 2014, 2017). We refer to Marden (1999), Oja (1999) or Oja (2010) for a systematic exposition and exhaustive list of references. All those concepts are extending the traditional univariate ones but none of them enjoys (DF)[6], let alone (DF$^+$).

---

[6]Biswas et al. (2014) is an exception, but fails on (DF$^+$)

(c) *Depth-based ranks.* Those ranks have been considered in Liu (1992), Liu and Singh (1993), He and Wang (1997), Zuo and He (2006), Zuo and Serfling (2000), among others; see Serfling (2002) for a general introduction on statistical depth, Hallin et al. (2010) for the related concept of quantile, Lòpez-Pintado and Romo (2012) for functional extensions, Zuo (2018) for a state-of-the art survey in a regression context. Depth-based ranks, in general, are distribution-free but fail to satisfy (DF$^+$).

(d) *Mahalanobis ranks and signs/interdirections.* When considered jointly with interdirections (Randles 1989), lift interdirections (Oja and Paindaveine 2005), Tyler angles or Mahalanobis signs (Hallin and Paindaveine 2002a, c), Mahalanobis ranks do satisfy (DF$^+$), but in elliptical models only—when f is limited to the family of elliptical densities. There, they have been used, quite successfully, in a variety of multivariate models, including one-sample location (Hallin and Paindaveine 2002a), k-sample location (Um and Randles 1998), serial dependence (Hallin and Paindaveine 2002b), linear models with VARMA errors (Hallin and Paindaveine 2004a, 2005a, 2006a), VAR order identification (Hallin and Paindaveine 2004b), shape (Hallin and Paindaveine 2006b; Hallin, Oja, and Paindaveine 2006), homogeneity of scatter (Hallin and Paindaveine 2008), principal and common principal components (Hallin, Paindaveine, and Verdebout 2010, 2013, 2014). Unfortunately, the tests developed in those references cease to be valid, and the related R-estimators no longer are root-n consistent, under non-elliptical densities.

None of those multivariate rank concepts, thus, enjoys distribution-freeness *and* (DF$^+$)—except, but only over the class of elliptically symmetric distributions, for the Mahalanobis ranks and signs. A few other concepts have been proposed as well, related to *cone orderings* (Belloni and Winkler 2011; Hamel and Kostner 2018), which require some subjective (or problem-specific) preliminary choices, and similarly fail to achieve distribution-freeness, hence (DF$^+$).

The lack, for $d \geq 2$, of a canonical ordering of $\mathbb{R}^d$ places an essential difference between dimensions $d = 1$ and $d \geq 2$. Whereas the same "exogenous" left-to-right ordering of $\mathbb{R}$ applies both in population and in the sample, pertinent orderings of $\mathbb{R}^d$ are bound to be "endogenous", that is, distribution-specific in populations, and data-driven (hence, random) in samples. This is the case for the concepts developed under (b)-(d) above; it also holds for the concept we are proposing in this paper. Each distribution, each sample, thus, is to produce its own ordering, inducing quantile and distribution functions, and classes of order-preserving transformations. As a result, datasets, at best, can be expected to produce, via adequate concepts of multivariate ranks and signs, consistent empirical versions of the underlying population ordering. That consistency typically takes the form of a Glivenko-Cantelli (GC) result connecting an empirical *center-outward distribution function* to its population version. A quintessential feature of Glivenko-Cantelli is its insensitivity to continuous order-preserving transformations of the data. That feature is not compatible with moment assumptions, since the existence of moments is not preserved under such transformations. Moment assumptions (as in Boeckel et al. (2018) or Chernozhukov et al. (2017) where (weak) consistency is established under compactly supported

distributions), therefore, are somewhat inappropriate in this context.

No ordering of $\mathbb{R}^d$ for $d \geq 2$ moreover can be expected to be of the one-sided "left-to-right" type, since "left" and "right" do not make sense anymore. A depth-type center-outward ordering is by far more sensible. All this calls for revisiting the traditional univariate concepts from a center-outward perspective, while disentangling (since they are to be based on distinct orderings) the population concepts from their sample counterparts.

1.3. Outline of the paper

This paper consists of a main text and an online appendix. Except for the proofs, the main text is self-contained and the reader familiar with measure transportation and statistical decision can skip most of the appendix. For those who are less familiar with those topics, however, we recommend the following plan for fruitful reading. After the introduction (Section 1), one may like to go to Appendix A.1 for a brief and elementary account of some classical facts in measure transportation, then to Appendix A.2 for a short review of the (scarce) literature on relation of that theory to multivariate ranks and quantiles. Appendix B is describing how the traditional univariate case, where the concepts of distribution and quantile functions, ranks, and signs are familiar, naturally enters the realm of measure trasportation once the usual distribution function F is replaced by the so-called center-outward one $2F - 1$. The paper then really starts with Section 2, where the main concepts—center-outward distribution and quantile functions, ranks, signs, quantile contours and quantile regions—are defined and their main properties—regularity of distribution and quantile functions, nestedness and connectedness of quantile regions, distribution-freeness of ranks and signs, their maximal ancillarity property and their Glivenko-Cantelli asymptotics—are stated. Proofs are given in Appendices D and D.1 and the relation, under ellipticity, to Mahalanobis ranks and signs is discussed in Appendix C. Up to that point, empirical distribution and quantile functions are defined at the observations only. Section 3 shows how to extend them into smooth functions defined over the entire space $\mathbb{R}^d$ while preserving their gradient of convex function nature, without which they no longer would qualify as distribution and quantile functions. This smooth extension is shown (Proposition 3.3) to satisfy an extended Glivenko-Cantelli property; proofs are concentrated in Appendix F. The tools we are using throughout are exploiting the concept of *cyclical monotonicity* and the approach initiated by McCann (1995).[7] Section 4 provides some numerical results. The algorithms we are using can handle samples of size as large as $n = 20000$ in dimension 2 (the complexity of the algorithms in $\mathbb{R}^d$ only depend on n, not on d); simulations demonstrate the power of empirical center-outward quantile functions as descriptive tools. Further numerical results, and a comparison with Tukey depth are given in Appendix H. Section 5 concludes with a discussion of some perspectives for further research.

---

[7]This fact is emphasized by a shift in the terminology: as our approach is no longer based on Monge-Kantorovich optimization ideas, we consistently adopt the terminology *center-outward* ranks and signs instead of *Monge-Kantorovich* ranks and signs.

1.4. Notation

Throughout, let μ_d stand for the Lebesgue measure over $\mathbb{R}^d$ equipped with its Borel σ-field $\mathcal{B}_d$. Denote by $\mathcal{P}_d$ the family of Lebesgue-absolutely continuous distributions over $(\mathbb{R}^d, \mathcal{B}_d)$, by $\mathcal{F}_d := \{f := d\mathrm{P}/d\mu_d, \mathrm{P} \in \mathcal{P}^d\}$ the corresponding family of densities, by $\mathcal{B}_d^n$ the n-fold product $\mathcal{B}_d \times \ldots \times \mathcal{B}_d$, by $\mathrm{P}^{(n)}$ or $\mathrm{P}_f^{(n)}$ the distribution of an i.i.d. n-tuple with marginals $\mathrm{P} = \mathrm{P}_f \in \mathcal{P}_d$, by $\mathcal{P}_d^{(n)}$ the corresponding collection $\{\mathrm{P}_f^{(n)}, f \in \mathcal{F}_d\}$; $\mathcal{P}_d^{(n)}$-a.s. means $\mathrm{P}^{(n)}$-a.s. for all $\mathrm{P} \in \mathcal{P}_d^{(n)}$. Write $\overline{\mathrm{spt}}(\mathrm{P})$ for the support of P, spt(P) for its interior, $\mathcal{S}_{d-1}$, $\mathbb{S}_d$, and $\overline{\mathbb{S}}_d$ for the unit sphere, the open, and the closed unit ball in $\mathbb{R}^d$, respectively.

2. Distribution and quantile functions, ranks and signs in $\mathbb{R}^d$

As announced in the introduction, our definitions of center-outward distribution and quantile functions are rooted in the main result of McCann (1985). Those definitions in Hallin (2017) are given under the assumption that P has a nonvanishing density with support $\mathbb{R}^d$. Under that assumption, one safely can define the center-outward distribution function as the unique gradient of a convex function $\nabla\phi$ pushing P forward to the uniform distribution over the unit ball. That gradient, moreover, is a homeomorphism between $\mathbb{R}^d \backslash \nabla\phi^{-1}(\{\mathbf{0}\})$ and the punctured unit ball $\mathbb{S}_d \backslash \{\mathbf{0}\}$ (Figalli 2019) and its inverse naturally qualifies as a quantile function—a very simple and intuitively clear characterization.

Things are more delicate when the support of P is a strict subset of $\mathbb{R}^d$, as uniqueness of $\nabla\phi$ then only holds P-a.s., and requires the slightly more elaborate definitions developed here. The two approaches, however, coincide in case P has a non vanishing density over $\mathbb{R}^d$.

2.1. Center-outward distribution and quantile functions in $\mathbb{R}^d$

Recall that a convex function ψ from $\mathbb{R}^d$ to $\mathbb{R} \cup \{\infty\}$ (a) is continuous on the interior of $\mathrm{dom}(\psi) := \{\mathbf{x} : \psi(\mathbf{x}) < \infty\}$ and (b) Lebesgue-a.e. differentiable, with gradient $\nabla\psi$, on $\mathrm{dom}(\psi)$. By abuse of language and notation, call gradient and denote as $\nabla\psi$ any function coinciding μ_d-a.e. with that gradient. A statement of McCann's main result adapted to our needs is the following. [8]

Theorem 2.1 (McCann 1985) *Let* P_1 *and* P_2 *denote two distributions in* $\mathcal{P}_d$. *Then, (i) the class of functions*

$$\nabla\Psi_{\mathrm{P}_1;\mathrm{P}_2} := \big\{\nabla\psi \,\big|\, \psi : \mathbb{R}^d \to \mathbb{R} \quad \text{convex, lower semi-continuous, and} \qquad (2.1)$$
$$\text{such that } \nabla\psi\#\mathrm{P}_1 = \mathrm{P}_2\big\}$$

is not empty; (ii) if $\nabla\psi'$ *and* $\nabla\psi''$ *are two elements of* $\nabla\Psi_{\mathrm{P}_1;\mathrm{P}_2}$, *they coincide* P_1-*a.s.;* [9] *(iii) if* P_1 *and* P_2 *have finite moments of order two, any element of* $\nabla\Psi_{\mathrm{P}_1;\mathrm{P}_2}$ *is an optimal quadratic transport pushing* P_1 *forward to* P_2.

---

[8]Below we are borrowing from the measure transportation literature the convenient notation $T\#\mathrm{P}_1 = \mathrm{P}_2$ for the distribution P_2 of $T(X)$ under $X \sim \mathrm{P}$—we say that T is *pushing forward* P_1 to P_2.

[9]That is, $\mathrm{P}_1(\{\mathbf{x} : \nabla\psi'(\mathbf{x}) \neq \nabla\psi''(\mathbf{x})\}) = 0$; in particular, $\nabla\psi_1(\mathbf{x}) = \nabla\psi_2(\mathbf{x})$ Lebesgue-a.e. for $\mathbf{x} \in \mathrm{spt}(\mathrm{P}_1)$.

Although not mentioned in McCann's main result (p. 310 of McCann (1995)), lower semi-continuity in (2.1) can be imposed without loss of generality (this follows, for instance, from his proof of uniqueness on p. 318).

Denoting by U_d the *spherical uniform* distribution over $\mathbb{S}_d$,[10] consider Theorem 2.1 and (2.1) for $\mathrm{P}_1 = \mathrm{U}_d$ and $\mathrm{P}_2 = \mathrm{P} \in \mathcal{P}_d$. Since the support of U_d is $\overline{\mathbb{S}}_d$ (which is convex and compact), ψ is uniquely determined over $\mathbb{S}_d$ if we impose, without loss of generality, $\psi(\mathbf{0}) = 0$.[11] Outside $\mathbb{S}_d$ (that is, on a set with U_d-probability zero), let us further impose again

$$\psi(\mathbf{u}) = \infty \ \text{ for } \ \|\mathbf{u}\| > 1 \quad \text{and} \quad \psi(\mathbf{u}) = \liminf_{\mathbb{S}_d \ni \mathbf{v} \to \mathbf{u}} \psi(\mathbf{v}) \ \text{ for } \ \|\mathbf{u}\| = \mathbf{1}. \quad (2.2)$$

The domain of ψ is $\mathrm{dom}(\psi) := \{\mathbf{u}|\psi(\mathbf{u}) < \infty\} = \overline{\mathbb{S}}_d$. A convex function is differentiable a.e. in the interior of its domain. Hence, the gradient $\nabla\psi$ of ψ satisfying (2.2) exists, is unique a.e. in $\mathbb{S}_d$, and still belongs to $\nabla\Psi_{\mathrm{U}_d;\mathrm{P}}$.

Inspired by the univariate case as described in Section B.3, we propose the following definitions of the center-outward quantile function of $\mathrm{P} \in \mathcal{P}_d$.

Definition 2.1 Call *center-outward quantile function* $\mathbf{Q}_\pm$ of $\mathrm{P} \in \mathcal{P}_d$ the a.e. unique element $\nabla\psi \in \nabla\Psi_{\mathrm{U}_d;\mathrm{P}}$ such that ψ satisfies (2.2).

In general, thus, $\mathbf{Q}_\pm$ is a *class* of Lebesgue-a.e. equal functions rather than a function. Each element in that class pushes U_d to P, hence fully characterizes P. Such a.e. uniqueness, in probability and statistics, is not uncommon: densities, conditional expectations, likelihoods, MLEs, ... all are defined up to sets of probability zero. As we shall see, however, strict uniqueness does hold for important families of distributions, for which ψ is everywhere differentiable over $\mathbb{S}_d$.

Next, let us proceed with the definition of the *center-outward distribution function* $\mathbf{F}_\pm$. Consider the Legendre transform

$$\phi(\mathbf{x}) := \psi^*(\mathbf{x}) := \sup_{\mathbf{u} \in \mathbb{S}_s} \left(\langle \mathbf{u}, \mathbf{x} \rangle - \psi(\mathbf{u}) \right) \qquad \mathbf{x} \in \mathbb{R}^d \quad (2.3)$$

of the a.e.-unique convex function ψ (satisfying $\psi(\mathbf{0}) = 0$ and (2.2)) of which $\mathbf{Q}_\pm$ is the gradient. Being the sup of a 1-Lipschitz function, ϕ also is 1-Lipschitz. It follows that ϕ is a.e. differentiable, with $\|\nabla\phi(\mathbf{x})\| \leq 1$, so that (Corollary (A.27) in Figalli (2017)), denoting by $\partial\phi(\mathbf{x})$ the subdifferential of ϕ at $\mathbf{x}$,[12]

$$\partial\phi(\mathbb{R}^d) := \bigcup_{\mathbf{x} \in \mathbb{R}^d} \partial\phi(\mathbf{x}) \subseteq \overline{\mathbb{S}}_d. \quad (2.4)$$

Moreover, since P has a density, Proposition 10 in McCann (1995) implies that

---

[10]Namely, the product of the uniform over the unit sphere $\mathcal{S}_{d-1}$ with a uniform over the unit interval of distances to the origin. While U_d coincides, for $d = 1$, with the Lebesgue-uniform over $\mathbb{S}_1$, this is no longer the case for $d > 1$; we nevertheless still call it *uniform over the unit ball*.

[11]Indeed, two convex functions with a.e. equal gradients on an open convex set are equal up to an additive constant: see Lemma 2.1 in del Barrio and Loubes (2019).

[12]Recall that the subdifferential of ϕ at $\mathbf{x} \in \mathbb{R}$ is the set $\partial\phi(\mathbf{x})$ of all $\mathbf{z} \in \mathbb{R}^d$ such that $\phi(\mathbf{y}) - \phi(\mathbf{x}) \geq \langle \mathbf{z}, \mathbf{y} - \mathbf{x} \rangle$ for all $\mathbf{y}$; ϕ is differentiable at $\mathbf{x}$ iff $\partial\phi(\mathbf{x})$ consists of a single point, $\nabla\phi(\mathbf{x})$.

$$\nabla\psi \circ \nabla\phi(\mathbf{x}) = \mathbf{x} \ \text{ P-a.s. and } \ \nabla\phi \circ \nabla\psi(\mathbf{u}) = \mathbf{u} \ \ \mathrm{U}_d\text{-a.s.} \tag{2.5}$$

In view of (2.4) and the second statement in (2.5), $\mathbf{F}_\pm := \nabla\phi$ takes values in $\overline{\mathbb{S}}_d$ and pushes P forward to U_d. Moreover, there exist subsets $\breve{\mathrm{spt}}(\mathrm{P})$ and $\breve{\mathbb{S}}_d$ of spt(P) and $\mathbb{S}_d$, respectively, such that (a) $\mathrm{P}\big(\breve{\mathrm{spt}}(\mathrm{P})\big) = 1 = \mathrm{U}_d\big(\breve{\mathbb{S}}_d\big)$, (b) the restriction to $\breve{\mathrm{spt}}(\mathrm{P})$ of $\nabla\phi =: \mathbf{F}_\pm$ and the restriction to $\breve{\mathbb{S}}_d$ of $\nabla\psi =: \mathbf{Q}_\pm$ are bijections, and (c) those restrictions are the inverse of each other. Accordingly, $\mathbf{F}_\pm$ qualifies as a center-outward distribution function.

Definition 2.2 Call $\mathbf{F}_\pm := \nabla\phi$ the *center-outward distribution function* of $\mathrm{P} \in \mathcal{P}_d$.

The following propositions summarize the main properties of $\mathbf{F}_\pm$ and $\mathbf{Q}_\pm$, some of which already have been mentioned in previous comments.

Proposition 2.1 *Let* $\mathbf{Z} \sim \mathrm{P} \in \mathcal{P}_d$ *and denote by* $\mathbf{F}_\pm$ *the center-outward distribution function of* P*. Then,*

- *(i)* $\mathbf{F}_\pm$ *takes values in* $\overline{\mathcal{S}}_d$ *and* $\mathbf{F}_\pm\#\mathrm{P} = \mathrm{U}_d$*:* $\mathbf{F}_\pm$*, thus, is a probability-integral transformation;*
- *(ii)* $\|\mathbf{F}_\pm(\mathbf{Z})\|$ *is uniform over* $[0,1]$*,* $\mathbf{S}(\mathbf{Z}) := \mathbf{F}_\pm(\mathbf{Z})/\|\mathbf{F}_\pm(\mathbf{Z})\|$ *uniform over* $\mathcal{S}_{d-1}$*, and they are mutually independent;*
- *(iii)* $\mathbf{F}_\pm$ *entirely characterizes* P*;*
- *(iv) for* $d=1$*,* $\mathbf{F}_\pm$ *coincides with* $2F-1$ *(F the traditional distribution function).*

For $q \in (0,1)$, define the *center-outward quantile region* and *center-outward quantile contour* of order as

$$\mathbb{C}(q) := \mathbf{Q}_\pm(q\,\overline{\mathbb{S}}_d) = \{\mathbf{z}\,\big|\,\|\mathbf{F}_\pm(\mathbf{Z})\| \leq q\} \text{ and } \mathcal{C}(q) := \mathbf{Q}_\pm(q\,\mathcal{S}_{d-1}) = \{\mathbf{z}\,\big|\,\|\mathbf{F}_\pm(\mathbf{Z})\| = q\}, \tag{2.6}$$

respectively.

Proposition 2.2 *Let* $\mathrm{P} \in \mathcal{P}_d$ *have center-outward quantile function* $\mathbf{Q}_\pm$*. Then,*

- *(i)* $\mathbf{Q}_\pm$ *pushes* U_d *forward to* P*, hence entirely characterizes* P*;*
- *(ii) the* center-outward quantile region $\mathbb{C}(q)$*,* $0 < q < 1$*, has* P*-probability content* q*;*
- *(iii)* $\mathbf{Q}_\pm(u)$ *coincides, for* $d = 1$*, with* $\inf\{x\big|F(x) \geq (1+u)/2)\}$*,* $u \in (-1,1)$*, and* $\mathbb{C}(q)$*,* $q \in (0,1)$*, with (F the traditional distribution function)*[13]

$$\big[\inf\{x\big|F(x) \geq (1-q)/2\}, \inf\{x\big|F(x) \geq (1+q)/2\}\big]\bigcap\overline{\mathrm{spt}}(\mathrm{P}). \tag{2.7}$$

The modulus $\|\mathbf{F}_\pm(\mathbf{x})\|$ thus is the order of the quantile contour and the P-probability content of the largest quantile region containing $\mathbf{x}$; the unit vector $\mathbf{S}(\mathbf{z}) := \mathbf{F}_\pm(\mathbf{z})/\|\mathbf{F}_\pm(\mathbf{z})\|$ has the interpretation of a multivariate sign. Note that the definition of $\mathbb{C}(0)$ so far has been postponed.

These properties are not entirely satisfactory, though, and a bijection between $\breve{\mathrm{spt}}(\mathrm{P})$ and $\breve{\mathbb{S}}_d$ is not enough for meaningful quantile concepts to exist. The terminology *quantile region* and *quantile contour*, indeed, calls for

---

[13]Since $\mathbf{Q}_\pm$ is only a.e. defined, one can as well use spt(P) in (2.7); this, however, no longer produces a closed region and may result in an empty set $\bigcap_{0<q<1}\mathbb{C}(q)$ of medians in (2.8).

a collection of connected, closed, and strictly nested regions $\mathbb{C}(q)$—i.e., such that $\mathbb{C}(q_1) \subsetneq \mathbb{C}(q) \subsetneq \mathbb{C}(q_2)$ for any $0 < q_1 < q < q_2 < 1$—with continuous boundaries $\mathcal{C}(q)$ of Hausdorff dimension $d - 1$; a reasonable[14] definition of a median set then is, with $\mathbb{C}(q)$ ($q \in (0,1)$) defined in (2.6),

$$\mathbb{C}(0) := \bigcap\nolimits_{0<q<1} \mathbb{C}(q). \tag{2.8}$$

Such attractive properties do not hold, unfortunately, and the median set $\mathbb{C}(0)$, as defined in (2.8) may be empty, unless $\mathbf{Q}_\pm$, hence $\mathbf{F}_\pm$, enjoy some continuity properties, which require regularity assumptions on P and its support: see Appendix H for examples. A sufficient condition, as we shall see, is the continuity of $\mathbf{u} \mapsto \mathbf{Q}_\pm(\mathbf{u})$, at least on $\mathbb{S}_d \setminus \{\mathbf{0}\}$.

To see this and understand the special role of $\mathbf{0}$, recall that $\mathbf{Q}_\pm$ is only a.e. defined. Hence, $\mathbf{Q}_\pm(\mathbf{0})$ can take any value compatible with the convexity of ψ—namely, any single point in the subdifferential $\partial\psi(\mathbf{0})$ of the uniquely defined ψ satisfying $\psi(\mathbf{0}) = 0$. As a consequence, continuity of $\mathbf{Q}_\pm$ is impossible unless $\partial\psi(\mathbf{0})$ (and all other subdifferentials—not just almost all of them) contains exactly one single point.

Continuity of the restriction of $\mathbf{Q}_\pm$ to a closed spherical annulus $q^+\overline{\mathbb{S}}_d \setminus q^-\mathbb{S}_d$ yields continuous contours $\mathcal{C}(q)$ and strictly nested closed regions $\mathbb{C}(q)$ for the orders $q \in [q^-, q^+]$. Letting $q^+ = 1 - \epsilon$ and $q^- = \epsilon$ with $\epsilon > 0$ arbitrarily small, continuity of $\mathbf{Q}_\pm$ everywhere except possibly at $\mathbf{0}$ thus yields continuous contours and strictly nested closed regions for the orders $q \in (0,1)$.

The definition of quantile regions implies that all possible values of $\mathbf{Q}_\pm(\mathbf{0})$ are contained in the intersection $\bigcap_{0<q<1} \mathbb{C}(q)$ of all regions of order $q > 0$; hence, $\partial\psi(\mathbf{0}) \subseteq \bigcap_{0<q<1} \mathbb{C}(q)$. Conversely, any point $\mathbf{u} \neq \mathbf{0}$ has a neighborhood $V(\mathbf{u})$ such that $\mathbf{0} \notin V(\mathbf{u})$. Assuming that $\mathbf{Q}_\pm$ is continuous everywhere but at $\mathbf{0}$, $\mathbf{Q}_\pm(V(\mathbf{u})) \cap \bigcap_{0<q<1} \mathbb{C}(q) = \emptyset$. Hence, $\partial\psi(\mathbf{0}) = \bigcap_{0<q<1} \mathbb{C}(q)$. As the subdifferential of a convex function ψ, $\partial\psi(\mathbf{0})$, hence $\bigcap_{0<q<1} \mathbb{C}(q)$, is closed and convex. Because P has a density and $\mathbf{0}$ is in the interior of ψ's domain, it also is compact and has Lebesgue measure zero (Lemma A.22 in Figalli (2017)).

It follows that by defining the median set as $\mathbb{C}(0) := \bigcap_{0<q<1} \mathbb{C}(q) = \partial\psi(\mathbf{0})$ (instead of $\mathbb{C}(0) := \mathbf{Q}_\pm(\mathbf{0})$, which is not uniquely determined), we do not need continuity at $\mathbf{0}$ to obtain strict nestedness of all quantile contours and regions—now including $\mathbb{C}(0)$—while (2.8), of course, is automatically satisfied.

This, which justifies giving up continuity at $\mathbf{0}$ (and only there), is not an unimportant detail: Proposition 2.3 below indeed shows that important classes of distributions yield quantile functions $\mathbf{Q}_\pm$ that are not continuous over the ball $\mathbb{S}_d$ but nevertheless enjoy continuity over the punctured ball $\mathbb{S}_d \setminus \{\mathbf{0}\}$.

Denote by $\mathcal{P}_d^{\mathrm{conv}}$ the class of distributions $\mathrm{P}_f \in \mathcal{P}_d$ such that *(a)* $\overline{\mathrm{spt}}(\mathrm{P}_f)$ is a convex set[15] and, *(b)* for all $D \in \mathbb{R}^+$, there exist constants $\Lambda_{D;f}$ and $\lambda_{D;f}$

---

[14]By analogy with the definition of $\mathbb{C}(q)$ for $q > 0$, one may be tempted to define $\mathbb{C}(0)$ as $\mathbf{Q}_\pm(\mathbf{0})$. This yields for $\mathbb{C}(0)$ an arbitrary point in the subdifferential $\partial\psi(\mathbf{0})$ which, unless that subdifferential consists of a single point, cannot satisfy (2.8).

[15]That convex set is not necessarily bounded.

in $(0, \infty)$ such that $\lambda_{D;f} \leq f(\mathbf{x}) \leq \Lambda_{D;f}$ for all $\mathbf{x} \in (D\,\mathbb{S}_d) \cap \overline{\mathrm{spt}}(\mathrm{P}_f)$. That class includes the class $\mathcal{P}_d^+$ of distributions with support $\mathrm{spt}(\mathrm{P}) = \mathbb{R}^d$ considered by Hallin (2017) and Figalli (2018).

The following result, which establishes the continuity properties of $\mathbf{F}_\pm$ and $\mathbf{Q}_\pm$ for $\mathrm{P} \in \mathcal{P}_d^{\mathrm{conv}}$, extends the main result obtained for $\mathcal{P}_d^+$ by Figalli (2018) and is borrowed, with some minor additions, from del Barrio et al. (2019).

Proposition 2.3 *Let* $\mathrm{P} \in \mathcal{P}_d^{\mathrm{conv}}$ *have density* f *and support* $\overline{\mathrm{spt}}(\mathrm{P})$. *Then, its center-outward distribution function* $\mathbf{F}_\pm = \nabla\phi$ *is continuous and single-valued on* $\mathbb{R}^d$ *and* $\|\mathbf{F}_\pm(\mathbf{x})\| = 1$ *for* $\mathbf{x} \notin \mathrm{spt}(\mathrm{P})$. *Furthermore, there exists a compact convex set* $K \subset \overline{\mathrm{spt}}(\mathrm{P})$ *with Lebesgue measure zero such that*

  (i) $\mathbf{F}_\pm$ *and the center-outward quantile function* $\mathbf{Q}_\pm = \nabla\psi$ *are homeomorphisms between* $\mathbb{S}_d \backslash \{\mathbf{0}\}$ *and* $\mathrm{spt}(\mathrm{P}) \backslash K$, *on which they are inverse of each other; for* $d = 1, 2$, *however,* K *contains a single point and the homeomorphisms are between* $\mathbb{S}_d$ *and* $\mathrm{spt}(\mathrm{P})$;

 (ii) *the quantile contours* $\mathcal{C}(q)$ *and regions* $\mathbb{C}(q)$, $0 < q < 1$ *defined by* $\mathbf{Q}_\pm$ *are such that* $\bigcap_{0<q<1} \mathbb{C}(q) = \partial\psi(\{\mathbf{0}\}) = K$; K *thus qualifies as the median set* $\mathbb{C}(0)$ *of* P *as defined in* (2.8).

If, moreover, $f \in \mathcal{C}_{\mathrm{loc}}^{k,\alpha}(\mathrm{spt}(\mathrm{P}_f))$ *for some* $k \geq 0$, *then*

(iii) (a) $\mathbf{Q}_\pm$ *and* $\mathbf{F}_\pm$ *are diffeomorphisms of class* $\mathcal{C}_{\mathrm{loc}}^{k+1,\alpha}$ *between* $\mathbb{S}_d \backslash \{\mathbf{0}\}$ *and* $\mathrm{spt}(\mathrm{P}) \backslash \mathbb{C}(0)$;

     (b) $f(\mathbf{z}) = c_d^{-1} \det\left[\mathbf{H}_\psi\big(\nabla\phi(\mathbf{z})\big)\right] \|\nabla\phi(\mathbf{z})\|^{1-d} I\big[\mathbf{z} \in \mathrm{spt}(\mathrm{P}_f) \backslash \mathbb{C}(0)\big]$ *where* c_d *is the area* $2\pi^{d/2}/\Gamma(d/2)$ *of the unit sphere* $\mathcal{S}_{d-1}$ *and* $\mathbf{H}_{\phi^*}(\mathbf{u})$ *the Hessian*[16] *of* ψ *computed at* $\mathbf{u}$.

Denote by $\mathcal{P}_d^\pm \subset \mathcal{P}_d$ the class of all distributions of the form $\mathrm{P} = \nabla\Upsilon$ where Υ is convex and $\nabla\Upsilon$ a homeomorphism from $\mathbb{S}_d \backslash \{\mathbf{0}\}$ to $\nabla\Upsilon(\mathbb{S}_d \backslash \{\mathbf{0}\})$ such that $\nabla\Upsilon(\{\mathbf{0}\})$ is a compact convex set of Lebesgue measure zero. By construction, such $\mathrm{P} \in \mathcal{P}_d^\pm$ has center-outward quantile function $\mathbf{Q}_\pm = \nabla\Upsilon$, center-outward distribution function $\mathbf{F}_\pm(\mathbf{x}) = (\nabla\Upsilon)^{-1}$ for $\mathbf{x}$ in the range of $\nabla\Upsilon$ and $\|\mathbf{F}_\pm(\mathbf{x})\| = 1$ outside that range, and satisfies Proposition 2.3; the latter actually can be rephrased as $\mathcal{P}_d^{\mathrm{conv}} \subset \mathcal{P}_d^\pm$, with the following immediate corollary in terms of quantile regions and contours.

Corollary 2.1 *For any* $\mathrm{P} \in \mathcal{P}_d^\pm$ *(hence, any* $\mathrm{P} \in \mathcal{P}_d^{\mathrm{conv}}$) *and* $q \in [0, 1)$, *the quantile regions* $\mathbb{C}(q)$ *are closed, connected, and nested, with continuous boundaries* $\mathcal{C}(q)$ *satisfying* $\mu_d(\mathcal{C}(q)) = 0$.

For any distribution $\mathrm{P} \in \mathcal{P}_d^\pm$, $\mathbf{F}_\pm$ thus induces a (partial) ordering of $\mathbb{R}^d$ similar to the ordering induced on the unit ball by the system of polar coordinates, and actually coincides with the "vector rank transformation" considered in Chernozhukov et al. (2017) when the reference distribution is U_d. The quantile contours $\mathcal{C}(q)$ also have the interpretation of depth contours associated with their Monge-Kantorovich depth. Their assumption of a compact support satisfying Cafarelli regularity conditions are sufficient (not necessary) for $\mathrm{P} \in \mathcal{P}_d^\pm$.

---

[16]That Hessian exists since $k \geq 0$ and $\nabla\phi(\mathbf{z}) \neq \mathbf{0}$ for $\mathbf{z} \in \mathrm{spt}(\mathrm{P}_f) \backslash \mathbb{C}(0)$.

2.2. Center-outward ranks and signs in $\mathbb{R}^d$

Turning to the sample situation, let $\mathbf{Z}^{(n)} := \big(\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}\big)$ denote an n-tuple of random vectors— observations or residuals associated with some parameter $\boldsymbol{\theta}$ of interest. We throughout consider the case that the $\mathbf{Z}_i^{(n)}$'s are (possibly, under parameter value $\boldsymbol{\theta}$) i.i.d. with density $f \in \mathcal{F}^d$, distribution P and center-outward distribution function $\mathbf{F}_{\pm}$.

For the empirical counterpart $\mathbf{F}_{\pm}^{(n)}$ of $\mathbf{F}_{\pm}$, we propose the following extension of the univariate concept described in Appendix B. Assuming $d \geq 2$, let n factorize into

$$n = n_R n_S + n_0, \qquad n_R, \ n_S, \ n_0 \in \mathbb{N}, \quad 0 \leq n_0 < \min(n_R, n_S) \qquad (2.9)$$

where $n_R \to \infty$ and $n_S \to \infty$ as $n \to \infty$ (implying $n_0/n \to 0$); (2.9) is extending to $d \geq 2$ the factorization of n into $n = \lfloor \frac{n}{2} \rfloor 2 + n_0$ with $n_0 = 0$ (n even) or $n_0 = 1$ (n odd) that leads, for $d = 1$, to the grids (B.6).

Next, consider a sequence of "regular grids" of $n_R n_S$ points in the unit ball $\mathbb{S}_d$ obtained as the intersection between

    – a "regular" n_S-tuple $\mathfrak{S}^{(n_S)} := (\mathbf{u}_1, \ldots \mathbf{u}_{n_S})$ of unit vectors, and

    – the n_R hyperspheres centered at $\mathbf{0}$, with radii $\dfrac{j}{n_R + 1}$, $j = 1, \ldots, n_R$,

along with n_0 copies of the origin whenever $n_0 > 0$. In theory, by a "regular" n_S-tuple $\mathfrak{S}^{(n_S)} = (\mathbf{u}_1, \ldots \mathbf{u}_{n_S})$, we only mean that the sequence of uniform discrete distributions over $\{\mathbf{u}_1, \ldots \mathbf{u}_{n_S}\}$ converges weakly, as $n_S \to \infty$, to the uniform distribution over $\mathcal{S}_{d-1}$. In practice, each n_S-tuple should be "as uniform as possible". For $d = 2$, perfect regularity can be achieved by dividing the unit circle into n_S arcs of equal length $2\pi/n_S$. Starting with $d = 3$, however, this typically is no longer possible. A random array of n_S independent and uniformly distributed unit vectors does satisfy (almost surely) the weak convergence requirement. More regular deterministic arrays (with faster convergence) can be considered, though, such as the *low-discrepancy sequences* of the type considered in numerical integration and Monte-Carlo methods (see, e.g., Niederreiter (1992), Judd (1998), Dick and Pillichshammer (2014), or Santner et al. (2003)), which are current practice in numerical integration and the design of computer experiments.

The resulting grid of $n_R n_S$ points then is such that the discrete distribution with probability masses $1/n$ at each gridpoint and probability mass n_0/n at the origin converges weakly to the uniform U_d over the ball $\mathbb{S}_d$. That grid, along with the n_0 copies of the origin, is called the *augmented grid* (n points).

We then define $\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)})$, $i = 1, \ldots, n$ as the solution of an optimal coupling problem between the observations and the augmented grid. Let $\mathcal{T}$ denote the set of all possible bijective mappings between $\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}$ and the n points of the augmented grid just described. Under the assumptions made, the $\mathbf{Z}_i^{(n)}$'s are all distinct with probability one, so that $\mathcal{T}$ contains $n!/n_0!$ classes of $n_0!$ indistinguishable couplings each (two couplings T_1 and T_2 are indistinguishable if $T_1(\mathbf{Z}_i^{(n)}) = T_2(\mathbf{Z}_i^{(n)})$ for all i).

Definition 2.3 Call *empirical center-outward distribution function* any of the mappings $\mathbf{F}_{\pm}^{(n)} \colon \big(\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}\big) \mapsto \big(\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_1^{(n)}), \ldots, \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_n^{(n)})\big) =: \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}^{(n)})$ satisfying

$$\sum_{i=1}^{n} \big\| \mathbf{Z}_i^{(n)} - \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)}) \big\|^2 = \min_{T \in \mathcal{T}} \sum_{i=1}^{n} \big\| \mathbf{Z}_i^{(n)} - T(\mathbf{Z}_i^{(n)}) \big\|^2 \qquad (2.10)$$

or, equivalently,

$$\sum_{i=1}^{n} \big\| \mathbf{Z}_i^{(n)} - \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)}) \big\|^2 = \min_{\pi} \sum_{i=1}^{n} \big\| \mathbf{Z}_{\pi(i)}^{(n)} - \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)}) \big\|^2 \qquad (2.10)$$

where the set $\{\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)}) \,|\, i = 1, \ldots, n\}$ consists of the n points of the augmented grid and π ranges over the $n!$ possible permutations of $\{1, 2, \ldots, n\}$.

Determining such a coupling is a standard optimal assignment problem, which takes the form of a linear program for which efficient algorithms are available (see Peyré and Cuturi (2019) for a recent survey).

Call *order statistic* $\mathbf{Z}_{(\,.\,)}^{(n)}$ of $\mathbf{Z}^{(n)}$ the *un*-ordered n-tuple of $\mathbf{Z}_i^{(n)}$ values—equivalently, an arbitrarily ordered version of the same. To fix the notation, let $\mathbf{Z}_{(\,.\,)}^{(n)} := \big(\mathbf{Z}_{(1)}^{(n)}, \ldots, \mathbf{Z}_{(n)}^{(n)}\big)$, where $\mathbf{Z}_{(i)}^{(n)}$ is such that its first component is the ith order statistic of the n-tuple of $\mathbf{Z}_i^{(n)}$'s first components. Under this definition, the points $\mathbf{z} \in \mathbb{R}^{nd}$ at which (2.10) possibly admits two minimizers or more lie in the union N of a finite number of linear subspaces of $\mathbb{R}^{nd}$ where some equidistance properties hold between $\mathbf{Z}_i^{(n)}$'s and gridpoints; therefore, N is $\mathbf{Z}_{(\,.\,)}^{(n)}$-measurable and has Lebesgue measure zero. Such multiplicities have no practical impact, thus, since (for a given grid) they take place on a unique null set N.

Another type of multiplicity occurs, even over $\mathbb{R}^{nd} \backslash N$: each of the minimizers $\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}^{(n)})$ of (2.10) indeed is such that the n-tuple

$$\big\{ \big(\mathbf{Z}_1^{(n)}, \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_1^{(n)})\big), \ldots, \big(\mathbf{Z}_n^{(n)}, \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_n^{(n)})\big) \big\} \qquad (2.11)$$

is one of the $n_0!$ indistinguishable couplings between the n observations and the n points of the augmented grid that minimize, over the $n!$ possible couplings, the sum of within-pairs squared distances. That multiplicity, which involves n_0 tied observations, does not occur for $n_0 = 0$ or 1: the mapping $\mathbf{z} \mapsto \big(\mathbf{z}_{(\,.\,)}, \mathbf{F}_{\pm}^{(n)}(\mathbf{z})\big)$ then is injective over $\mathbb{R}^{nd} \backslash N$. For $n_0 > 1$, it is easily taken care of by replacing, in the grid, the $n_0 > 1$ copies of $\mathbf{0}$ with n_0 i.i.d. points uniformly distributed over $(n_R + 1)^{-1} \mathbb{S}_d$—a convenient tie-breaking device (see footnote 9 in Appendix D.2) restoring the injectivity over $\mathbb{R}^{nd} \backslash N$ of $\mathbf{z} \mapsto \big(\mathbf{z}_{(\,.\,)}, \mathbf{F}_{\pm}^{(n)}(\mathbf{z})\big)$.

Reinterpreting (2.10) as an expected (conditional on the order statistic—see Section 2.4 for a precise definition) transportation cost, the same optimal coupling(s) also constitute(s) the optimal L^2 transport mapping the empirical distribution to the uniform discrete distribution over the augmented grid (and, conversely, the two problems being entirely symmetric, the optimal L^2 transport mapping the uniform discrete distribution over the augmented grid to the empirical distribution). Classical results (McCann (1995) again) then show that optimality is achieved (i.e., (2.10) is satisfied) iff the so-called *cyclical monotonicity* property holds for the n-tuple (2.11).

Definition 2.4 A subset S of $\mathbb{R}^d \times \mathbb{R}^d$ is said to be *cyclically monotone* if, for any finite collection of points $\{(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_k, \mathbf{y}_k)\} \subseteq S$,

$$\langle \mathbf{y}_1, \ \mathbf{x}_2 - \mathbf{x}_1 \rangle + \langle \mathbf{y}_2, \ \mathbf{x}_3 - \mathbf{x}_2 \rangle + \ldots + \langle \mathbf{y}_k, \ \mathbf{x}_1 - \mathbf{x}_k \rangle \leq 0. \qquad (2.12)$$

The subdifferential of a convex function does enjoy cyclical monotonicity, which heuristically can be interpreted as a discrete version of the fact that a smooth convex function has a positive semi-definite second-order differential.

Note that a finite subset $S = \{(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_n, \mathbf{y}_n)\}$ of $\mathbb{R}^d \times \mathbb{R}^d$ is cyclically monotone iff (2.12) holds for $k = n$—equivalently, iff, among all pairings of $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $(\mathbf{y}_1, \ldots, \mathbf{y}_n)$, S maximizes $\sum_{i=1}^n \langle \mathbf{x}_i, \mathbf{y}_i \rangle$ (an empirical covariance), or minimizes $\sum_{i=1}^n \|\mathbf{y}_i - \mathbf{x}_i\|^2$ (an empirical distance). In other words, a finite subset S is cyclically monotone iff the couples $(\mathbf{x}_i, \mathbf{y}_i)$ are a solution of the optimal assignment problem with assignment cost $\|\mathbf{y}_i - \mathbf{x}_i\|^2$. The L^2 transportation cost considered here is thus closely related to the concept of convexity and the geometric property of cyclical monotonicity; it does not play the statistical role of an estimation loss function, though—the L^2 distance between the empirical transport and its population counterpart (the expectation of which might be infinite), indeed, is never considered.

Associated with our definition of an empirical center-outward distribution function $\mathbf{F}_\pm^{(n)}$ are the following concepts of

- *center-outward ranks* $R_{\pm,i}^{(n)} := (n_R + 1)\|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_i^{(n)})\|$,
- *empirical center-outward quantile contours* and *regions*

$$\mathcal{C}_{\pm;\mathbf{Z}^{(n)}}^{(n)}\Big(\frac{j}{n_R+1}\Big) := \big\{\mathbf{Z}_i^{(n)}|R_{\pm,i}^{(n)} = j\big\} \text{ and } \mathbb{C}_{\pm;\mathbf{Z}^{(n)}}^{(n)}\Big(\frac{j}{n_R+1}\Big) := \big\{\mathbf{Z}_i^{(n)}|R_{\pm,i}^{(n)} \leq j\big\},$$

respectively, where $j/(n_R+1)$, $j = 0, 1, \ldots, n_R$, is an empirical probability contents, to be interpreted as a quantile order,

- *center-outward signs* $\mathbf{S}_{\pm,i}^{(n)} := \mathbf{F}_\pm^{(n)}(\mathbf{Z}_i^{(n)})I\Big[\mathbf{F}_\pm^{(n)}(\mathbf{Z}_i^{(n)}) \neq \mathbf{0}\Big]/\|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_i^{(n)})\|$, and *center-outward sign curves* $\{\mathbf{Z}_i^{(n)}|\mathbf{S}_{\pm,i}^{(n)} = \mathbf{u}\}$, $\mathbf{u} \in \mathfrak{S}^{(n_S)}$.

The contours, curves, and regions defined here are finite collections of observed points; the problem of turning them into continuous contours enclosing compact regions and continuous lines is treated in Section 3.

Up to this point, we have defined multivariate generalizations of the univariate concepts of center-outward distribution and quantile functions, center-outward ranks and signs, all reducing to their univariate analogues in case $d = 1$. However, it remains to show that those multivariate extensions are adequate in the sense that they enjoy in $\mathbb{R}^d$ the characteristic properties that make the inferential success of their univariate counterparts—namely,

(GC) a Glivenko-Cantelli-type asymptotic relation between $\mathbf{F}_\pm^{(n)}$ and $\mathbf{F}_\pm$, and

(DF$^+$) the (essential) maximal ancillarity property described for $d = 1$ in Section 1.1.

This is the objective of Sections 2.3 and 2.4.

2.3. Glivenko-Cantelli

With the definitions adopted in Sections 2.1 and 2.2, the traditional Glivenko-Cantelli theorem, under center-outward form (B.7), holds, essentially *ne varietur*, in $\mathbb{R}^d$ under $\mathrm{P} \in \mathcal{P}_d^{\pm}$.

Proposition 2.4 *Let* $\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}$ *be i.i.d. with distribution* $\mathrm{P} \in \mathcal{P}_d^{\pm}$. *Then,*

$$\max_{1 \leq i \leq n} \left\| \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)}) - \mathbf{F}_{\pm}(\mathbf{Z}_i^{(n)}) \right\| \longrightarrow 0 \quad a.s. \ as \ n \to \infty. \tag{2.13}$$

The particular case of elliptical distributions is considered in Appendix C.

Proposition 2.4 considerably reinforces, under more general assumptions (no second-order moments), an early strong consistency result by Cuesta-Albertos et al. (1997). It readily follows from the more general Proposition 3.3, which extends (2.13) under sup form to cyclically monotone interpolations of $\mathbf{F}_{\pm}^{(n)}$.

2.4. Distribution-freeness and maximal ancillarity

Proposition 2.5 provides the multivariate extension of the usual distributional properties of univariate order statistics and ranks. Note that, contrary to Proposition 2.4, it holds for $\mathrm{P} \in \mathcal{P}_d$. See Appendices D.2 and D.1 for a proof and details on sufficiency, ancillarity, and (strong) essential maximal ancillarity.

Proposition 2.5 *Let* $\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}$ *be i.i.d. with distribution* $\mathrm{P} \in \mathcal{P}_d$, *center-outward distribution function* $\mathbf{F}_{\pm}$, *order statistic* $\mathbf{Z}_{(\cdot)}^{(n)}$, *and empirical center-outward distribution function* $\mathbf{F}_{\pm}^{(n)}$. *Then,*

(i) $\mathbf{Z}_{(\cdot)}^{(n)}$ *is sufficient and complete, hence minimal sufficient, for* $\mathcal{P}_d^{(n)}$;

(ii)(DF) $\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}^{(n)}) := \left(\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_1^{(n)}), \ldots, \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_n^{(n)}) \right)$ *is uniformly distributed over the* $n!/n_0!$ *permutations with repetitions (the origin counted as* n_0 *indistinguishable points) of the grid described in Section 2.2;*

(iii) for $n_0 = 0$, *the vectors of center-outward ranks* $\left(R_{\pm,1}^{(n)}, \ldots, R_{\pm,n}^{(n)} \right)$ *and signs* $\left(\mathbf{S}_{\pm,1}^{(n)}, \ldots, \mathbf{S}_{\pm,n}^{(n)} \right)$ *are mutually independent; for* $n_0 > 0$, *the same independence holds for the* $(n_R n_S)$-*tuple of ranks and signs associated with the (random) set* $\{ i_1 < \ldots < i_{n_R n_S} \}$ *such that* $\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_{i_j}^{(n)}) \neq \mathbf{0}$;

(iv) for all $\mathrm{P} \in \mathcal{P}_d$, $\mathbf{Z}_{(\cdot)}^{(n)}$ *and* $\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}^{(n)})$ *are mutually* P-*independent, and*

(v) for $n_0 \leq 1$ *or after adequate tie-breaking (cf. comment below),* $\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}^{(n)})$ *is strongly* $\mathcal{P}_d^{(n)}$-*essentially maximal ancillary.*

In *(iii)* and *(v)*, n_0 plays a special role. In *(iii)*, the fact that the sign, for the n_0 observations mapped to the origin, is not a unit vector induces, for $n_0 \geq 1$, a (very mild) dependence between signs and ranks which, however, does not affect joint distribution-freeness. In *(v)*, $n_0 \leq 1$ implies that $\mathbf{z} \mapsto \left(\mathbf{z}_{(\cdot)}, \mathbf{F}_{\pm}^{(n)}(\mathbf{z}) \right)$ is injective over $\mathbb{R}^{nd} \setminus N$. As previously explained, injectivity is easily restored via a simple tie-breaking device: *(v)* then is satisfied irrespective of n_0. Note that the proportion n_0/n of points involved anyway tends to zero as $n \to \infty$.

More important is the interpretation of essential maximal ancillarity in terms of finite-sample semiparametric efficiency in case $\mathbf{Z}_i$ is the $\boldsymbol{\theta}$-residual $\mathbf{Z}_i(\boldsymbol{\theta})$ in

some semiparametric model with parameter of interest $\boldsymbol{\theta}$ and nuisance f (see Section 1.1). Another crucial consequence of *(v)* is the following corollary.

Corollary 2.2 *Denote by $\tilde{\mathcal{B}}_{\pm}^{(n)}$ the sub-σ-field generated by the mapping $\tilde{\mathbf{F}}_{\pm}^{(n)}$ associated with some other deterministic[17] n-points grid—whether over the unit ball, the unit cube, or any other fixed domain. Then, there exists $M \in \mathcal{B}_d^n$ such that* $\mathrm{P}^{(n)}(M) = 0$ *for all* $\mathrm{P} \in \mathcal{P}_d$ *and* $\mathcal{B}_{\pm}^{(n)} \cap \left(\mathbb{R}^{nd} \setminus M\right) = \tilde{\mathcal{B}}_{\pm}^{(n)} \cap \left(\mathbb{R}^{nd} \setminus M\right)$.

It follows (see Appendix D.1) that $\mathcal{B}_{\pm}^{(n)}$ and $\tilde{\mathcal{B}}_{\pm}^{(n)}$ are strongly essentially equivalent σ-fields. Ranks and signs associated with distinct grids, thus, essentially generate the same sub-σ-fields, which considerably attenuates the impact of grid choices; see Appendix D.2 for details and a proof.

3. Smooth interpolation under cyclical monotonicity constraints

So far, Definition 2.3 only provides a value of $\mathbf{F}_{\pm}^{(n)}$ at the sample values $\mathbf{Z}_i^{(n)}$. If $\mathbf{F}_{\pm}^{(n)}$ is to be extended to $\mathbf{z} \in \mathbb{R}^d$, an interpolation $\overline{\mathbf{F}}_{\pm}^{(n)}$, similar for instance to the one shown, for $d = 1$, in Figure 5 of Appendix B, has to be constructed. Such interpolation should belong to the class of gradients of convex functions from $\mathbb{R}^d$ to $\mathbb{S}_d$, so that the resulting contours $\mathcal{C}_{\pm;\mathbf{Z}^{(n)}}^{(n)}$ have the nature of continuous quantile contours. Moreover, it still should enjoy (now under a $\sup_{\mathbf{z} \in \mathbb{R}^d}$ form similar to (B.2)) the Glivenko-Cantelli property.[18] Constructing such interpolations is considerably more delicate for $d \geq 2$ than in the univariate case.

Empirical center-outward distribution functions $\mathbf{F}_{\pm}^{(n)}$, as defined in Definition 2.3, are cyclically monotone (discrete) mappings from the random sample (or n-tuple of residuals) $\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}$ to a (nonrandom) regular grid of $\mathbb{S}_d$; hence, $\mathbf{F}_{\pm}^{(n)}$ is defined at the observed points only. Although such discrete $\mathbf{F}_{\pm}^{(n)}$ perfectly fulfills its statistical role as a sufficient sample summary carrying the same information as the sample itself, one may like to define an empirical center-outward distribution function as an object of the same nature—a smooth cyclically monotone mapping from $\mathbb{R}^d$ to $\mathbb{S}_d$—as its population counterpart $\mathbf{F}_{\pm}$. This brings into the picture the problem of the existence and construction, within the class of gradients of convex functions, of a continuous extension $\mathbf{x} \mapsto \overline{\mathbf{F}}_{\pm}^{(n)}(x)$ of the discrete $\mathbf{F}_{\pm}^{(n)}$, yielding a Glivenko-Cantelli theorem of the $\sup_{\mathbf{x} \in \mathbb{R}^d}$ form— namely, $\sup_{\mathbf{x} \in \mathbb{R}^d} \|\overline{\mathbf{F}}_{\pm}^{(n)}(\mathbf{x}) - \mathbf{F}_{\pm}(\mathbf{x})\| \to 0$ a.s. as $n \to \infty$—rather than the $\max_{1 \leq i \leq n}$ form established in Proposition 2.4. That problem reduces to the more general problem of smooth interpolation under cyclical monotonicity (see Definition 2.4) constraints, which we now describe.

Let $\boldsymbol{\mathcal{X}}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and $\boldsymbol{\mathcal{Y}}_n = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ denote two n-tuples of points in $\mathbb{R}^d$. Assuming that there exists a *unique* bijection $T : \boldsymbol{\mathcal{X}}_n \to \boldsymbol{\mathcal{Y}}_n$ such that the

---

[17]Deterministic here means nonrandom or randomly generated from a probability space that has no relation to the observations.

[18]It should be insisted, though, that the $\max_{1 \leq i \leq n}$ form (2.13) of Glivenko-Cantelli is not really restrictive, as interpolations do not bring any additional information, and are mainly intended for (graphical or virtual) depiction of quantile contours.

set $\{(\mathbf{x}, T(\mathbf{x}))|\ \mathbf{x} \in \boldsymbol{\mathcal{X}}_n\}$ is cyclically monotone, there is no loss of generality in relabeling the elements of $\boldsymbol{\mathcal{Y}}_n$ so that $\mathbf{y}_i = T(\mathbf{x}_i)$. Accordingly, we throughout are making the following assumption.

Assumption (A). The n-tuples $\boldsymbol{\mathcal{X}}_n$ and $\boldsymbol{\mathcal{Y}}_n$ are such that $T\colon \mathbf{x}_i \mapsto T(\mathbf{x}_i) = \mathbf{y}_i$ for $i = 1, \dots, n$ is the unique cyclically monotone bijective map from $\boldsymbol{\mathcal{X}}_n$ to $\boldsymbol{\mathcal{Y}}_n$.

Our goal, under Assumption (A), is to construct a smooth (at least continuous) cyclically monotone map $\overline{T}\colon \mathbb{R}^d \to \mathbb{R}^d$ such that $\overline{T}(\mathbf{x}_i) = T(\mathbf{x}_i) = \mathbf{y}_i$, $i = 1, \dots, n$.

It is well known that the subdifferential of a convex function ψ from $\mathbb{R}^d$ to $\mathbb{R}$ enjoys cyclical monotonicity. A classical result by Rockafellar (1966) establishes the converse: any finite cyclically monotone subset S of $\mathbb{R}^d \times \mathbb{R}^d$ lies in the subdifferential of some convex function. Our result reinforces this characterization by restricting to differentiable convex functions. Note that a differentiable convex function ψ is automatically continuously differentiable, with unique (at all $\mathbf{x}$) subgradient $\nabla\psi(\mathbf{x})$ and subdifferential $\{(\mathbf{x}, \nabla\psi(\mathbf{x}))|\mathbf{x} \in \mathbb{R}^d\}$. When ψ is convex and differentiable, the mapping $x \mapsto \nabla\psi(x)$ thus enjoys cyclical monotonicity. We show in Corollary 3.1 that, conversely, any subset $S = \{(\mathbf{x}_i, \mathbf{y}_i)|i = 1, \dots, n\}$ of $\mathbb{R}^d \times \mathbb{R}^d$ enjoying cyclical monotonicity is the subdifferential (at $\mathbf{x}_1, \dots, \mathbf{x}_n$) of some (continuously) differentiable convex function ψ.

Note that Assumption (A) holds if and only if identity is the unique minimizer of $\sum_{i=1}^n \|\mathbf{x}_i - \mathbf{y}_{\sigma(i)}\|^2$ among the set of all permutations σ of $\{1, \dots, n\}$. Letting $c_{i,j} := \|\mathbf{x}_i - \mathbf{y}_j\|^2$, the same condition can be recast in terms of uniqueness of the solution of the linear program

$$\min_{\pi} \sum_{i=1}^n \sum_{j=1}^n c_{i,j}\pi_{i,j} \quad \text{s.t.} \ \sum_{i=1}^n \pi_{i,j} = \sum_{j=1}^n \pi_{i,j} = \frac{1}{n}, \ \pi_{i,j} \geq 0, \ i,j = 1, \dots, n. \ (3.1)$$

Clearly, $\sigma(i) = i$ minimizes $\sum_{i=1}^n \|\mathbf{x}_i - \mathbf{y}_{\sigma(i)}\|^2$ iff $\pi_{i,i} = 1/n$, $\pi_{i,j} = 0$ for $j \neq i$ is the unique solution of (3.1).

Our solution to the cyclically monotone interpolation problem is constructed in two steps. First (Step 1), we extend T to a piecewise constant cyclically monotone map defined on a set in $\mathbb{R}^d$ whose complementary has Lebesgue measure zero. Being piecewise constant, that map cannot be smooth. To fix this, we apply (Step 2) a regularization procedure yielding the required smoothness while keeping the interpolation feature. For Step 1, we rely on the following result (see Appendix F.1 for the proof).

Proposition 3.1 *Assume that $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^d$ and $\mathbf{y}_1, \dots, \mathbf{y}_n \in \mathbb{R}^d$ are such that $i \neq j$ implies $\mathbf{x}_i \neq \mathbf{x}_j$ and $\mathbf{y}_i \neq \mathbf{y}_j$. Then,*

(i) the map $T(\mathbf{x}_i) = \mathbf{y}_i$, $i = 1, \dots, n$ is cyclically monotone if and only if there exist real numbers $\psi_1, \dots, \psi_n$ such that

$$\langle \mathbf{x}_i, \mathbf{y}_i \rangle - \psi_i = \max_{j=1,\dots,n} (\langle \mathbf{x}_i, \mathbf{y}_j \rangle - \psi_j), \quad i = 1, \dots, n;$$

(ii) furthermore, T is the unique cyclically monotone map from $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ to $\{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ if and only if there exist real numbers $\psi_1, \dots, \psi_n$ such that

$$\langle \mathbf{x}_i, \mathbf{y}_i \rangle - \psi_i > \max_{j=1,\dots,n, j \neq i} (\langle \mathbf{x}_i, \mathbf{y}_j \rangle - \psi_j), \quad i = 1, \dots, n. \qquad (3.2)$$

Remark 3.1 The condition, in Proposition 3.1, that $\mathbf{y}_1, \ldots, \mathbf{y}_n$ are distinct in general is not satisfied in the case of empirical center-outward distribution functions, where, typically, $\mathbf{y}_1 = \cdots = \mathbf{y}_{n_0}$ with $\mathbf{y}_1 \neq \mathbf{y}_i$ for $i > n_0$ and n_0 ranging between 0 and $\min(n_R, n_S) - 1$. This can be taken care of by means of the tie-breaking device described in Section 2.2. The proof (see Appendix F.1), however, is easily adapted to show that the map $T(\mathbf{x}_i) = \mathbf{y}_i$, $i = 1, \ldots, n$ is cyclically monotone if and only if there exist real numbers $\psi_1, \psi_{n_0+1}, \ldots, \psi_n$ such that, setting $\psi_i = \psi_1$, $i = 2, \ldots, n_0$,

$$\langle \mathbf{x}_i, \mathbf{y}_i \rangle - \psi_i = \max_{j=1,\ldots,n} (\langle \mathbf{x}_i, \mathbf{y}_j \rangle - \psi_j), \quad i = 1, \ldots, n.$$

Similarly, the map $T(\mathbf{x}_i) = \mathbf{y}_i$, $i = 1, \ldots, n$ is the unique cyclically monotone map from $\boldsymbol{\mathcal{X}}_n$ to $\{\mathbf{y}_1, \mathbf{y}_{n_0+1} \ldots, \mathbf{y}_n\}$ mapping n_0 points in $\boldsymbol{\mathcal{X}}_n$ to $\mathbf{y}_1$ if and only if there exist real numbers $\psi_1, \psi_{n_0+1}, \ldots, \psi_n$ such that

$$\langle \mathbf{x}_i, \mathbf{y}_1 \rangle - \psi_1 > \langle \mathbf{x}_i, \mathbf{y}_j \rangle - \psi_j, \quad i = 1, \ldots, n_0, \, j = n_0 + 1, \ldots, n,$$

$$\langle \mathbf{x}_i, \mathbf{y}_i \rangle - \psi_i > \langle \mathbf{x}_i, \mathbf{y}_j \rangle - \psi_j, \quad i = n_0 + 1, \ldots, n, \, j = 1, n_0 + 1, \ldots, n, \, j \neq i.$$

Details are omitted.

As a consequence of Proposition 3.1, we can extend T to a cyclically monotone map from $\mathbb{R}^d$ to $\mathbb{R}^d$ as follows. Under Assumption (A), we can choose $\psi_1, \ldots, \psi_n$ such that (3.2) holds. Consider the convex map

$$\mathbf{x} \mapsto \varphi(\mathbf{x}) := \max_{1 \leq j \leq n} (\langle \mathbf{x}, \mathbf{y}_j \rangle - \psi_j). \tag{3.3}$$

Now the sets $C_i = \{x \in \mathbb{R}^d \, | \, (\langle \mathbf{x}, \mathbf{y}_i \rangle - \psi_i) > \max_{j \neq i} (\langle \mathbf{x}, \mathbf{y}_j \rangle - \psi_j)\}$ are open convex sets such that φ is differentiable in C_i, with $\nabla \varphi(\mathbf{x}) = \mathbf{y}_i$, $\mathbf{x} \in C_i$. The complement of $\bigcup_{i=1}^n C_i$ has Lebesgue measure zero. Thus, we can extend T to $\mathbf{x} \in \bigcup_{i=1}^n C_i$, hence to almost all $\mathbf{x} \in \mathbb{R}^d$, by setting $\overline{T}(\mathbf{x}) := \nabla \varphi(\mathbf{x})$.

By construction, $\mathbf{x}_i \in C_i$, hence $\overline{T}$ is an extension of T. Theorem 12.15 in Rockafellar and Wets (1998) implies that $\overline{T}$ is cyclically monotone. We could (in case $\bigcup_{i=1}^n C_i \subsetneq \mathbb{R}^d$) extend $\overline{T}$ from $\bigcup_{i=1}^n C_i$ to $\mathbb{R}^d$ while preserving cyclical monotonicity, but such extension of $\overline{T}$ cannot be continuous. Hence, we do not pursue that idea and, rather, try to find a smooth extension of T. For this, consider the Moreau envelopes

$$\varphi_\varepsilon(\mathbf{x}) := \inf_{\mathbf{y} \in \mathbb{R}^d} \left[\varphi(\mathbf{y}) + \frac{1}{2\varepsilon} \|\mathbf{y} - \mathbf{x}\|^2 \right], \quad \mathbf{x} \in \mathbb{R}^d, \, \varepsilon > 0 \tag{3.4}$$

of φ (as defined in (3.3)): see, e.g., Rockafellar and Wets (1998). The following theorem shows that, for sufficiently small $\varepsilon > 0$, $\nabla \varphi_\varepsilon$—the so-called Yosida regularization of $\nabla \varphi$ (Yosida 1964)—provides a continuous, cyclically monotone interpolation of $(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_n, \mathbf{y}_n)$, as desired.

Proposition 3.2 *Let Assumption (A) hold, and consider φ as in (3.3), with constants $\psi_1, \ldots, \psi_n$ satisfying (3.2). Let φ_ε as in (3.4). Then, there exists $e > 0$ such that, for every $0 < \varepsilon \leq e$, the map φ_ε is continuously differentiable and $T_\varepsilon := \nabla \varphi_\varepsilon$ is a continuous, cyclically monotone map such that $T_\varepsilon(\mathbf{x}_i) = \mathbf{y}_i$ for all $i = 1, \ldots, n$ and $\|T_\varepsilon(\mathbf{x})\| \leq \max_{i=1,\ldots,n} \|\mathbf{y}_i\|$ for all $\mathbf{x} \in \mathbb{R}^d$.*

The main conclusion of Proposition 3.2 (see Appendix F.2 for the proof) remains true in the setup of Remark 3.1, and we still can guarantee the existence of a convex, continuously differentiable φ such that $\nabla\varphi(\mathbf{x}_i) = \mathbf{y}_1$ for $i = 1, \ldots, n_0$ and $\nabla\varphi(\mathbf{x}_i) = \mathbf{y}_i$ for $i = n_0 + 1, \ldots, n$. More generally, the following corollary, which heuristically can be interpreted as a discrete version of the fact that a smooth convex function has a positive semi-definite second-order differential, is an immediate consequence.

Corollary 3.1 *Any cyclically monotone subset* $\{(\mathbf{x}_i, \mathbf{y}_i)|i = 1, \ldots, n\}$ *of* $\mathbb{R}^d \times \mathbb{R}^d$ *such that* $\mathbf{x}_i \neq \mathbf{x}_j$ *for* $i \neq j$ *lies in the subdifferential (at* $\mathbf{x}_i$, $i = 1, \ldots, n$*) of some (continuously) differentiable convex function* ψ.

Remark 3.2 It is important to note that, in spite of what intuition may suggest, and except for the univariate case ($d = 1$), linear interpolation does not work in this problem; see Remark F.1 in the appendix for a counterexample.

Remark 3.3 The interpolating function T_ε given by the proof of Proposition 3.2 is not only continuous but, in fact, Lipschitz with constant $1/\varepsilon$ (see, e.g., Exercise 12.23 in [108]). Looking for the smoothest possible interpolation we should, therefore, take the largest possible ε for which the interpolation result remains valid. Let us assume that $\|\mathbf{y}_i\| \leq 1$, $i = 1, \ldots, n$ (this does not imply any loss of generality; we could adequately normalize the data to get this satisfied, then backtransform the interpolating function). Set

$$\varepsilon_0 := \frac{1}{2} \min_{1 \leq i \leq n} \left((\langle \mathbf{x}_i, \mathbf{y}_i \rangle - \psi_i) - \max_{j \neq i}(\langle \mathbf{x}_i, \mathbf{y}_j \rangle - \psi_j) \right). \tag{3.5}$$

Then, arguing as in the proof of Proposition 3.2, we see that $B(\mathbf{x}_i, \varepsilon_0) \subset C_i$. Let $\varepsilon > 0$ and $\delta > 0$ be such that $\varepsilon + \delta < \varepsilon_0$. Then, for $\mathbf{x} \in B(\mathbf{x}_i, \delta)$, we have $\mathbf{x} - \varepsilon\mathbf{y}_i \in B(\mathbf{x}_i, \varepsilon_0)$, and we can mimic the argument in the proof to conclude that, for $\mathbf{x} \in B(\mathbf{x}_i, \delta)$, we have $\varphi_\varepsilon(\mathbf{x}) = \langle \mathbf{x}, \mathbf{y}_i \rangle - \psi_i - \frac{\varepsilon}{2}\|\mathbf{y}_i\|^2$, and, consequently, $T_\varepsilon(\mathbf{x}_i) = \mathbf{y}_i$ for every $\varepsilon < \varepsilon_0$ with ε_0 given by (3.5). By continuity of the Yosida regularization (see Theorem 2.26 in Rockafellar and Wets (1998)), we conclude that $T_{\varepsilon_0}(\mathbf{x}_i) = \mathbf{y}_i$, $i = 1, \ldots, n$. We summarize our findings in the following result.

Corollary 3.2 *Let Assumption (A) hold. Assume further that* $\|\mathbf{y}_i\| \leq 1$ *for all* $i = 1, \ldots, n$. *Let* $\varphi(\mathbf{x}) := \max_{1 \leq j \leq n}(\langle \mathbf{x}, \mathbf{y}_j \rangle - \psi_j)$ *with* $\psi_1, \ldots, \psi_n$ *defined as in (3.2),* φ_ε *as in (3.3), and* ε_0 *as in (3.5). Then* $T_{\varepsilon_0} := \nabla\varphi_{\varepsilon_0}$ *is a Lipschitz continuous, cyclically monotone map, with Lipschitz constant* $1/\varepsilon_0$, *such that* $T_{\varepsilon_0}(\mathbf{x}_i) = \mathbf{y}_i$, $i = 1, \ldots, n$ *and* $\|T_{\varepsilon_0}(\mathbf{x})\| \leq 1$ *for every* $\mathbf{x} \in \mathbb{R}^d$.

To conclude, let us turn to the choice of the weights ψ_i that satisfy condition (3.2), as required by our construction. In view of Corollary 3.2 and the discussion in Remark 3.3, choosing the weights that maximize ε_0 in (3.5) results in smoother interpolations. The optimal smoothing value then is half of the maximum in the linear program

$$\max_{\psi, \varepsilon} \varepsilon \qquad \text{s.t.} \quad \langle \mathbf{x}_i, \mathbf{y}_i - \mathbf{y}_j \rangle \geq \psi_i - \psi_j + \varepsilon, \quad i, j \in \{1, \ldots, n\}, \ i \neq j; \tag{3.6}$$

the optimal ψ_j's are the corresponding weights. The dual of (3.6) is

$$\min_{z_{i,j},\, i\neq j} \sum_{i,j=1,\ldots,n;\, i\neq j} z_{i,j}\langle \mathbf{x}_i, \mathbf{y}_i - \mathbf{y}_j\rangle \tag{3.7}$$

$$\text{s.t.} \sum_{j=1,\ldots,n;\, j\neq i} (z_{i,j} - z_{j,i}) = 0, \quad \sum_{i,j=1,\ldots,n;\, i\neq j} z_{i,j} = 1, \quad z_{i,j} \geq 0, \;\; i,j = 1,\ldots,n.$$

Now, (3.7) is a circulation problem over a complete graph with n vertices. By the Flow Decomposition Theorem (see, e.g., Theorem 3.5 and Property 3.6 in Ahuja et al. (1993)), any circulation is of the form $z_{i,j} = \sum_{W\in\mathcal{W}} \delta_{ij}(W)f(W)$ where $\mathcal{W}$ denotes the set of all cycles in the graph, $\delta_{ij}(W) = 1$ if the arc connecting i and j belongs to cycle W ($\delta_{ij}(W) = 0$ otherwise), and $f(W) \geq 0$ is the flow along cycle W. Writing $c_{i,j} = \langle \mathbf{x}_i, \mathbf{y}_i - \mathbf{y}_j\rangle$ and $c(W) = \sum_{i,j} \delta_{ij}(W)c_{i,j}$ (where $c(W)$ is the cost of moving one mass unit along the cycle W), the objective function in (3.7) takes the form

$$\sum_{i,j=1,\ldots,n;\, i\neq j} c_{i,j}z_{i,j} = \sum_{W\in\mathcal{W}} c(W)f(W),$$

with the constraint $\sum_{W\in\mathcal{W}} |W|f(W) = 1$ where $|W|$ denotes the length (number of arcs) in the cycle W. Putting $\tilde{f}(W) := |W|f(W)$, (3.7) can be rewritten as

$$\min_{\tilde{f}(W)} \sum_{W\in\mathcal{W}} \tilde{f}(W)\frac{c(W)}{|W|} \qquad \text{s.t.} \sum_{W\in\mathcal{W}} \tilde{f}(W) = 1, \quad \tilde{f}(W) \geq 0.$$

It follows that the optimal solution to (3.7) is $z_{i,j} = \delta_{ij}(\widehat{W})/|\widehat{W}|$, where $\widehat{W}$ is a *minimum mean cost cycle*, that is, a minimizer among all cycles of $c(W)/|W|$. The computation of the minimum mean cost cycle can be carried out in polynomial time using, for instance, Karp's algorithm (Karp (1978)). For this, we fix a vertex in the graph (vertex 1, say; this choice does not affect the final ouput) and write $d_{k,i}$ for the length of the shortest path from 1 to i in k steps (where the lentgh of the path $(i_1, i_2, \cdots, i_k)$ is $c_{i_1,i_2} + \cdots + c_{i_{k-1},i_k}$ and $d_{k,i} = +\infty$ if there is no path with k steps from 1 to i). The lengths $d_{k,i}$ for $0 \leq k \leq n$ and $1 \leq i \leq n$ can be computed recursively starting from $d_{0,1} = 0$, $d_{0,i} = \infty$ for $i \neq 1$, and $d_{k+1,i} = \min_j(d_{k,j} + c_{j,i})$ with $c_{i,i} = \infty$. Then, the minimum cycle mean is $\varepsilon^* = \min_{1\leq i\leq n}\max_{0\leq k\leq n-1}(d_{n,i} - d_{k,i})/(n-k)$, which can be computed in $O(n^3)$ steps (see Theorem 1 and subsequent comments in Karp (1978)). We observe that Assumption (A) is equivalent to $\varepsilon^* > 0$.

We still need to compute the optimal weights ψ_i. For this, we can consider the graph with modified costs $\tilde{c}_{i,j} := c_{i,j} - \varepsilon^*$ and compute the length $\tilde{d}_i$ of the shortest path (of any length) from vertex 1 to i. It is easy to see that a shortest path of length at most $(n-1)$ exists. Hence we can compute the shortest k-step distances $\tilde{d}_{k,i}$ as above, and $\tilde{d}_i = \min_{0\leq k\leq n-1} \tilde{d}_{k,i}$. Finally, we set $\psi = -\tilde{d}_i$. Now, by optimality, $\tilde{d}_j \leq \tilde{d}_i + \tilde{c}_{i,j}$, that is, $c_{i,j} \geq \psi_i - \psi_j + \varepsilon^*$. This shows that $(\psi_1,\ldots,\psi_n,\varepsilon^*)$ is an optimal solution to (3.6) which, moreover, can be computed in $O(n^3)$ computer time.

For $n = 2$, it is easily seen that the optimum in (3.7) (hence in (3.6)) is $\varepsilon_0 = \langle \mathbf{x}_1 - \mathbf{x}_2, \mathbf{y}_1 - \mathbf{y}_2\rangle/4 > 0$. The optimal weights can be chosen

as $\psi_i = \langle(\mathbf{x}_1 + \mathbf{x}_2), \mathbf{y}_i\rangle/2$, $i = 1, 2$. In the one-dimensional case, if $n = 2$, uniqueness of T holds iff $x_1 < x_2$ and $y_1 < y_2$. A simple computation yields

$$T_\varepsilon(x) = \begin{cases} y_1 & \text{for} & \big(x - (x_1 + x_2)/2\big)/\varepsilon & \leq y_1, \\ \big(x - (x_1 + x_2)/2\big)/\varepsilon & \text{for} & y_1 \leq & \big(x - (x_1 + x_2)/2\big)/\varepsilon & \leq y_2 \\ y_2 & \text{for} & y_2 \leq & \big(x - (x_1 + x_2)/2\big)/\varepsilon. \end{cases}$$

We see that T_ε is an extension of $x_i \mapsto y_i$, $i = 1, 2$ if and only if $x_2 - x_1 \geq -2\varepsilon y_1$ and $x_2 - x_1 \geq 2\varepsilon y_2$, which implies that $\varepsilon \leq (x_2 - x_1)/(y_2 - y_1)$—equivalently, $1/\varepsilon$ larger than or equal to $(y_2 - y_1)/(x_2 - x_1)$, the minimal Lipschitz constant of any Lipschitz extension of $x_i \mapsto y_i$. This yields, for $y_1 = -1$, $y_2 = 1$,

$$\varepsilon_0 = (x_2 - x_1)/2 = (y_2 - y_1)/(x_2 - x_1)$$

and T_{ε_0} is the Lipschitz extension of $x_i \mapsto y_i$ with minimal Lipschitz constant.

We now turn back to the smooth extension of the empirical center-outward distribution function $\mathbf{F}_\pm^{(n)}$ of Section 2.2. Proposition 3.2 (and subsequent comments in case $n_0 > 1$) allows us to extend $\mathbf{F}_\pm^{(n)}$ to a Lipschitz-continuous gradient of convex function over $\mathbb{R}^d$, denoted as $\overline{\mathbf{F}}_\pm^{(n)}$. The following result (proof in Appendix F.3) extends to $\overline{\mathbf{F}}_\pm^{(n)}$ the Glivenko-Cantelli result of Proposition 2.4. We state (and prove) it for the value ε_0 (3.5) of the smoothing constant; with obvious modifications, it also holds for any admissible ε.

Proposition 3.3 *(Glivenko-Cantelli) Let $\overline{\mathbf{F}}_\pm^{(n)}$ denote the smooth interpolation, with smoothing constant ε_0, of $\mathbf{F}_\pm^{(n)}$ computed from a sample of observations with distribution $\mathrm{P} \in \mathcal{P}_d^\pm$ and center-outward distribution function $\mathbf{F}_\pm$. Then,*

$$\sup_{\mathbf{x} \in \mathbb{R}^d} \|\overline{\mathbf{F}}_\pm^{(n)}(\mathbf{x}) - \mathbf{F}_\pm(\mathbf{x})\| \to 0 \quad \text{a.s. as } n \to \infty.$$

Remark 3.4 Throughout, we focused on a smooth interpolation of $\mathbf{F}_\pm^{(n)}$, applying Proposition 3.2 to the cyclically monotone n-tuple $\big(\mathbf{Z}_i^{(n)}, \mathbf{F}_\pm^{(n)}(\mathbf{Z}_i^{(n)})\big)$, $i = 1, \ldots, n$. For $n_0 \leq 1$ (or after implementing the tie-breaking device described in Section 2.2), the resulting $\overline{\mathbf{F}}_\pm^{(n)}$ is invertible, yielding a smooth interpolation— denote it as $\overline{\mathbf{Q}}_\pm^{(n)} := \big(\overline{\mathbf{F}}_\pm^{(n)}\big)^{-1}$—of the empirical quantile function $\mathbf{Q}_\pm^{(n)}$. For $n_0 > 1$, the restriction of $\overline{\mathbf{F}}_\pm^{(n)}$ to $\mathbb{R}^d \setminus \big(\mathbf{F}_\pm^{(n)}\big)^{-1}(\mathbf{0})$ (which has Lebesgue measure one) can be considered instead. In all cases, strong consistency still holds for $\overline{\mathbf{Q}}_\pm^{(n)}$; uniformity is lost, however, unless $\mathrm{spt}(\mathrm{P})$ itself is compact.

Remark 3.5 Another interpolation of $\mathbf{Q}_\pm^{(n)}$ is considered in Chernozhukov et al. (2017), based on the so-called α-*hull* method (see, e.g., Pateiro-López and Rodríguez-Casal (2010)). Although producing visually nice results (Figure 2, same reference), that method does not take into account any cyclical monotonicity constraints. The resulting contours therefore do not have the nature of quantile contours. Moreover, contrary to $\overline{\mathbf{Q}}_\pm^{(n)}$, the α-hull interpolation does not yield a homeomorphism; α-hull contours need not be closed, and the resulting quantile regions need not be connected: see Appendix H.4 for an example.

An alternative "multivariate step function" extension of $\mathbf{F}_{\pm}^{(n)}$ is proposed in Appendix G.

4. Some numerical results

This section provides some two-dimensional numerical illustrations of the results of this paper. The codes we used were written in R, and can handle sample sizes as high as $n = 20000$ (with $n_R = 100$ and $n_S = 200$, for instance) on a computer with 32Gb RAM. The algorithm consists of three main steps.

(Step 1) Determine the optimal assignment between the sample points and the regular grid. This could be done with a cubic implementation of the Hungarian algorithm like the one included in the `clue` R package (for a detailed account of the Hungarian algorithm and the complexity of different implementations, see, e.g., Chapter 4 in Burkhard et al. (2009)). Faster algorithms are available, though, as Bertsekas' *auction algorithm* or its variant, the *forward/reverse auction algorithm*, (Chapter 4 in Bertsekas (199), implemented in the R package `transport`. These auction algorithms depend on some parameter $\epsilon > 0$ and give in $O(n^2)$ time a solution to the assignment problem which is within $n\epsilon$ of being optimal. If the costs are integers and $n\epsilon < 1$, the solution given by the auction algorithm is optimal. Else, Step 2 below provides a check for the optimality of the solution given by the auction algorithm. If the check is negative, the algorithm is iterated with a smaller value of ϵ.

(Step 2) Compute the optimal value ε_0 of the regularization parameter and the optimal weights ψ_i. This is achieved via Karp's algorithm and the subsequent computation of shortest path distances as described in the discussion after Corollary 3.2. If $\varepsilon^* < 0$, then the solution of the assignment problem was not optimal and we return to Step 1 with a smaller value of ϵ. If not, we go to Step 3.

(Step 3) Compute the Yosida regularization based on a projected gradient descent method.

In Figure 1, we illustrate the convergence (as formulated by the Glivenko-Cantelli result of Proposition 3.3), of empirical contours to their population counterparts as the sample size increases. The problem is that analytical expressions for the population contours are not easily derived, except for spherical distributions. We therefore investigate the case of i.i.d. observations with bivariate $\mathcal{N}(0, \text{Id})$ distributions, and increasing samples sizes $n = 200, \ldots 10000$.

Inspection of Figure 1 clearly shows the expected consistency. Empirical contours are nicely nested, as they are supposed to be. For sample sizes as big as $n = 500$, and despite the fact that the underlying distribution is light-tailed, the .90 empirical contour still exhibits significant "spikes" out and in the theoretical circular contour. Those spikes reflect the intrinsic variability of an empirical quantile of order .90 based on about n_R observations; they rapidly

Figures 2–4 consider various Gaussian mixtures. Gaussian mixtures generate a variety of possibly multimodal and non-convex empirical dataclouds. In Figure 2, we simulated $n = 2000$ observations from a symmetric mixture of two spherical Gaussians. Figure 3 clearly demonstrates the quantile contour nature of our interpolations, as opposed to level contours. Level contours in

the right-hand panel clearly would produce disconnected regions separating the two modes of the mixture. Here, the contours remain nested—a fundamental monotonicity property of quantiles. The low-probability region between the two component populations is characterized by a "flat profile" of

Figure 3 similarly considers a mixture of three Gaussian distributions producing, in the central and right panels, distinctively nonconvex datasets. Picking that nonconvexity is typically difficult, and none of the traditional depth contours (most of them are intrinsically convex) are able to do it. Our interpolations do pick it, the inner contours much faster than the outer ones, as n increases. The very idea of a smooth interpolation indeed leads to bridging empty regions with nearly piecewise linear solutions. This is particularly clear with the .90 contour in the right-hand panel: the banana shape of the distribution is briefly sketched at the inception of the concave part, but rapidly turns into an essentially linear interpolation in the "central part of the banana". That phenomenon disappears as $n \to \infty$ and the "empty" regions eventually fill in.

Attention so far has been given to quantile contours, neglecting an important feature of center-outward quantile functions: being vector-valued, they also carry essential directional information. That information is contained in the empirical sign curves—the images, by the interpolated empirical quantile function, of the radii of the underlying regular grid. In the spherical case, those sign curves are quite uninformative and we did not plot them in Figures1 and 2. In the highly non-spherical Gaussian mixture of Figure 3, those sign curves are conveying an essential information.

Figure 4 is providing the full picture for $n = 20000$ (see also Figure 7 in Appendix H.1). The sign curves to the left and to the right of the vertical direction are vigorously combed to the left and the right. Since each curvilinear sector comprised between two consecutive sign curves roughly has the same probability contents, Figure 4 provides graphical evidence of a very low density in the central concavity bridged by the contours, thus producing a clear visualization of the banana shape of the dataset. Such figures, rather than contours alone, are the descriptive plots associated with empirical center-outward quantile functions. See Appendix H.1 for a comparison with Tukey depth.

5. Conclusions and perspectives

Unlike the earlier proposals, our concepts of distribution and quantile functions, ranks, and signs are satisfying the properties that make their univariate counterparts efficient and meaningful tools for statistical inference. In principle, they are paving the way to a solution of the long-standing open problem of distribution-free inference in multivariate analysis, offering a unique combination of strict distribution-freeness and semiparametric efficiency. A preliminary version (Hallin 2017) of this paper already triggered several important applications: De Valk and Segers (2018), Shi et al. (2019), Deb and Sen (2019), Ghosal and Sen (2019), Hallin, La Vecchia, and Liu (2019), Hallin, Hlubinka, and Hudecová (2020), ... A number of questions remain open, though. In particular, *(i)* Several issues remain to be studied about the concepts themselves: how in

finite samples should we choose the factorization of n into $n_R n_S + n_0$? should we consider cross-validation? how do those grids compare to random grids?

n=200   n=500   n=1000

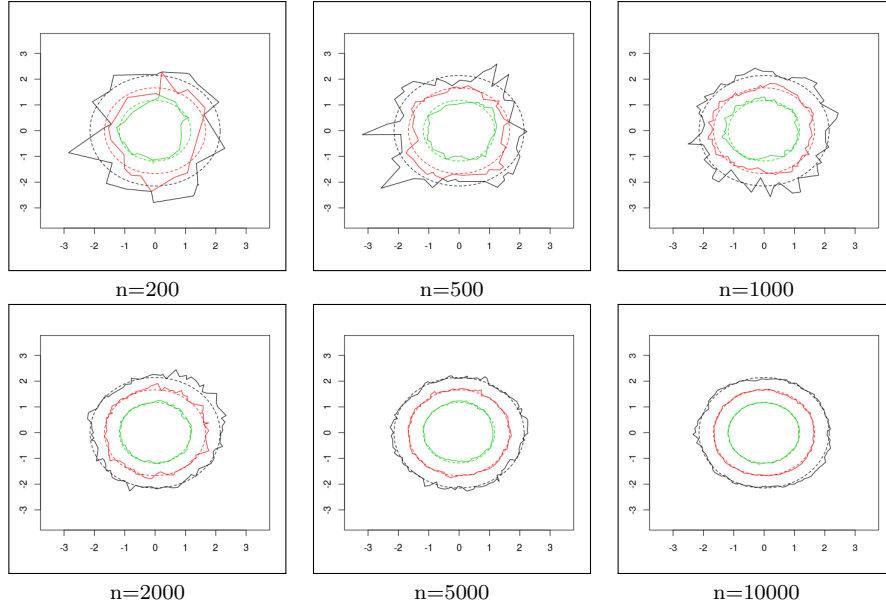n=2000   n=5000   n=10000

Fig 1: Smoothed empirical center-outward quantile contours (probability contents .50 (green), .75 (red), .90 (black)) computed from $n = 200, 500, 1000, 2000, 5000, 10000$ i.i.d. observations from a bivariate $\mathcal{N}(\mathbf{0}, \mathbf{I})$ distribution, along with their theoretical counterparts.

$$\frac{1}{2}\mathcal{N}\left(\begin{pmatrix} -1 \\ 0 \end{pmatrix}, \mathrm{Id}\right) + \frac{1}{2}\mathcal{N}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathrm{Id}\right) \qquad \frac{1}{2}\mathcal{N}\left(\begin{pmatrix} -2 \\ 0 \end{pmatrix}, \mathrm{Id}\right) + \frac{1}{2}\mathcal{N}\left(\begin{pmatrix} 2 \\ 0 \end{pmatrix}, \mathrm{Id}\right) \qquad \frac{1}{2}\mathcal{N}\left(\begin{pmatrix} -4 \\ 0 \end{pmatrix}, \mathrm{Id}\right) + \frac{1}{2}\mathcal{N}\left(\begin{pmatrix} 4 \\ 0 \end{pmatrix}, \mathrm{Id}\right)$$
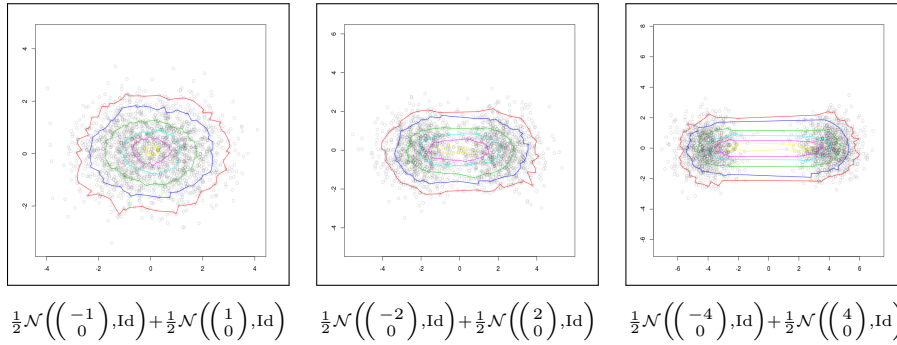
Fig 2: Smoothed empirical center-outward quantile contours (probability contents .02 (yellow), .20 (cyan), .25 (light blue) .50 (green), .75 (dark blue), .90 (red)) computed from $n = 2000$ i.i.d. observations from mixtures of two bivariate Gaussian distributions.

$$\frac{3}{8}\mathcal{N}(\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_1)+\frac{3}{8}\mathcal{N}(\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_2) \qquad \frac{3}{8}\mathcal{N}(-3\boldsymbol{\mu}_h,\boldsymbol{\Sigma}_1)+\frac{3}{8}\mathcal{N}(3\boldsymbol{\mu}_h,\boldsymbol{\Sigma}_2) \qquad \frac{3}{8}\mathcal{N}(-8\boldsymbol{\mu}_h,\boldsymbol{\Sigma}_1)+\frac{3}{8}\mathcal{N}(8\boldsymbol{\mu}_h,\boldsymbol{\Sigma}_2)$$
$$+\frac{1}{4}\mathcal{N}(\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_3) \qquad\qquad +\frac{1}{4}\mathcal{N}(-\tfrac{5}{2}\boldsymbol{\mu}_v,\boldsymbol{\Sigma}_3) \qquad\qquad +\frac{1}{4}\mathcal{N}(-5\boldsymbol{\mu}_v,\boldsymbol{\Sigma}_3)$$
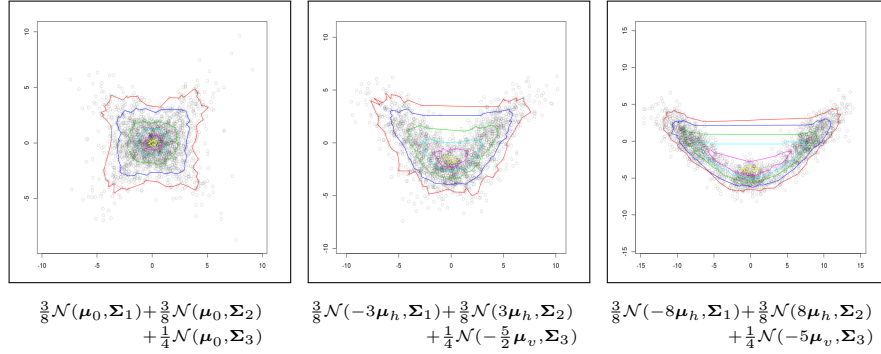
Fig 3: Smoothed empirical center-outward quantile contours (probability contents .02 (yellow), .20 (cyan), .25 (light blue) .50 (green), .75 (dark blue), .90 (red)) computed from $n = 2000$ i.i.d. observations from mixtures of three bivariate Gaussian distributions, with $\boldsymbol{\mu}_0 = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$, $\boldsymbol{\mu}_h = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$, $\boldsymbol{\mu}_v = \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$, $\boldsymbol{\Sigma}_1 = \left(\begin{smallmatrix} 5 & -4 \\ -4 & 5 \end{smallmatrix}\right)$, $\boldsymbol{\Sigma}_2 = \left(\begin{smallmatrix} 5 & 4 \\ 4 & 5 \end{smallmatrix}\right)$, $\boldsymbol{\Sigma}_3 = \left(\begin{smallmatrix} 4 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.
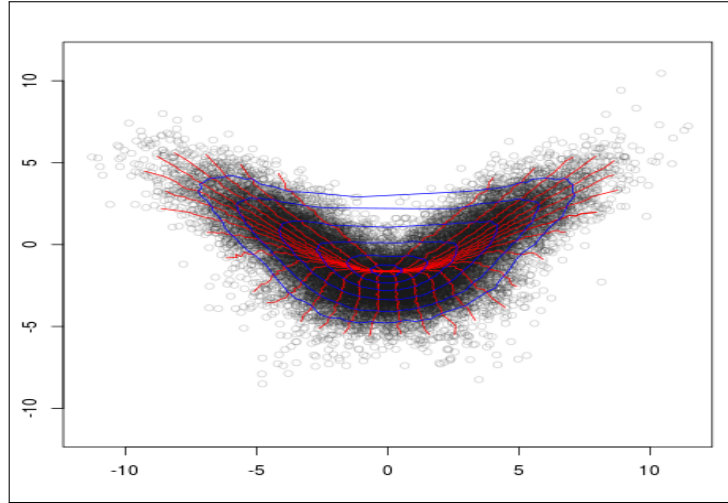
Fig 4: Center-outward quantile contours and sign curves for the same Gaussian mixture as in the middle panel of Figure 3, with $n = 20000$ (right).

(ii) How should we construct efficient rank tests in specific problems? Proposition C.1 suggests replacing, in the optimal test statistics derived under elliptic symmetry, the Mahalanobis ranks and signs with the center-outward ones. Can we similarly construct one-step R-estimators? This, which requires Hájek-type asymptotic representation results, would result in a fairly complete toolkit of distribution-free (hence "universally valid") semiparametrically efficient-at-

elliptical-densities rank-based procedures for multivariate analysis and multivariate time series.

(iii) Can goodness-of-fit tests be based, e.g. on Kolmogorov-Smirnov or Cramér-von Mises distances between center-outward distribution functions?

(iv) Turning to quantiles, what are the properties of $\mathbf{Q}_{\pm}^{(n)}(\mathbf{0})$ (for $n_0 \neq 0$) as a multivariate median? can we construct multivariate median or sign tests? can we, on the model of Carlier et al. (2016) or Hallin et al. (2010, 2015), perform multiple-output quantile regression? construct multivariate growthcharts (as in McKeague et al. (2011))? How?

(v) Center-outward quantile contours are obvious candidates as multivariate value-at-risk concepts, playing a central role in risk management; in that context, still in dimension $d = 1$, the primitives of ordinary distribution or quantile functions enter the definitions of a number of relevant notions such as Lorenz curves, average values at risk, or expected shortfall, see Gushchin and Borzykh (2017). The potential functions ψ and ϕ are natural multivariate extensions of those primitives, and likely to provide useful multivariate extensions.

(vi) What happens in high dimension $(d \to \infty)$? in functional spaces? on spheres (directional data) and other Riemannian manifolds?

Finally, these new empirical distribution and quantile functions are calling for a study of the corresponding empirical processes with further results such as Donsker and iterated logarithm theorems, or Bahadur representations.

References

[1] Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. (1993). *Network Flows*, Prentice-Hall.

[2] del Barrio, E., González-Sanz, A., and Hallin, M. (2019). A note on the regularity of center-outward distribution and quantile functions, arXiv:1912.10719.

[3] del Barrio, E. and Loubes, J.M. (2018). Central Limit Theorem for empirical transportation cost in general dimension, *Ann. Probab.* **47**, 926–951.

[4] Basu, D. (1955). On statistics independent of a complete sufficient statistic, *Sankhya* **15**, 377–380.

[5] Basu, D. (1959). The family of ancillary statistics, *Sankhyā Ser. A* **21**, 247–256.

[6] Belloni, A. and Winkler, R.L. (2011). On multivariate quantiles under partial orders. *Ann. Statist.* **39**, 1125–1179.

[7] Bertsekas, D. P. (1991). *Linear Network Optimization. Algorithms and Codes.* MIT Press, Cambridge, MA.

[8] Bickel, P.J. (1965). On some asymptotically nonparametric competitors of Hotelling's T^2, *Ann. Math. Statist.* **36**, 160-173.

[9] Biswas, M., Mukhopadhyay, M., and Ghosh, A.K. (2014). A distribution-free two-sample run test applicable to high-dimensional data, *Biometrika* **101**, 913–926.

[10] Boeckel, M., Spokoiny, V., and Suvorikova, A. (2018). Multivariate Brenier cumulative distribution functions and their application to nonparametric testing, arXiv:1809.04090.

[11] Burkhard, R., Dell'Amico, M., and Martello, S. (2009). *Assignment Problems*, SIAM.

[12] Chakraborty, A. and Chaudhuri, P. (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths, *Ann. Statist.* **42**, 1203–1231.

[13] Chakraborty, A. and Chaudhuri, P. (2017). Tests for high-dimensional data based on means, spatial signs and spatial ranks, *Ann. Statist.* **45**, 771–799.

[14] Chakraborty, B. and Chaudhuri, P. (1996). On a transformation-retransformation technique for constructing an affine equivariant multivariate median, *Proc. Amer. Math. Soc.* **124**, 2539–2547.

[15] Chakraborty, B. and Chaudhuri, P. (1998). On an adaptive transformation-retransformation estimate of multivariate location, *J. R. Stat. Soc. Ser. B* **60**, 145–157.

[16] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data, *J. Amer. Statist. Assoc.* **91**, 862–872.

[17] Chaudhuri, P. and Sengupta, D. (1993). Sign tests in multidimension : inference based on the geometry of the data cloud, *J. Amer. Statist. Assoc.* **88**, 1363–1370.

[18] Chernozhukov, V., Galichon, A., Hallin, M. and Henry, M. (2017). Monge-Kantorovich depth, quantiles, ranks, and signs, *Ann. Statist.* **45**, 223–256.

[19] Choi, K. and Marden, J. (1997). An approach to multivariate rank tests in multivariate analysis of variance, *J. Amer. Statist. Assoc.* **92**, 1581–1590.

[20] Cuesta-Albertos, J.A., Matrán, C., and Tuero-Diaz, A. (1997). Optimal transportation plans and convergence in distribution, *J. Multivariate Anal.* **60**, 72–83.

[21] Deb, N. and Sen, B. (2019). Multivariate rank-based distribution-free nonparametric testing using measure transportation, arXiv:1909.08733.

[22] De Valk, C. and Segers, J. (2018). Stability and tail limits of transport-based quantile contours, arXiv:1811.12061.

[23] Dick, J. and Pillichshammer, F. (2014). Discrepancy theory and quasi-Monte Carlo integration, in W. Chen, A. Srivastava and G. Travaglini, Eds, *A Panorama of Discrepancy Theory*, 539–620, Springer, New York.

[24] Figalli, A. (2017). *The Monge-Ampère Equation and its Applications*, Zurich Lectures in Advanced Mathematics, EMS, Zurich.

[25] Figalli, A. (2018). On the continuity of center-outward distribution and quantile functions, *Nonlin. Anal.: Theory, Methods & Appl.* **177**, 413-421.

[26] Genest, Chr. and Rivest, P.L. (2001). On the multivariate probability-integral transformation, *Statist. Probab. Lett.* **53**, 391–399.

[27] Ghosal, P. and Sen, B. (2019). Multivariate ranks and quantiles using optimal transportation and applications to goodness-of-fit testing, arXiv:1905.05340.

[28] Gushchin, A.A. and Borzykh, D.A. (2017). Integrated quantile functions: properties and applications, *Mod. Stoch. Theory Appl.* **4**, 285–314.

[29] Hallin, M. (2017). On distribution and quantile functions, ranks, and signs in $\mathbb{R}^d$, https://ideas.repec.org/p/eca/wpaper/2013-258262.html.

[30] Hallin, M., Hlubinka, D., and Hudecová, S. (2020). Efficient center-outward

rank tests for multiple-output regression, unpublished manuscript.

[31] Hallin, M., Ingenbleek, J-Fr., and Puri, M.L. (1989). Asymptotically most powerful rank tests for multivariate randomness against serial dependence. *J. Multivariate Anal.* **30**, 34–71.

[32] Hallin, M., La Vecchia, D., and Liu, H. (2019). Center-outward R-estimation for semiparametric VARMA models, arxiv.org/abs/1910.08442.

[33] Hallin, M. and Mehta, Ch. (2015). R-Estimation for asymmetric independent component analysis. *J. Amer. Statist. Assoc.* **110**, 218–232.

[34] Hallin, M., Oja, H., and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape II. Optimal R-estimation of shape, *Ann. Statist.* **34**, 2757–2789.

[35] Hallin, M. and Paindaveine, D. (2002a). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.* **30**, 1103–1133.

[36] Hallin, M. and Paindaveine, D. (2002b). Optimal procedures based on interdirections and pseudo-Mahalanobis ranks for testing multivariate elliptic white noise against ARMA dependence. *Bernoulli* **8**, 787–815.

[37] Hallin, M. and Paindaveine, D. (2002c). Multivariate signed ranks : Randles' interdirections or Tyler's angles? In *Statistical Data Analysis Based on the L_1 Norm and Related Procedures*, Y. Dodge, Ed., Birkhäuser, 271–282.

[38] Hallin, M. and Paindaveine, D. (2004a). Rank-based optimal tests of the adequacy of an elliptic VARMA model. *Ann. Statist.* **32**, 2642–2678.

[39] Hallin, M. and Paindaveine, D. (2004b). Multivariate signed rank tests in vector autoregressive order identification. *Statist. Sci.* **19**, 697–711.

[40] Hallin, M. and Paindaveine, D. (2005). Affine-invariant aligned rank tests for multivariate general linear models with ARMA errors, *J. Multivariate Anal.* **93**, 122–163.

[41] Hallin, M. and Paindaveine, D. (2006a). Semiparametrically efficient rank-based inference for shape: I Optimal rank-based tests for sphericity, *Ann. Statist.* **34**, 2707–2756.

[42] Hallin, M. and Paindaveine, D. (2006b). Parametric and semiparametric inference for shape: the role of the scale functional, *Statist. Decisions* **24**, 1001–1023.

[43] Hallin, M. and Paindaveine, D. (2008). Optimal rank-based tests for homogeneity of scatter, *Ann. Statist.* **36**, 1261-1298.

[44] Hallin, M., Paindaveine, D., and Šiman, M. (2010). Multivariate quantiles and multiple-output regression quantiles: from L^1 optimization to halfspace depth" (with discussion), *Ann. Statist.* **38**, 635–669.

[45] Hallin, M., Paindaveine, D., and Verdebout, Th. (2010). Optimal rank-based testing for principal components, *Ann. Statist.* **38**, 3245–3299.

[46] Hallin, M., Paindaveine, D., and Verdebout, Th. (2013). Optimal rank-based tests for common principal components, *Bernoulli* **19**, 2524–2556.

[47] Hallin, M., Paindaveine, D., and Verdebout, Th. (2014). Efficient R-estimation of principal and common principal components, *J. Amer. Statist. Assoc.* **109**, 1071–1083.

[48] Hallin, M. and Werker, B.J.M. (1998). Optimal testing for semiparametric

autoregressive models: from Gaussian Lagrange multipliers to regression rank scores and adaptive tests. In S. Ghosh, Ed., *Asymptotics, Nonparametrics, and Time Series*. M. Dekker, New York, 295–358.

[49] Hallin, M. and Werker, B.J.M. (2003). Semiparametric efficiency, distribution-freeness, and invariance, *Bernoulli* **9**, 137–165.

[50] Hamel, A. and Kostner, D. (2018). Cone distribution functions and quantiles for multivariate random variables, *J. Multivariate Anal.* **167**, 97–113.

[51] He, X. and Wang, G. (1997). Convergence of depth contours for multivariate datasets, *Ann. Statist.* **25**, 495–504.

[52] Hodges, J.L. (1955). A bivariate sign test, *Ann. Math. Statist.* **26**, 523-527.

[53] Ilmonen, P. and Paindaveine, D. (2011). Semiparametrically efficient inference based on signed ranks in independent component models, *Ann. Statist.* **39**, 2448–2476.

[54] Judd, K.L. (1998). *Numerical Methods in Economics*, MIT Press, Cambridge, MA.

[55] Karp, R. M. (1978). A characterization of the minimum cycle mean in a digraph. *Discrete Math.* **23**, 309–311.

[56] Koltchinskii, V. (1997). M-Estimation, convexity and quantiles, *Ann. Statist.* **25**, 435–477.

[57] Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*, Springer, New York.

[58] Lehmann, E.L. and Scholz, F.W. (1992). Ancillarity, in M. Ghosh and P.K. Pathak, Eds, *Issues in Statistical Inference: Essays in Honor of D. Basu*, IMS Lecture Notes - Monograph Series **17**, 32–51.

[59] Liu, R. Y. (1992). Data depth and multivariate rank tests, in Y. Dodge, Ed., L^1 *Statistics and Related Methods*, 279–294, North-Holland, Amsterdam.

[60] Liu, R.Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests, *J. Amer. Statist. Assoc.* **88**, 257–260.

[61] Lòpez-Pintado, S., and Romo, J. (2012). On the concept of depth for functional data, *J. Amer. Statist. Assoc.* **104**, 718–734.

[62] Marden, J. (1999). Multivariate rank tests, in S. Ghosh, Ed., *Design of Experiments and Survey Sampling*, M. Dekker, New York, 401–432.

[63] McCann, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps, *Duke Math. J.* **80**, 309–324.

[64] McKeague, I.W., López-Pintado, S., Hallin, M., and Šiman, M. (2011). Analyzing growth trajectories. *J. Dev. Or. Health and Disease* **2**, 322-329.

[65] Monge, G. (1781). *Mémoire sur la Théorie des Déblais et des Remblais*, *Mémoires Acad. Royale Sci.* **1781**, 666–704.

[66] Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *J. Nonparametr. Statist.* **5**, 201-213.

[67] Möttönen, J., Oja, H., and Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests, *Ann. Statist.* **25**, 542–552.

[68] Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, PA.

[69] Nordhausen, K., Oja, H., and Paindaveine, D. (2009). Signed-rank tests for

location in the symmetric independent component model. *J. Multivariate Anal.* **100**, 821–834.

[70] Oja, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates: a review. *Scand. J. Statist.* **26**, 319–343.

[71] Oja, H. (2010). *Multivariate Nonparametric Methods with R: an approach based on spatial signs and ranks*. Springer, New York.

[72] Oja, H., and Paindaveine, D. (2005). Optimal signed-rank tests based on hyperplanes, *J. Statist. Plann. Inference* **135**, 300–323.

[73] Oja, H. and Randles, R.H. (2004). Multivariate nonparametric tests. *Statist. Sci.* **19**, 598–605.

[74] Pateiro-López, B., and Rodríguez-Casal, A. (2010). Generalizing the convex hull of a sample: the R package α-hull. *J. Statist. Software* **34**.

[75] Peyré, G. and Cuturi, M. (2019). Computational optimal transport. *Foundations and Trends in Machine Learning* **11**, 355–607.

[76] Puri, M.L. and Sen, P.K. (1966). On a class of multivariate multisample rank-order tests (1966). *Sankhya A* **28**, 353–376.

[77] Puri, M.L. and Sen, P.K. (1967). On the theory of rank order tests for location in the multivariate one-sample problem. *Ann. Math. Statist.* **38**, 1216–1228.

[78] Puri, M.L. and Sen, P.K. (1969). A class of rank order tests for a general linear hypothesis. *Ann. Math. Statist.* **40**, 1325–1343.

[79] Puri, M.L. and Sen, P.K. (1971). *Nonparametric Methods in Multivariate Analysis*. John Wiley & Sons, New York.

[80] Randles, R.H. (1989). A distribution-free multivariate sign test based on interdirections, *J. Amer. Statist. Assoc.* **84**, 1045–1050.

[81] Rockafellar, R.T. (1966). Characterization of the subdifferential of convex functions, *Pacific J. Math.* **17**, 497–510.

[82] Rockafellar, R.T. and Wets, R.J.B. (1998). *Variational Analysis*, Springer.

[83] Santner, T.J., Williams, B.J. and Notz, W.I. (2003). *The Design and Analysis of Computer Experiments*, Springer-Verlag, New York.

[84] Segers, J., van den Akker, R., and Werker, B.J.M. (2014). Semiparametric Gaussian copula models: geometry and efficient R-estimation, *Ann. Statist.* **42**, 1911–1940.

[85] Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications, *Stat. Neerl.* **56**, 214–232.

[86] Shi, H., Drton, M., and Han, F. (2019). Distribution-free consistent independence tests via Hallin's multivariate ranks, arXiv:1909.10024.

[87] Um, Y. and Randles, R.H. (1998). Nonparametric tests for the multivariate multi-sample location problem. *Statist. Sinica* **8**, 801–812.

[88] Yosida, K. (1964). *Functional Analysis*, Springer Verlag.

[89] Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function, *Ann. Statist.* **28**, 461–482.

[90] Zuo, Y. (2018). On general notions of depth for regression, arXiv:1805.02046v1.

[91] Zuo, Y. and He, X. (2006). On limiting distributions of multivariate depth-based rank sum statistics and related tests, *Ann. Statist.* **34**, 2879–2896.

Appendix A: Measure transportation in a nutshell

A.1. Measure transportation, from Monge to McCann

Starting from a very practical problem—*How should one best move given piles of sand to fill up given holes of the same total volume?*—Gaspard Monge (1746-1818), with his 1781 *Mémoire sur la Théorie des Déblais et des Remblais*, initiated a profound mathematical theory anticipating different areas of differential geometry, linear programming, nonlinear partial differential equations, fluid mechanics, and probability.

In modern notation, the simplest and most intuitive—if not most general—formulation of Monge's problem is (in probabilistic form) as follows. Let P_1 and P_2 denote two probability measures over (for simplicity) $(\mathbb{R}^d, \mathcal{B}^d)$ and let $L : \mathbb{R}^{2d} \to [0, \infty]$ be a Borel-measurable loss function such that $L(\mathbf{x}_1, \mathbf{x}_2)$ represents the cost of transporting $\mathbf{x}_1$ to $\mathbf{x}_2$. The objective is to find a measurable (transport) map $T_{P_1;P_2} : \mathbb{R}^d \to \mathbb{R}^d$ solving the minimization problem

$$\inf_T \int_{\mathbb{R}^d} L\big(\mathbf{x}, T(\mathbf{x})\big) \mathrm{d}P_1 \qquad \text{subject to} \quad T \# P_1 = P_2 \qquad (A.8)$$

where T ranges over the set of measurable map from $\mathbb{R}^d$ to $\mathbb{R}^d$ and $T \# P_1$ is the so-called *push forward of* P_1 *by* T.[1] For simplicity, and with a slight abuse of language, we will say that T *is mapping* P_1 *to* P_2. A map $T_{P_1;P_2}$ achieving the infimum in (A.8) is called an *optimal transport map*, in short, an *optimal transport*, of P_1 to P_2. In the sequel, we shall restrict to the quadratic (or L^2) loss function $L(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2$.

The problem looks simple but it is not. Monge himself (who moreover was considering the more delicate loss $L(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\|_2$) did not solve it, and relatively little progress was made until the 1940s and the pathbreaking duality approach of Kantorovich. Relaxing the problem into

$$\inf_\gamma \int_{\mathbb{R}^d \times \mathbb{R}^d} L\big(\mathbf{x}, \mathbf{y}\big) \mathrm{d}\gamma(\mathbf{x}, \mathbf{y}) \qquad \text{subject to} \quad \gamma \in \Gamma(P_1, P_2) \qquad (A.9)$$

where $\Gamma(P_1, P_2)$ denotes the collection of all distributions over $\mathbb{R}^d \times \mathbb{R}^d$ with marginals P_1 and P_2, Kantorovich established that the solutions of (A.9) are of the form $\big(\text{Identity} \times T\big) \# P_1$ where T solves Monge's problem (A.8).

The topic attracted a renewed surge of interest some thirty years ago. For the L^2 transportation cost, Cuesta-Albertos and Matrán (1989) showed (under the assumption of finite second-order moments) the existence of solutions of the Monge problem and Rachev and Rüschendorf (1990) characterized them in terms of gradients of convex (potential) functions. Brenier (1991) with his celebrated *polar factorization theorem* independently obtained the same results and, moreover, established the (a.s.) uniqueness of the solution.

---

[1]In statistics, a more classical but heavier notation for $T \# P_1$ would be $P_1^{T\mathbf{X}}$ or $\overline{T}P_1$, where $\overline{T}$ is the transformation of $\mathcal{P}$ *induced by* T; see Chapter 6 of Lehmann and Romano (2005).

Measure transportation ever since has been among the most active domains of mathematical analysis, with applications in various fields, from fluid mechanics to economics (see Galichon (2016)), learning, and statistics (Carlier et al. (2016); Panaretos and Zemel (2016, 2018); Álvarez et al. (2018) and del Barrio et al. (2018)). It was popularized recently by the French Fields medalist Cédric Villani, with two authoritative monographs (Villani 2003, 2009), where we refer to for background reading, along with the two volumes by Rachev and Rüschendorf (1998), where the scope is somewhat closer to probabilistic and statistical concerns.

Whether described as in (A.8), or relaxed into the more general coupling form (A.9), the so-called Monge-Kantorovich problem remains an optimization problem, though, which only makes sense under densities for which expected costs are finite—under finite variances, thus, for quadratic loss. When defining concepts of distribution and quantile functions, ranks and signs, one clearly would like to avoid such assumptions. This is made possible thanks to a remarkable result by McCann (1995), hereafter *McCann's theorem*, the nature of which is geometric rather than analytical. Contrary to Monge, Kantorovitch, and Brenier, McCann (1995) does not require any moment restrictions and avoids using Kantorovich duality. McCann's main theorem implies that, for any given absolutely continuous distributions P_1 and P_2 over $\mathbb{R}^d$, there exists convex functions $\psi : \mathbb{R}^d \to \mathbb{R}$ with a.e. gradients[2] $\nabla\psi$ pushing P_1 forward to P_2; although ψ may not be unique, $\nabla\psi$ is P_1-a.s. uniquely determined[3]. Under the existence of finite moments of order two, $\nabla\psi$ moreover is with a L^2-optimal (in the Monge-Kantorovich sense) transport pushing P_1 forward to P_2.

A.2. Measure transportation, quantiles, and ranks: a review of the literature

Measure transportation ideas only recently made their way to statistical applications. Most of them are related to Wasserstein distance (see Panaretos and Zemel (2019)) and somewhat disconnected from the problems considered here. They are the basis, however, of Carlier et al. (2016)'s method of *vector quantile regression* and Chernozhukov et al. (2017)'s concept of *Monge-Kantorovich depth* and related quantiles, ranks and signs, two papers of which Ekeland et al. (2012) can be considered a precursor. While Carlier et al. (2016) consider mappings to the unit cube, Chernozhukov et al. (2017) deal with mappings to general reference distributions, including the uniform over the unit ball. On the other hand, they emphasize the consistent estimation of Monge-Kantorovich depth/quantile contours, with techniques that strongly exploit Kantorovich's duality approach and require compactly supported distributions, hence finite moments of all orders.

In the present paper, we privilege mappings to the (spherical) uniform over the unit ball, which enjoys better invariance/equivariance properties than the

---

[2]Recall that a convex function is a.e. differentiable.
[3]This means that, if ψ_1 and ψ_2 are convex and such that $\nabla\psi_1 \# P_1 = P_2 = \nabla\psi_2 \# P_1$, then $P_1 [\{\mathbf{x} : \nabla\psi_1(\mathbf{x}) \neq \nabla\psi_2(\mathbf{x})\}] = 0$.

unit cube—the latter indeed is not coordinate-free, and possesses edges and vertices, which are "very special points". Mappings to the unit ball naturally extend the structure of elliptical models, which is central to classical Multivariate Analysis, and is induced by linear sphericizing transformation—transports to spherical distributions. The same spherical structure also is the basis of the Mahalanobis ranks and signs approach developed in Hallin and Paindaveine (2002a, b, c, etc.). Adopting McCann's geometric point of view, we manage to waive moment assumptions which, as we already stressed, are not natural in the context. Moreover, we are focusing on the inferential virtues of ranks and signs, which are rooted in their independence with respect to the order statistic. The focus, applicability and decision-theoretic nature of our approach, in that respect, are quite different from Chernozhukov et al. (2017).

This paper results from merging two working papers, Hallin (2017) (essentially, Sections 1 and 2, with the Glivenko-Cantelli and Basu factorization results of Sections 2.3 and 2.4) and del Barrio et al. (2018) (essentially, Sections 3 and 4, with the cyclically monotone interpolation of Section 3, the extended Glivenko-Cantelli result of Proposition 3.3, and the numerical illustrations of Section 4).

Inspired by Chernozhukov et al. (2017), Boeckel et al. (2018) propose, under the name of ν-*Brenier distribution function* (ν a distribution over $\mathbb{R}^d$ with convex compact support[4]), a very general concept the empirical version of which satisfies a Glivenko-Cantelli property under compactly supported absolutely continuous distributions. Their empirical ν-Brenier distribution functions, however, are obtained by mapping the sample to an independent random sample of ν and therefore do not provide (even for $d = 1$) a neat interpretation in terms of ranks and signs. Yet another approach is taken in a recent paper by Faugeras and Rüschendorf (2018), who propose combining a mapping in the Chernozhukov et al. (2017) style with a preliminary copula transform. This copula transform takes care of the compact support/second-order moment restriction, but results in a concept that crucially depends on the original coordinate system.

The ideas developed in Chernozhukov et al. (2017) and Hallin (2017), on the other hand, have been successfully adopted by Shi, Drton, and Han (2019), who exploit the distribution-freeness properties of center-outward ranks in the construction of distribution-free tests of independence between random vectors (a long-standing open problem). Deb and Sen (2019) obtain similar results using different reference uniform distributions, different empirical transports, and different asymptotic techniques. In both cases, the key properties are distribution-freeness (and the Basu factorization property (DF^+) which, however, is not explicitly mentioned) of center-outward ranks. Ghosal and Sen (2019) also propose population concepts of distribution and quantile functions that are similar to those of Hallin (2017). Their empirical versions, however, are quite different, as their objective, contrary to this paper, is quantile reconstruction rather than a multivariate theory of rank-based inference. In particular, their empiri-

---

[4]The authors suggest the Lebesgue-uniform rather than the spherical uniform distribution over the unit ball.

cal distribution and quantile functions are based on semi-discrete transportation (pushing the empirical distribution of the sample forward to a continuous reference such as U_d or the Lebesgue-uniform over the unit cube). The resulting ranks and signs then are losing the distribution-freeness properties that are central to our approach. The computational benefit is that their empirical quantile functions, contours, and regions are obtained directly via the semi-discrete optimal transport[5] instead of cyclically monotone interpolation as in Section 3.

Recently, optimal *center-outward R-estimators* have been derived (Hallin, La Vecchia, and Lu 2019) for VARMA models, optimal *center-outward rank tests* are proposed by Hallin, Hlubinka, and Hudecova (2019) for multiple-output regression and MANOVA, while center-outward quantile-based methods for the measurement of multivariate risk are proposed in del Barrio, Beirlant, Buitendag, and Hallin (2019). Applications to the study of tail behavior and extremes can be found in De Valk and Segers (2018).

Appendix B: Distribution and quantile functions, ranks, and signs in dimension one

B.1. *Traditional univariate concepts*

The population and empirical concepts of distribution function, hence those of ranks, signs, order statistics, and quantiles, are well understood in dimension one. Before introducing multivariate extensions, let us briefly revisit some of their essential properties.

Let F denote the distribution function of a random variable Z with distribution $P \in \mathcal{P}_1$. It is well known that F is a *probability-integral transformation* ($Z \sim P_f$ iff $F(Z) \sim U_{[0,1]}$, where $U_{[0,1]}$ is the uniform over $[0,1]$), that is, in the terminology of measure transportation, F *pushes* P *forward to* $U_{[0,1]}$: $F\#P = U_{[0,1]}$.

Denote by $\mathbf{Z}^{(n)} := \left(Z_1^{(n)}, \ldots, Z_n^{(n)}\right)$ an n-tuple of random variables—observations or residuals associated with some parameter $\boldsymbol{\theta}$ of interest (see Section 1.1). Denoting by $R_i^{(n)}$ the rank of $Z_i^{(n)}$ among $Z_1^{(n)}, \ldots, Z_n^{(n)}$, the value at $Z_i^{(n)}$ of the empirical distribution $F^{(n)}$ of $\mathbf{Z}^{(n)}$ is $F^{(n)}(Z_i^{(n)}) := R_i^{(n)}/(n+1)$, where the denominator $(n+1)$ is adopted rather than n in order for $F^{(n)}(Z_i^{(n)})$ to take values in $(0,1)$ rather than $[0,1]$. Note that the mapping $Z_i^{(n)} \mapsto R_i^{(n)}/(n+1)$ is monotone nondecreasing from the sample to the regular grid

$$\{1/(n+1),\ 2/(n+1), \ldots, n/(n+1)\}. \tag{B.1}$$

The empirical distribution function $F^{(n)}$ then can be defined over $\mathbb{R}$ as an arbitrary non-decreasing interpolation of this discrete mapping. Usual practice is adopting a right-continuous step function interpolation, but that choice carries

---

[5]As a consequence, their empirical distribution functions only are continuous, while ours are at least Lipschitz-continuous (see Corollary 3.2) and the related discussion.

no information and has no particulat statistical justification: any other choice is as legitimate.

Intimately related with the concept of ranks is the *dual* concept of *order statistic* $\mathbf{Z}_{(.)}^{(n)} := \left(Z_{(1)}^{(n)}, \ldots, Z_{(n)}^{(n)}\right)$, with the rth order statistic $Z_{(r)}^{(n)}$, $r = 1 \ldots, n$ implicitly defined by $Z_{(R_i^{(n)})}^{(n)} = Z_i^{(n)}$, $i = 1, \ldots, n$. Assume that $\mathbf{Z}^{(n)}$ is an i.i.d. n-tuple with unspecified distribution $\mathrm{P} \in \mathcal{P}_1$. Then, $\mathbf{Z}_{(.)}^{(n)}$ is *minimal sufficient* and *complete* for f, while the vector $\mathbf{R}^{(n)} := \left(R_1^{(n)}, \ldots, R_n^{(n)}\right)$ of ranks is uniform over the $n!$ permutations of $\{1, \ldots, n\}$, hence distribution-free. Clearly, there is a one-to-one correspondence between $\left(\mathbf{Z}_{(.)}^{(n)}, \mathbf{R}^{(n)}\right)$ and $\mathbf{Z}^{(n)}$, so that (DF$^+$) follows from Basu's Theorem (see Section 1.1). The Glivenko-Cantelli theorem moreover tells us that, irrespective of the nondecreasing interpolation $F^{(n)}$ adopted,

$$\sup_{z \in \mathbb{R}} \left| F^{(n)}(z) - F(z) \right| \longrightarrow 0 \quad \text{a.s. as } n \to \infty \tag{B.2}$$

which, for $\mathrm{P} \in \mathcal{P}_1$, is equivalent to the apparently weaker forms

$$\sup_{z \in \mathrm{spt}(\mathrm{P})} \left| F^{(n)}(z) - F(z) \right| \longrightarrow 0 \quad \text{a.s. as } n \to \infty \tag{B.3}$$

where $\mathrm{spt}(\mathrm{P})$ denotes the support of P and

$$\max_{1 \le i \le n} \left| F^{(n)}(Z_i^{(n)}) - F(Z_i^{(n)}) \right| \longrightarrow 0 \quad \text{a.s. as } n \to \infty. \tag{B.4}$$

Finally, note that $\mathrm{P} \in \mathcal{P}_1$ (as well as F) is entirely characterized by the restriction of F to $\mathrm{spt}(\mathrm{P})$ and the fact that it is monotone nondecreasing (i.e., the gradient of a convex function).

B.2. *Univariate center-outward concepts*

The strong left-to-right orientation of the real line underlying the definition of F, the ranks, and $F^{(n)}$, however, cannot be expected to extend to dimension two and higher. For the purpose of multidimensional generalization, we therefore consider slightly modified concepts, based on a center-outward orientation. Define the *center-outward distribution function* of $Z \sim \mathrm{P}_f \in \mathcal{P}_1$ as $\mathbf{F}_{\pm} := 2F - 1$.

Clearly, being linear transformations of each other, F and $\mathbf{F}_{\pm}$ carry the same information about P_f. Just as F, $\mathbf{F}_{\pm}$ is a probability-integral transformation, now to the uniform distribution U_1 over the unit ball $\mathbb{S}_1 = (-1, 1)$: $\mathbf{F}_{\pm} \# \mathrm{P}_f = \mathrm{U}_1$.

Boldface is used in order to emphasize the interpretation of $\mathbf{F}_{\pm}$ as a vector-valued quantity: while $\|\mathbf{F}_{\pm}(z)\| = |2F(z) - 1|$ is the U_1-probability contents of the interval $(\pm\|\mathbf{F}_{\pm}(z)\|)$ (the one-dimensional ball with radius $\|\mathbf{F}_{\pm}(z)\|$), the unit vector $\mathbf{S}_{\pm}(z) := \mathbf{F}_{\pm}(z)/\|\mathbf{F}_{\pm}(z)\|$ ($\mathbf{S}_{\pm}(0)$ can be defined arbitrarily) is a direction (a point on the unit sphere $\mathcal{S}_0 = \{-1, 1\}$) or a sign—the sign of the deviation of z from the median $\mathrm{Med}(\mathrm{P}) := F^{-1}(1/2) = \mathbf{F}_{\pm}^{-1}(0)$ of P (possibly, an interval $[\mathrm{Med}^-(\mathrm{P}), \mathrm{Med}^+(\mathrm{P})]$ that does not intersect $\mathrm{spt}(\mathrm{P})$).

Inverting $\mathbf{F}_\pm$ yields a (possibly set-valued) *center-outward quantile function* $\mathbf{Q}_\pm$. The sets $\left\{\mathbf{Q}_\pm(u)\,\big|\,|u|\!=\!p\right\} = \left\{z_p^-, z_p^+\right\}$ and intervals $\left\{\mathbf{Q}_\pm(u)\,\big|\,|u| \leq p\right\} = \left[z_p^-, z_p^+\right]$, with z_p^- and z_p^+ such that $\mathrm{P}\!\left([z_p^-, \mathrm{Med}^-(\mathrm{P})]\right) = \mathrm{P}\!\left([\mathrm{Med}^+(\mathrm{P}), z_p^+]\right) = p/2$, accordingly have the interpretation of *quantile contours* and *quantile regions*, respectively, with quantile level (probability contents) $0 \leq p < 1$. Those quantile regions are closed, connected, and nested.

Turning to a sample $Z_1^{(n)}, \ldots, Z_n^{(n)}$ (with probability one, n distinct values), consider the $\lfloor n/2 \rfloor$ observations sitting to the right of the median. Ordering them from left to right yields ranks $R_{\pm;i}^{(n)}$, say, with values $1, \ldots, \lfloor n/2 \rfloor$; give them sign $\mathbf{S}_{\pm;i}^{(n)} = \mathbf{1}$ (the positive unit vector). Similarly rank the $\lfloor n/2 \rfloor$ observations sitting to the left of the median from right to left, obtaining ranks $R_{\pm;i}^{(n)}$; give them sign $\mathbf{S}_{\pm;i}^{(n)} = -\mathbf{1}$.[6] Call $R_{\pm;i}^{(n)}$ and $\mathbf{S}_{\pm;i}^{(n)}$ *center-outward ranks* and *signs*, respectively, and define the *empirical center-outward distribution function* as

$$\mathbf{F}_\pm^{(n)}(Z_i^{(n)}) := \mathbf{S}_{\pm;i}^{(n)} \frac{R_{\pm;i}^{(n)}}{\lfloor n/2 \rfloor + 1} = \begin{cases} 2F^{(n)}(Z_i^{(n)}) - 1 & n \text{ odd} \\[2mm] \dfrac{n+1}{n+2}\left(2F^{(n)}(Z_i^{(n)}) - 1\right) + \dfrac{1}{n+2} & n \text{ even,} \end{cases}$$
(B.5)

with values on the regular grids

$$\frac{-\lfloor n/2 \rfloor}{\lfloor n/2 \rfloor + 1}, \ldots, \frac{-2}{\lfloor n/2 \rfloor + 1}, \frac{-1}{\lfloor n/2 \rfloor + 1}, \ 0 \ , \frac{1}{\lfloor n/2 \rfloor + 1}, \frac{2}{\lfloor n/2 \rfloor + 1}, \ldots, \frac{\lfloor n/2 \rfloor}{\lfloor n/2 \rfloor + 1}$$

(n odd), and
(B.6)

$$\frac{-\lfloor n/2 \rfloor}{\lfloor n/2 \rfloor + 1}, \ldots, \frac{-2}{\lfloor n/2 \rfloor + 1}, \frac{-1}{\lfloor n/2 \rfloor + 1}, \ \frac{1}{\lfloor n/2 \rfloor + 1}, \frac{2}{\lfloor n/2 \rfloor + 1}, \ldots, \frac{\lfloor n/2 \rfloor}{\lfloor n/2 \rfloor + 1}$$

(n even), respectively. Those grids are the intersection of the two unit vectors $\mathbf{u} = \pm\mathbf{1}$ and the collection of $\lfloor n/2 \rfloor$ "circles" with center at the origin and radii $R/(\lfloor n/2 \rfloor + 1)$, $R = 1, \ldots, \lfloor n/2 \rfloor$—along ($n$ odd) with the origin itself.

Under the assumptions made, each sign $\mathbf{S}_{\pm;i}^{(n)}$ is uniform over the unit sphere $\mathcal{S}_0$, and independent of the ranks $R_{\pm;i}^{(n)}$; each rank is uniformly distributed over the integers $(0, 1, 2, \ldots, \lfloor n/2 \rfloor)$ or $(1, 2, \ldots, \lfloor n/2 \rfloor = n/2)$ according as n is odd or even; the n-tuple $\left(\mathbf{F}_\pm^{(n)}(Z_1^{(n)}), \ldots, \mathbf{F}_\pm^{(n)}(Z_n^{(n)})\right)$ is uniform over the $n!$ permutations of the grids (B.6).

In view of (B.5) and (B.6), the Glivenko-Cantelli result (B.4) for $F^{(n)}$ straightforwardly extends to $\mathbf{F}_\pm^{(n)}$:

$$\max_{1 \leq i \leq n} \left\| \mathbf{F}_\pm^{(n)}(Z_i^{(n)}) - \mathbf{F}_\pm(Z_i^{(n)}) \right\| \longrightarrow 0 \quad \text{a.s. as } n \to \infty \tag{B.7}$$

If $\mathbf{F}_\pm^{(n)}$ is to be defined over the whole real line, any nondecreasing interpolation $\overline{\mathbf{F}}_\pm^{(n)}$ of the n couples $(Z_i^{(n)}, \mathbf{F}_\pm^{(n)}(Z_i^{(n)}))$ provides a solution, all of them yielding Glivenko-Cantelli statements under $\sup_{z \in \mathbb{R}}$ or $\sup_{z \in \mathrm{spt}(\mathrm{P})}$ form (similar

---

[6]In case n is odd and the median is $Z_{i_0}^{(n)}$, put $\mathbf{S}_{\pm;i_0}^{(n)} = \mathbf{0}$ and $R_{\pm;i_0}^{(n)} = 0$.

to (B.2) and (B.3)). Among them is the continuous-from-the-left on the left-hand side of the (empirical) median, and continuous-from-the-right on the right-hand side of the median piecewise constant interpolation shown in Figure 5.

Clearly, the traditional ranks $R_i^{(n)}$ and the empirical center-outward values $\mathbf{F}_\pm^{(n)}(Z_i^{(n)})$, $i = 1, \ldots, n$, generate the same σ-field and both enjoy (DF$^+$): all classical rank statistics thus can be rewritten in terms of $\mathbf{F}_\pm^{(n)}$. Traditional ranks and center-outward ranks and signs, therefore, are strictly equivalent statistics.

B.3. Relation to measure transportation

The probability-integral transformation $z \mapsto \mathbf{F}_\pm(z)$ from $\mathbb{R}$ to the unit ball $\mathbb{S}_1$ (the interval $(-1, 1)$) is pushing $P_f \in \mathcal{P}_1$ forward to the uniform distribution U_1 over $\mathbb{S}_1$. As a continuous monotone non-decreasing function, it is the gradient (here, the derivative) of a convex function ψ_f, say, which, therefore, is everywhere continuously differentiable. Actually, it is the unique monotone function pushing P forward to U_d. It follows from McCann's theorem (see Section A.1) that $\nabla\psi_f$ coincides, P_f-almost surely—hence, over $\mathrm{spt}(P_f)$—with any monotone nondecreasing function (any gradient of a convex function) $\nabla\psi$ pushing P_f forward to U_1. It follows that $\psi_f - \psi$ is a constant, hence that $\nabla\psi_f = \nabla\psi$ everywhere. It follows that such a gradient is uniquely determined on $\mathrm{spt}(P_f)$, and that $\mathbf{F}_\pm$ on $\mathrm{spt}(P_f)$ can be characterized as the unique gradient of a convex function pushing P_f forward to U_1. The (noninformative) values of $\mathbf{F}_\pm$ outside $\mathrm{spt}(P_f)$ then are easily obtained by imposing monotonicity and range $[0, 1]$.

The huge advantage of this characterization is that it does not involve the canonical ordering of $\mathbb{R}$, hence readily extends to dimension $d \geq 2$. The extension, actually, would be entirely straightforward for distributions P_f with nonvanishing densities f (hence support $\mathbb{R}^d$). More general cases require some additional care with P_f-a.s. uniqueness, though—while the support of $P_f \in \mathcal{P}_1$ consists at most of a countable collection of intervals, the support of of $P_f \in \mathcal{P}_d$ is potentially much weirder. Everywhere continuous differentiability of the potential ψ, in particular, will not survive in higher dimension.

Appendix C: Center-outward and Mahalanobis ranks and signs

Recall that a d-dimensional random vector $\mathbf{X}$ has elliptical distribution $P_{\boldsymbol{\mu},\boldsymbol{\Sigma},f}$ with location $\boldsymbol{\mu} \in \mathbb{R}^d$, positive definite symmetric $d \times d$ scatter matrix $\boldsymbol{\Sigma}$ and radial density f iff $\mathbf{Z} := \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ has spherical distribution $P_{\mathbf{0},\boldsymbol{I},f}$, which holds iff $\mathbf{F}_{\mathrm{ell}}(\mathbf{Z}) := \mathbf{Z}F(\|\mathbf{Z}\|)/\|\mathbf{Z}\| \sim U_d$, where F, with density f, is the distribution function of $\|\mathbf{Z}\|$ (the *radial distribution* and *radial density*). Elliptical distributions with nonvanishing radial densities clearly belong to the class $\mathcal{P}_d^{\mathrm{conv}}$, with support $\mathbb{R}^d$.

The mapping $\mathbf{Z} \mapsto \mathbf{F}_{\mathrm{ell}}(\mathbf{Z})$ is a probability-integral transformation. Chernozhukov et al. (2017) show (Section 2.4) that it actually coincides with $\mathbf{Z}$'s center-outward distribution function $\mathbf{F}_\pm$. Letting $\mathbf{X}_i^{(n)}$, $i = 1, \ldots, n$ be i.i.d. with
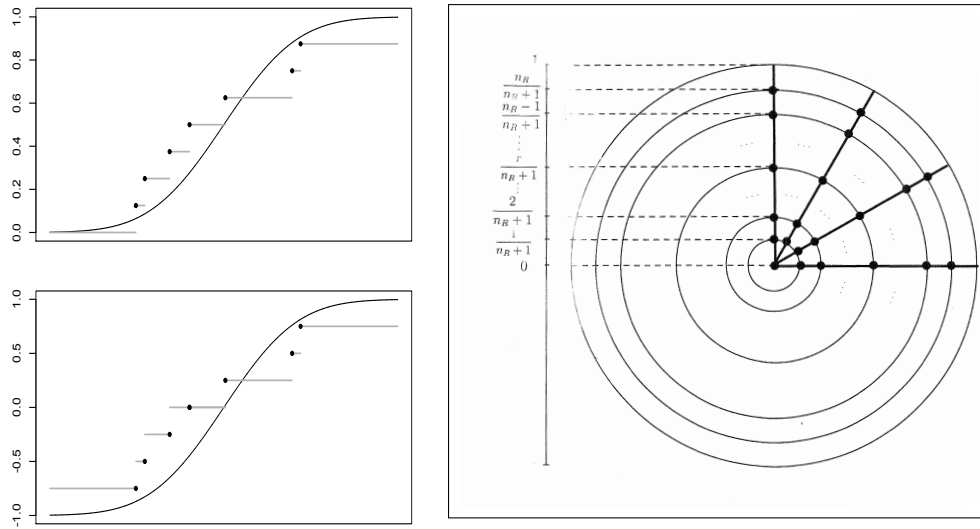
Fig 5: A classical distribution function F and its empirical counterpart $F^{(n)}$ for $n = 7$ (top left panel), along with (bottom left panel) their center-outward versions $\mathbf{F}_{\pm}$ and $\mathbf{F}_{\pm}^{(n)}$, the latter with left-continuous piecewise constant interpolation on the left-hand side of the (empirical) median, right-continuous piecewise constant interpolation on the right-hand side of the median; a regular grid of $n = n_R n_S$ points over $\mathbb{S}_2$ (right panel).

elliptical distribution $\mathrm{P}_{\boldsymbol{\mu},\boldsymbol{\Sigma},f}$, denote by $\hat{\boldsymbol{\mu}}^{(n)}$ and $\hat{\boldsymbol{\Sigma}}^{(n)}$ consistent estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively: the empirical version of $\mathbf{F}_{\text{ell}}$, based on Mahalanobis ranks and signs (the ranks $R_i^{(n)}$ of the residual moduli $\|\mathbf{Z}_i^{(n)}\| := \|\hat{\boldsymbol{\Sigma}}^{(n)-1/2}(\mathbf{X}_i^{(n)} - \hat{\boldsymbol{\mu}}^{(n)})\|$ and the corresponding unit vectors $\mathbf{Z}_i^{(n)}/\|\mathbf{Z}_i^{(n)}\|$) is, for the ith observation, $\mathbf{F}_{\text{ell}}^{(n)}(\mathbf{Z}_i^{(n)}) := \big(R_i^{(n)}/(n+1)\big)\mathbf{U}_i^{(n)}$.

Proposition C.1 *Let $\mathbf{X}_i^{(n)}$, $i = 1, \ldots, n$ be i.i.d. with elliptical distribution $\mathrm{P}_{\boldsymbol{\mu},\boldsymbol{\Sigma},f}$, and assume that $\hat{\boldsymbol{\mu}}^{(n)}$ and $\hat{\boldsymbol{\Sigma}}^{(n)}$ are strongly consistent estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively. Then, $\mathbf{F}_{ell}$ and $\mathbf{F}_{\pm}$ coincide, and*

$$\max_{1 \leq i \leq n} \|\mathbf{F}_{ell}^{(n)}(\mathbf{Z}_i^{(n)}) - \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)})\|, \quad \text{hence also} \quad \max_{1 \leq i \leq n} \|\mathbf{F}_{ell}^{(n)}(\mathbf{Z}_i^{(n)}) - \mathbf{F}_{\pm}(\mathbf{Z}_i^{(n)})\|$$

tend to zero a.s., as $n \to \infty$, where $\mathbf{F}_{\pm}$ denotes the center-outward distribution function of $\mathrm{P}_{\mathbf{0},\boldsymbol{I},f}$.

This result connects the center-outward ranks and signs with the well-studied elliptical ones. The consistency of $\mathbf{F}_{\text{ell}}^{(n)}$, however, requires ellipticity, whereas $\mathbf{F}_{\pm}^{(n)}$ remains consistent under any $\mathrm{P} \in \mathcal{P}_d^{\pm}$. Note also that $\mathbf{F}_{\text{ell}}^{(n)}$ determines n ellipsoidal contours, while $\mathbf{F}_{\pm}^{(n)}$ only determines n_R of them (which, moreover, for finite n do not define an ellipsoid).

Appendix D: Proofs for Section 2

D.1. Proofs of Propositions 2.1 and 2.2

Proof of Proposition 2.1. Part (i) is satisfied by construction (see the conclusion following (2.5)). For Part (ii), since $\mathbf{F}_{\pm}\#\mathrm{P} = \mathrm{U}_d$, the joint distribution of $\|\mathbf{F}_{\pm}(\mathbf{Z})\|$ and $\mathbf{S}(\mathbf{Z}) = \mathbf{F}_{\pm}(\mathbf{Z})/\|\mathbf{F}_{\pm}(\mathbf{Z})\|$ are those of $\|\mathbf{U}\|$ and $\mathbf{U}/\|\mathbf{U}\|$, where $\mathbf{U} \sim \mathrm{U}_d$; the claim follows. Turning to Part (iii), for any Borel set C of $\mathbb{R}^d$, we have $\mathrm{P}\big(C\big) = \mathrm{P}\big(C \cap \mathrm{spt}(\mathrm{P})\big)$. Now, the fact that the restrictions of $\mathbf{F}_{\pm}$ and $\mathbf{Q}_{\pm}$ to $\mathrm{spt}(\mathrm{P})$ and $\mathbb{S}_d$, respectively, are the inverse of each other, pushing P forward to U_d and U_d back to P, entails

$$\mathrm{P}\big(C \cap \mathrm{spt}(\mathrm{P})\big) = \mathrm{P}\big(\mathbf{Q}_{\pm} \circ \mathbf{F}_{\pm}(C \cap \mathrm{spt}(\mathrm{P}))\big) = \mathrm{U}_d\big(\mathbf{F}_{\pm}(C \cap \mathrm{spt}(\mathrm{P}))\big)$$
$$= \mathrm{U}_d\big(\mathbf{F}_{\pm}(C) \cap \mathbf{F}_{\pm}(\mathrm{spt}(\mathrm{P}))\big) = \mathrm{U}_d\big(\mathbf{F}_{\pm}(C) \cap \mathbb{S}_d\big);$$

the claim follows. Finally, Part (iv) readily follows from the fact that, in dimension $d = 1$, $2F - 1$ is the only monotone mapping from $\mathbb{R}$ to $\mathbb{S}_1 = (-1, 1)$ pushing $\mathrm{P} \in \mathcal{P}_1$ forward to $\mathrm{U}_1 = \mathrm{U}_{(-1,1)}$: See Appendix B.2. □

Proof of Proposition 2.2. Parts (i) and (ii) are direct consequences of the definition of $\mathbf{Q}_{\pm}$. As for Part (iii), it follows Proposition 2.1(iv) by adapting the traditional definition of a quantile function as a general inverse. □

D.2. Proofs of Proposition 2.3, Proposition 2.5, and Corollary 2.2

Proof of Proposition 2.3. Parts (i), (ii) and (iii) of the proposition are proved in del Barrio et al. (2019). Hence, we only have to prove the claims about $\mathbf{F}_\pm(\mathbf{x})$ for $\mathbf{x} \notin \mathrm{spt}(\mathrm{P})$. Since ϕ is a finite convex function on $\mathbb{R}^d$, it has a nonempty subgradient at every point. Let $\mathbf{x} \notin \mathrm{spt}(\mathrm{P})$ and consider $\mathbf{u} \in \partial\phi(\mathbf{x})$. Since ϕ is 1-Lipschitz, we have $\|\mathbf{u}\| \leq 1$. We claim that $\|\mathbf{u}\| = 1$. To show this, assume that, on the contrary, $\|\mathbf{u}\| < 1$. Then, from part (i) of the proposition, we have that $\mathbf{u} = \nabla\phi(\mathbf{x}_0)$ for some $\mathbf{x}_0 \in \mathrm{spt}(\mathrm{P})$. But this means that both $\mathbf{x}$ and $\mathbf{x}_0$ are in $\partial\phi^*(\mathbf{u})$. Then, by convexity, $(1-t)\mathbf{x} + t\mathbf{x}_0 \in \partial\phi^*(\mathbf{u})$ for every $t \in [0,1]$. Equivalently, $\mathbf{u} \in \partial\phi((1-t)\mathbf{x}+t\mathbf{x}_0)$ for every $t \in [0,1]$. Since $\mathrm{spt}(\mathrm{P})$ is open, this means that different points in $\mathrm{spt}(\mathrm{P})$ are mapped through $\nabla\phi$ to $\mathbf{u}$, contradicting the injectivity of $\nabla\phi$ in $\mathrm{spt}(\mathrm{P})\backslash K$. We conclude that, necessarily, $\|\mathbf{u}\| = 1$.

Next, let us assume that $\mathbf{x} \notin \mathrm{spt}(\mathrm{P})$ is such that $\mathbf{u_1} \neq \mathbf{u_2} \in \partial\phi(\mathbf{x})$. Then, for every $t \in [0,1]$, $(1-t)\mathbf{u_1} + t\mathbf{u_2} \in \partial\phi(\mathbf{x})$. But $\|(1-t)\mathbf{u_1} + t\mathbf{u_2}\| < 1$ for $t \in (0,1)$ unless $\mathbf{u_1} = \mathbf{u_2}$. This proves that $\partial\phi(\mathbf{x})$ is a singleton, hence that ϕ is differentiable at $\mathbf{x}$. The fact that the gradient of a convex function is continuous in the differentiability set completes the proof. $\quad\square$

Proof of Proposition 2.5. Part *(i)*. Sufficiency of $\mathbf{Z}^{(n)}_{(\cdot)}$—equivalently, sufficiency of the sub-σ-field $\mathcal{B}^{(n)}_{(\cdot)}$ of permutationally invariant[7] events of $\mathcal{B}^n_d$—follows from a trivial application of the classical Fisher-Neyman factorization criterion for dominated families (Corollary 2.6.1 in Lehmann and Romano (2005)). Completeness is established (under the name of *symmetric completeness*) in Lemma 3 of Bell et al. (1960)[8] and minimal sufficiency follows from the fact (see, e.g., Proposition 1.4.8 in Pfanzagl (2011)) that complete sufficient σ-fields are automatically minimal sufficient.

Part *(ii)*. Assume $n_0 = 0$ or 1. Conditionally on $\mathbf{Z}^{(n)}_{(\cdot)}$, $\mathbf{F}^{(n)}_\pm(\mathbf{Z}^{(n)})$ takes values in the set of the $n!$ permutations of the n gridpoints. Because of the permutational symmetry of the $\mathbf{Z}^{(n)}$ likelihood, all those values are equally likely, hence have conditional probability $1/n!$. Since that (uniform) conditional distribution does not depend on $\mathbf{Z}^{(n)}_{(\cdot)}$, it is also unconditional. If $n_0 > 1$, the situation is exactly the same, except that the $n_0!$ permutations involving the n_0 copies of the origin are undistinguishable, so that the $n!$ permutations of the grid reduce to $n!/n_0!$ permutations with repetitions. This, however, can be avoided by breaking the n_0 ties at the origin.[9]

---

[7]Permutation here means permutation among the n d-dimensional subspaces of $\mathcal{B}^n_d$.

[8]That lemma establishes completeness of the order statistic for nonparametric families of the form $\{\mathrm{P}^n | \mathrm{P} \ll \mathrm{P}_1\}$ where P_1 is non-atomic. In that notation, $\mathcal{P}_d = \{\mathrm{P}^n | \mathrm{P} \ll \mathcal{N}(\mathbf{0},\mathbf{I})\}$ where $\mathcal{N}(\mathbf{0},\mathbf{I})$ indeed is non-atomic; the result thus applies to $\mathcal{P}^{(n)}_d$.

[9]For instance, one may replace the n_0 copies of $\mathbf{0}$ with an i.i.d. n-tuple of gridpoints (distinct with probability one) simulated from a uniform over $[1/2(n_R + 1)]\mathbb{S}_d$ or a uniform over $[1/2(n_R + 1)]\mathcal{S}_{d-1}$. Uniformity (conditional on the simulation results) over the $n!$ permutations of the resulting n gridpoints, hence distribution-freeness, is recovered. Some of the resulting ranks, however, are losing their nature as integers—much in the same way as the *midranks* resulting from traditional univariate tie-breaking (see, e.g., Section III.8 in Hájek and Šidák (1967)).

Part *(iii)*: Assume $n_0 = 0$. The result readily follows from the uniformity of $\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}^{(n)})$ over the $n!$ permutations of the gridpoints, which are indexed by a product of the set of n_R integers $\{1, \ldots n_R\}$ and the set of n_S unit vectors. It no longer holds for $n_0 \geq 1$, even after performing the tie-break prodedure just described: indeed, $0 \leq R_{\pm,i}^{(n)} < 1$ tells something about $\mathbf{S}_{\pm,i}^{(n)}$. However, the proportion n_0/n of ties tends to zero as $n \to \infty$; moreover, the non-independence between ranks and (multivariate) signs has no decision-theoretic consequences as long as joint distribution-freeness holds.

Part *(iv)*, in view of *(i)* and *(ii)*, is an immediate consequence of the classical Basu theorem—Basu's Second Theorem in Appendix D.1 below.

Turning to Part *(v)*, either assume that $n_0 \leq 1$ or, for $n_0 > 1$, assume that the previously described tie-breaking grid randomization device has been performed, so that the grid does not exhibit any multiplicity at the origin. Then, the mapping $\mathbf{z} \mapsto \big(\mathbf{z}_{(\,.\,)}, \mathbf{F}_{\pm}^{(n)}(\mathbf{z})\big)$ is injective for $\mathbf{z} \in \mathbb{R}^{nd} \backslash N$. Distribution-freeness (Part *(ii)* of the proposition) entails the ancillarity of the σ-field $\mathcal{B}_{\pm}^{(n)}$ generated by $\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}^{(n)})$. In view of Corollary E.1, the completeness of the sufficient σ-field $\mathcal{B}_{(\,.\,)}^{(n)}$ generated by the order statistic $\mathbf{Z}_{(\,.\,)}^{(n)}$, and the ancillarity of $\mathcal{B}_{\pm}^{(n)}$, we only have to show that the σ-field $\sigma\big(\mathbf{Z}_{(\,.\,)}^{(n)}, \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}^{(n)})\big)$ is strongly $\mathcal{P}_d^{(n)}$-essentially equivalent to the Borel σ-field $\mathcal{B}^{nd}$. This readily follows, however, from the injectivity, over $\mathbb{R}^{nd} \backslash N$, of $\mathbf{z} \mapsto \big(\mathbf{z}_{(\,.\,)}, \mathbf{F}_{\pm}^{(n)}(\mathbf{z})\big)$. The claim follows. □

Proof of Corollary 2.2. It follows from the injectivity of the restriction to $\mathbb{R}^{nd} \backslash N$ of $\mathbf{z} \mapsto \big(\mathbf{z}_{(\,.\,)}, \mathbf{F}_{\pm}^{(n)}(\mathbf{z})\big)$ that the Borel σ-field $\mathcal{B}^{nd}$ is strongly $\mathcal{P}_d^{(n)}$-essentially equivalent (for some null set N) to $\sigma\big(\mathbf{Z}_{(\,.\,)}^{(n)}, \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}^{(n)})\big)$. For the same reason, $\mathcal{B}^{nd}$ is strongly $\mathcal{P}_d^{(n)}$-essentially equivalent (for some null set $\tilde{N}$) to $\sigma\big(\mathbf{Z}_{(\,.\,)}^{(n)}, \tilde{\mathbf{F}}_{\pm}^{(n)}(\mathbf{Z}^{(n)})\big)$. These strong essential equivalences still hold true with N and $\tilde{N}$ replaced with $M := N \cup \tilde{N}$. It follows that, for $\mathbf{z} \in \mathbb{R}^{nd} \backslash M$, a bijection exists between $\big(\mathbf{z}_{(\,.\,)}, \mathbf{F}_{\pm}^{(n)}(\mathbf{z})\big)$ and $\big(\mathbf{z}_{(\,.\,)}, \tilde{\mathbf{F}}_{\pm}^{(n)}(\mathbf{z})\big)$, hence between the permutations $\mathbf{F}_{\pm}^{(n)}(\mathbf{z})$ and $\tilde{\mathbf{F}}_{\pm}^{(n)}(\mathbf{z})$ of the two n-points grids. The result follows. □

Appendix E: Minimal sufficiency and maximal ancillarity

This appendix collects, for ease of reference, some classical and less classical definitions and results about sufficiency and ancillarity which are scattered across Basu's papers; some of them (such as the concept of *strong essential equivalence*) are slightly modified to adapt our needs.

The celebrated result commonly known as Basu's Theorem was first established as Theorem 2 in Basu (1955). The same paper also contains a Theorem 1, of which Theorem 2 can be considered a partial converse. Call them Basu's First and Second Theorems, respectively.

Proposition E.1 (Basu's First Theorem) *Let S be sufficient for a family $\mathcal{P}$ of distributions over some abstract space $(\mathcal{X}, \mathcal{A})$. Then, if a statistic W is P-independent of S for all $\mathrm{P} \in \mathcal{P}$, it is distribution-free over $\mathcal{P}$.*

Proposition E.2 (Basu's Second Theorem) *Let T be (boundedly) complete and sufficient for a family $\mathcal{P}$ of distributions over some abstract space $(\mathcal{X}, \mathcal{A})$. Then, if a statistic W is distribution-free over $\mathcal{P}$, it is P-independent of T for all $\mathrm{P} \in \mathcal{P}$.*

Basu's original proof of Proposition E.1 was flawed, however, and Basu's First Theorem does not hold with full generality. Basu (1958) realized that problem and fixed it by imposing on $\mathcal{P}$ a sufficient additional condition of *connectedness*. Some twenty years later, that condition has been replaced (Koehn and Thomas 1975) with a considerably weaker necessary and sufficient one (same notation as in Proposition E.1).

Proposition E.3 (Koehn and Thomas 1975) *Basu's First Theorem holds true if and only if $\mathcal{P}$ does not admit a measurable* splitting set, *namely, a set $A \in \mathcal{A}$ along with a partition $\mathcal{P} = \mathcal{P}_0 \oplus \mathcal{P}_1$ of $\mathcal{P}$ into two nonempty subsets $\mathcal{P}_0$ and $\mathcal{P}_1$ such that $\mathrm{P}(A) = 0$ for all $\mathrm{P} \in \mathcal{P}_0$ and $\mathrm{P}(A) = 1$ for all $\mathrm{P} \in \mathcal{P}_0$.*

Recall that a sub-σ-field $\mathcal{A}_0$ of $\mathcal{A}$ such that $\mathrm{P}_1(A) = \mathrm{P}_2(A)$ for all $A \in \mathcal{A}_0$ and all $\mathrm{P}_1, \mathrm{P}_2$ in $\mathcal{P}$ is called *ancillary*. Clearly, the σ-field $\mathcal{A}_V$ generated by a distribution-free statistic V is ancillary. Contrary to sufficient σ-fields (the smaller, the better), it is desirable for ancillary σ-field to be a large as possible. While minimal sufficient σ-fields, when they exist, are unique, maximal ancillary σ-fields typically exist, but are neither unique nor easily characterized—due, mainly, to null-sets issues.

Basu (1959) therefore introduced the notions of $\mathcal{P}$-essentially equivalent and $\mathcal{P}$-essentially maximal sub-σ-fields.

Definition E.1 *Two sub-σ-fields $\mathcal{A}_1$ and $\mathcal{A}_2$ of $\mathcal{A}$ are said to be $\mathcal{P}$-essentially equivalent if, for any $A_1 \in \mathcal{A}_1$, there exists an $A_2 \in \mathcal{A}_2$ and, for any $A_3 \in \mathcal{A}_2$, an $A_4 \in \mathcal{A}_1$ such that $\mathrm{P}(A_1 \Delta A_2) = 0 = \mathrm{P}(A_3 \Delta A_4)$ for any $\mathrm{P} \in \mathcal{P}$. An ancillary sub-$\sigma$-field essentially equivalent to a maximal ancillary sub-σ-field is called* essentially maximal.

The same reference then establishes the following sufficient condition for an ancillary statistic to be essentially maximal.

Proposition E.4 (Basu's Third Theorem) *Denote by $\mathcal{A}_{\mathrm{suff}}$ a (boundedly) complete and sufficient (for a family $\mathcal{P}$ of distributions over $(\mathcal{X}, \mathcal{A})$) sub-$\sigma$-field of $\mathcal{A}$. Then, any ancillary sub-σ-field $\mathcal{A}_{\mathrm{anc}}$ such that $\sigma\big(\mathcal{A}_{\mathrm{suff}} \cup \mathcal{A}_{\mathrm{anc}}\big) = \mathcal{A}$ is essentially maximal ancillary.*

Let us slightly reinforce Definition E.1 and the concepts of essentially equivalent and essentially maximal sub-σ-fields.

Definition E.2 *Two sub-σ-fields $\mathcal{A}_1$ and $\mathcal{A}_2$ of $\mathcal{A}$ are said to be* strongly $\mathcal{P}$-essentially equivalent *if there exists $N \in \mathcal{A}$ such that $\mathrm{P}(N) = 0$ for all $\mathrm{P} \in \mathcal{P}$ and $\mathcal{A}_1 \cap (\mathcal{X} \backslash N) = \mathcal{A}_1 \cap (\mathcal{X} \backslash N)$. An ancillary sub-$\sigma$-field strongly $\mathcal{P}$-essentially equivalent to a maximal ancillary sub-σ-field is called* strongly $\mathcal{P}$-essentially maximal.

Clearly, strong essential equivalence and maximal ancillarity imply essential equivalence and maximal ancillarity, respectively. The following slightly modified version of Basu's Third Theorem then readily follows.

Corollary E.1 *Denote by $\mathcal{A}_{\mathrm{suff}}$ a (boundedly) complete and sufficient, for a family $\mathcal{P}$ of distributions over $(\mathcal{X}, \mathcal{A})$, sub-$\sigma$-field of $\mathcal{A}$. Then, any ancillary sub-σ-field $\mathcal{A}_{\mathrm{anc}}$ such that $\sigma(\mathcal{A}_{\mathrm{suff}} \cup \mathcal{A}_{\mathrm{anc}})$ is strongly $\mathcal{P}$-essentially equivalent to $\mathcal{A}$ is strongly $\mathcal{P}$-essentially maximal ancillary.*

Appendix F: Proofs for Section 3

F.1. Proof of Proposition 3.1

Duality yields, for the linear program (3.1),

$$
\begin{aligned}
&\min_{\pi} \sum_{i=1}^{n} \sum_{j=1}^{n} c_{i,j} \pi_{i,j} & = \quad &\max_{a,b} \frac{1}{n} \sum_{i=1}^{n} a_i + \frac{1}{n} \sum_{j=1}^{n} b_j \\
&\text{s.t. } \sum_{i=1}^{n} \pi_{i,j} = \sum_{j=1}^{n} \pi_{i,j} = \frac{1}{n}, & &\text{s.t. } a_i + b_j \leq c_{i,j}, \; i,j = 1,\ldots,n. \\
&\quad \pi_{i,j} \geq 0, \; i,j = 1,\ldots,n
\end{aligned}
\tag{F.8}
$$

Moreover, $\pi = \{\pi_{i,j} \,|\, i,j = 1,\ldots,n\}$ is a minimizer for the left-hand side program, and $(a,b) = (a_1,\ldots,a_n,b_1,\ldots,b_n)$ a maximizer for the right-hand side one, if and only if they satisfy the corresponding constraints and

$$
\sum_{i=1}^{n} \sum_{j=1}^{n} c_{i,j} \pi_{i,j} = \frac{1}{n} \sum_{i=1}^{n} a_i + \frac{1}{n} \sum_{j=1}^{n} b_j.
$$

With the change of variables $a_i =: \|\mathbf{x}_i\|^2 - 2\varphi_i$, $b_j =: \|\mathbf{y}_j\|^2 - 2\psi_j$, the dual programs (F.8) take the form

$$
\begin{aligned}
&\max_{\pi} \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_{i,j} \langle \mathbf{x}_i, \mathbf{y}_j \rangle & = \quad &\min_{\varphi,\psi} \frac{1}{n} \sum_{i=1}^{n} \varphi_i + \frac{1}{n} \sum_{j=1}^{n} \psi_j \\
&\text{s.t. } \sum_{i=1}^{n} \pi_{i,j} = \sum_{j=1}^{n} \pi_{i,j} = \frac{1}{n}, & &\text{s.t. } \varphi_i + \psi_j \geq \langle \mathbf{x}_i, \mathbf{y}_j \rangle, \; i,j = 1,\ldots,n \\
&\quad \pi_{i,j} \geq 0, \; i,j = 1,\ldots,n
\end{aligned}
\tag{F.9}
$$

where π is a maximizer for the left-hand side program and (φ, ψ) a minimizer for the right-hand side one if and only if they satisfy the corresponding constraints and

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \pi_{i,j}\langle \mathbf{x}_i, \mathbf{y}_j \rangle = \frac{1}{n}\sum_{i=1}^{n}\varphi_i + \frac{1}{n}\sum_{j=1}^{n}\psi_j.$$

Let (φ, ψ) be a minimizer for the right-hand side program in (F.9). Then, replacing φ_i with $\tilde{\varphi}_i := \max_{j=1,\dots,n}(\langle \mathbf{x}_i, \mathbf{y}_j \rangle - \psi_j)$ yields a new feasible solution $(\tilde{\varphi}, \psi)$ satisfying $\varphi_i \geq \tilde{\varphi}_i$. Optimality of (φ, ψ) thus implies that $\varphi_i = \tilde{\varphi}_i$, so that, at optimality,

$$\varphi_i = \max_{j=1,\dots,n}(\langle \mathbf{x}_i, \mathbf{y}_j \rangle - \psi_j), \quad i = 1,\dots,n. \tag{F.10}$$

Now, if (3.1) is minimal, then $\pi_{i,i} = 1/n$, $\pi_{i,j} = 0$, $j \neq i$ is the unique maximizer in the left-hand side linear program in (F.9). Therefore, (φ, ψ) is a minimizer for the right-hand side program if and only if

$$\frac{1}{n}\sum_{i=1}^{n}(\varphi_i + \psi_i - \langle \mathbf{x}_i, \mathbf{y}_i \rangle) = 0.$$

In view of (F.10) this implies that

$$\langle \mathbf{x}_i, \mathbf{y}_i \rangle - \psi_i = \max_{j=1,\dots,n}(\langle \mathbf{x}_i, \mathbf{y}_j \rangle - \psi_j), \quad i = 1,\dots,n. \tag{F.11}$$

Conversely, assume that the weights $\psi_1, \dots, \psi_n$ are such that (F.11) holds. Then, letting $\varphi_i = \max_{j=1,\dots,n}(\langle \mathbf{x}_i, \mathbf{y}_j \rangle - \psi_j)$, we have that (φ, ψ) is a feasible solution for which

$$\frac{1}{n}\sum_{i=1}^{n}(\varphi_i + \psi_i - \langle \mathbf{x}_i, \mathbf{y}_i \rangle) = 0,$$

which, in view of the discussion above, implies that the map

$$T : \mathbf{x}_i \mapsto T(\mathbf{x}_i) = \mathbf{y}_i$$

is cyclically monotone. This completes the proof of Part *(i)* of the proposition.

As for Part *(ii)*, T is the unique cyclically monotone map from $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ to $\{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ if and only if, for any choice of indices $\{i_0, i_1, \dots, i_m\}$ in $\{1, \dots, n\}$, we have

$$\langle \mathbf{x}_{i_0}, \mathbf{y}_{i_0} - \mathbf{y}_{i_1} \rangle + \langle \mathbf{x}_{i_1}, \mathbf{y}_{i_0} - \mathbf{y}_{i_2} \rangle + \cdots + \langle \mathbf{x}_{i_m}, \mathbf{y}_{i_m} - \mathbf{y}_{i_0} \rangle > 0, \tag{F.12}$$

while (3.2) holds if and only if there exist real numbers $\psi_1, \dots, \psi_n$ such that

$$\langle \mathbf{x}_i, \mathbf{y}_i - \mathbf{y}_j \rangle > \psi_i - \psi_j \quad \text{for all } i \neq j.$$

On the other hand, defining $f_{i,j}(\psi) := \psi_i - \psi_j - \langle \mathbf{x}_i, \mathbf{y}_i - \mathbf{y}_j \rangle$ for $i \neq j$, we can apply Farkas' Lemma (see, e.g., Theorem 21.1. in Rockafellar (1970)) to see that

either there exists $\psi \in \mathbb{R}^n$ such that $f_{i,j}(\psi) < 0$ for all $i \neq j$ (equivalently, (3.2) holds), or there exist nonnegative weights $\lambda_{i,j}$, not all zero, such that

$$\sum_{i \neq j} \lambda_{i,j} f_{i,j}(\psi) \geq 0 \quad \text{for all } \psi \in \mathbb{R}^n.$$

Consider the graph with vertices $\{1, \ldots, n\}$ and (directed) edges corresponding to those pairs (i, j) for which $\lambda_{i,j} > 0$. There cannot be a vertex of degree one in the graph since, in that case, $\sum_{i \neq j} \lambda_{i,j} f_{i,j}(\psi)$ could not be bounded from below. Hence, the graph contains at least one cycle, that is, there exist $i_0, i_1, \ldots, i_m$ such that $\lambda_{i_0,i_1}, \lambda_{i_1,i_2}, \ldots,$ and λ_{i_m,i_0} all are strictly positive. Part *(i)* of the lemma then implies the existence of $\bar{\psi}_1, \ldots, \bar{\psi}_n$ such that $f_{i,j}(\bar{\psi}) \leq 0$ for all $i \neq j$. But then $0 \leq \sum_{i \neq j} \lambda_{i,j} f_{i,j}(\bar{\psi}) \leq 0$, which implies that $f_{i,j}(\bar{\psi}) = 0$ for each pair i, j with $\lambda_{i,j} > 0$, so that

$$f_{i_0,i_1}(\bar{\psi}) + f_{i_1,i_2}(\bar{\psi}) + \cdots + f_{i_m,i_0}(\bar{\psi}) = 0.$$

This in turn entails (observe that the sum $\bar{\psi}_i - \bar{\psi}_j$ along a cycle $i_0, i_1, \ldots, i_m, i_0$ vanishes)

$$\langle \mathbf{x}_{i_0}, \mathbf{y}_{i_0} - \mathbf{y}_{i_1} \rangle + \langle \mathbf{x}_{i_1}, \mathbf{y}_{i_1} - \mathbf{y}_{i_2} \rangle + \cdots + \langle \mathbf{x}_{i_m}, \mathbf{y}_{i_m} - \mathbf{y}_{i_0} \rangle = 0. \tag{F.13}$$

But (F.13) contradicts (F.12), which implies that if T is the unique cyclically monototone map from $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ to $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$, then (3.2) holds. Conversely, if (3.2) holds, then, for every cycle $i_0, i_1, \ldots, i_m, i_0$, we have

$$\langle \mathbf{x}_{i_0}, \mathbf{y}_{i_0} - \mathbf{y}_{i_1} \rangle + \langle \mathbf{x}_{i_1}, \mathbf{y}_{i_0} - \mathbf{y}_{i_2} \rangle + \cdots + \langle \mathbf{x}_{i_m}, \mathbf{y}_{i_m} - \mathbf{y}_{i_0} \rangle$$

$$> (\psi_{i_0} - \psi_{i_1}) + (\psi_{i_1} - \psi_{i_2}) + \cdots + (\psi_{i_m} - \psi_{i_0}) = 0,$$

and T is the unique cyclically monotone map from $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ to $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$. This completes the proof. □

F.2. *Proof of Proposition 3.2*

The map φ_ε is convex and continuously differentiable since φ is convex (see, e.g., Theorem 2.26 in Rockafellar and Wets (1998)). Hence $T_\varepsilon := \nabla\varphi_\varepsilon$ is a cyclically monotone, continuous map for every $\varepsilon > 0$. Setting

$$\tilde{\varepsilon}_0 = \min_{1 \leq i \leq n} \left((\langle \mathbf{x}_i, \mathbf{y}_i \rangle - \psi_i) - \max_{j \neq i}(\langle \mathbf{x}_i, \mathbf{y}_j \rangle - \psi_j) \right),$$

let $\varepsilon_0 = \frac{1}{2}\tilde{\varepsilon}_0 \min(1, 1/\max_{1 \leq i \leq n} \|\mathbf{y}_i\|)$. Note that $\tilde{\varepsilon}_0$, by (3.1), is strictly positive; hence, so is ε_0. If $\mathbf{x}$ lies in the ε_0-ball $B(\mathbf{x}_i, \varepsilon_0)$ centered at $\mathbf{x}_i$, then, if $j \neq i$,

$$\begin{aligned} \langle \mathbf{x}, \mathbf{y}_i \rangle - \psi_i &= \langle \mathbf{x}_i, \mathbf{y}_i \rangle - \psi_i + \langle \mathbf{x} - \mathbf{x}_i, \mathbf{y}_i \rangle > \langle \mathbf{x}_i, \mathbf{y}_j \rangle - \psi_j + \tilde{\varepsilon}_0 - \varepsilon_0\|\mathbf{y}_i\| \\ &\geq \langle \mathbf{x}_i, \mathbf{y}_j \rangle - \psi_j + \frac{1}{2}\tilde{\varepsilon}_0 \geq \langle \mathbf{x}, \mathbf{y}_j \rangle - \psi_j. \end{aligned}$$

This shows that $B(\mathbf{x}_i, \varepsilon_0) \subset C_i$ and

$$\varphi(\mathbf{x}) = \langle \mathbf{x}, \mathbf{y}_i \rangle - \psi_i, \quad \mathbf{x} \in B(\mathbf{x}_i, \varepsilon_0).$$

Assume now that

$$0 < \varepsilon \leq \frac{1}{2}\varepsilon_0 \min\left(1, \frac{1}{\max_{1 \leq i \leq n} \|\mathbf{y}_i\|}\right),$$

and let $\mathbf{x} \in B(\mathbf{x}_i, \varepsilon)$. The map $\mathbf{y} \mapsto \langle \mathbf{y}, \mathbf{y}_i \rangle - \psi_i + \frac{1}{2\varepsilon}\|\mathbf{y} - \mathbf{x}\|^2$ attains its global minimum at $\mathbf{y} = \mathbf{x} - \varepsilon\mathbf{y}_i \in B(\mathbf{x}_i, \varepsilon_0)$. For any $\mathbf{y}$, we have

$$
\begin{aligned}
\varphi(\mathbf{y}) + \frac{1}{2\varepsilon}\|\mathbf{y} - \mathbf{x}\|^2 &\geq \langle \mathbf{y}, \mathbf{y}_i \rangle - \psi_i + \frac{1}{2\varepsilon}\|\mathbf{y} - \mathbf{x}\|^2 \\
&\geq \varphi(\mathbf{x} - \varepsilon\mathbf{y}_i) + \frac{1}{2\varepsilon}\|\mathbf{x} - \varepsilon\mathbf{y}_i - \mathbf{x}\|^2 \\
&= \langle \mathbf{x}, \mathbf{y}_i \rangle - \psi_i - \frac{\varepsilon}{2}\|\mathbf{y}_i\|^2.
\end{aligned}
$$

This proves that

$$\varphi_\varepsilon(\mathbf{x}) = \langle \mathbf{x}, \mathbf{y}_i \rangle - \psi_i - \frac{\varepsilon}{2}\|\mathbf{y}_i\|^2, \quad \mathbf{x} \in B(\mathbf{x}_i, \varepsilon);$$

in particular, we conclude that $T_\varepsilon(\mathbf{x}_i) = \mathbf{y}_i$.

Turning to the last claim, note that

$$T_\varepsilon(\mathbf{x}) = \frac{1}{\varepsilon}(\mathbf{x} - \mathbf{y}_0),$$

where $\mathbf{y}_0$ is the unique minimizer of $\mathbf{y} \mapsto \varphi(\mathbf{y}) + \|\mathbf{y} - \mathbf{x}\|^2/2\varepsilon$ (again by Theorem 2.26 in Rockafellar and Wets (1998)). But $\mathbf{y}_0$ is such a minimizer if and only if $0 \in \partial\varphi(\mathbf{y}_0) + \frac{1}{\varepsilon}(\mathbf{y}_0 - \mathbf{x})$, that is, if and only if $T_\varepsilon(\mathbf{x}) \in \partial\varphi(\mathbf{y}_0)$, where $\partial\varphi(\mathbf{y}_0)$ denotes the subdifferential of φ at $\mathbf{y}_0$. Now (this is Theorem 25.6 in Rockafellar (1970)), for every $\mathbf{x} \in \mathbb{R}^d$, $\partial\varphi(\mathbf{x})$ is the closure of the convex hull of the set of limit points of sequences of the type $\nabla\varphi(\mathbf{x}_n)$ with $\mathbf{x}_n \to \mathbf{x}$. The map φ is differentiable in the regions C_i, with gradient $\mathbf{y}_i$. Hence, for every $\mathbf{x}$, $T_\varepsilon(\mathbf{x})$ belongs to the convex hull of $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$. This completes the proof, $\qquad\square$

Remark F.1 (Remark 3.2 continued) It is important to note that, in spite of what intuition may suggest, and except for the one-dimensional case ($d = 1$), linear interpolation does not work in this problem. Assume that $n \geq d + 1$ and that $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ are in general position. Denoting by $\mathcal{C}$ the convex hull of $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, there exists a partition of $\mathcal{C}$ into d-dimensional simplices determined by points in $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$: every point in $\mathcal{C}$ thus can be written in a unique way as a linear convex combination of the points determining the simplex it belongs to (with obvious modification for boundary points). Therefore, for all $\mathbf{x} \in \mathcal{C}$, there exist uniquely defined coefficients $\lambda_i^{\mathbf{x}} \in [0, 1]$, $i = 1, \ldots, n$,

with $\sum_i \lambda_i^{\mathbf{x}} = 1$ and $\#\{i|\lambda_i^{\mathbf{x}} \neq 0\} \leq d+1$, such that $\mathbf{x} = \sum_{i=1}^k \lambda_i^{\mathbf{x}} \mathbf{x}_i$. A "natural" linear interpolation of T on $\mathcal{C}$ would be $\mathbf{x} \mapsto \sum_{i=1}^k \lambda_i^{\mathbf{x}} \mathbf{y}_i$, $x \in \mathcal{C}$. For $d = 1$, this map is trivially monotone increasing, hence cyclically monotone. Starting with $d = 2$, however, this is no longer true, as the following counterexample shows. Let (for $d = 2$)

$$\mathbf{x}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \ \mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \ \mathbf{x}_3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

$$\mathbf{y}_1 = \begin{pmatrix} -5 \\ -.01 \end{pmatrix}, \ \mathbf{y}_2 = \begin{pmatrix} .5 \\ .01 \end{pmatrix}, \ \mathbf{y}_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

It is easily checked that the map $\mathbf{x}_i \mapsto \mathbf{y}_i$, $i = 1, 2, 3$ is the only cyclically monotone one pairing those points. Now, let us consider the points

$$\mathbf{x}_0 = .8\mathbf{x}_1 + .1\mathbf{x}_2 + .1\mathbf{x}_3 \quad \text{and} \quad \mathbf{y}_0 = .8\mathbf{y}_1 + .1\mathbf{y}_2 + .1\mathbf{y}_3.$$

The computation of all possible 24 pairings shows that the only cyclically monotone mapping between the sets $\{\mathbf{x}_0, \ldots, \mathbf{x}_3\}$ and $\{\mathbf{y}_0, \ldots, \mathbf{y}_3\}$ is

$$\mathbf{x}_0 \mapsto \mathbf{y}_2, \quad \mathbf{x}_1 \mapsto \mathbf{y}_1, \quad \mathbf{x}_2 \mapsto \mathbf{y}_0, \quad \mathbf{x}_3 \mapsto \mathbf{y}_3$$

where, obviously, $\mathbf{x}_0$ is not paired with $\mathbf{y}_0$ (nor $\mathbf{x}_2$ with $\mathbf{y}_2$).

F.3. Proof of Proposition 3.3

The proof of Proposition 3.3 relies on the following two preliminary propositions.

Proposition F.1 *Let $\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}$ be i.i.d. with distribution $P \in \mathcal{P}_d$ and denote by $\mu^{(n)}$ the corresponding empirical distribution. Then,*

$$\gamma^{(n)} := (\text{identity} \times \mathbf{F}_{\pm}^{(n)}) \# \mu^{(n)} \text{ converges weakly to } \gamma = (\text{identity} \times \mathbf{F}_{\pm}) \# P$$

as $n \to \infty$, $P - a.s.$, where $\mathbf{F}_{\pm}$ is P's center-outward distribution.

Proposition F.2 *Let $P \in \mathcal{P}_d^{\pm}$ have center-outward distribution function $\mathbf{F}_{\pm}$ and let $\mathbf{x}_n = \lambda_n \mathbf{u}_n$ with $0 < \lambda_n \to \infty$, $\|\mathbf{u}_n\| = 1$, and $\mathbf{u}_n \to \mathbf{u}$ as $n \to \infty$: then, $\mathbf{F}_{\pm}(\mathbf{x}_n) \to \mathbf{u}$.*

The proof of Proposition F.1 involves four lemmas, three from McCann (1995) and one from Rockafellar (1966), which we reproduce here for the sake of completeness. Throughout this section, μ and ν denote elements of the set $\mathcal{P}(\mathbb{R}^d)$ of all probability distributions on $\mathbb{R}^d$, $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ the set of all probability distributions on $\mathbb{R}^d \times \mathbb{R}^d$, and $\Gamma(\mu, \nu)$ the set of probability distributions in $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with given marginals μ and ν in $\mathcal{P}(\mathbb{R}^d)$. A measure γ in $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is said to have *cyclically monotone support* if there exists a cyclically monotone closed Borel set S in $\mathbb{R}^d \times \mathbb{R}^d$ such that $\gamma(S) = 1$.

Lemma F.1 (McCann 1995, Corollary 14). *Let $\mu,\ \nu \in \mathcal{P}(\mathbb{R}^d)$, and suppose that one of those two measures vanishes on all sets of Hausdorff dimension $d-1$. Then, there exists one and only one measure $\gamma \in \Gamma(\mu,\nu)$ having cyclically monotone support.*

Lemma F.2 (McCann 1995, Lemma 9). *Let $\gamma^{(n)} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ converge weakly as $n \to \infty$ to $\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. Then,*

(i) *if $\gamma^{(n)}$ has cyclically monotone support for all n, so does γ;*
(ii) *if $\gamma^{(n)} \in \Gamma(\mu^{(n)}, \nu^{(n)})$ where $\mu^{(n)}$ and $\nu^{(n)}$ converge weakly, as $n \to \infty$, to μ and ν, respectively, then $\gamma \in \Gamma(\mu,\nu)$.*

Lemma F.3 (McCann 1995, Proposition 10). *Suppose that $\gamma \in \Gamma(\mu,\nu)$ is supported on the subdifferential $\partial\psi$ of some convex function ψ on $\mathbb{R}^d$ (meaning that the support of γ is a subset of $\partial\psi$). Assume that μ vanishes on Borel sets of Hausdorff dimension $d-1$. Then, $\nabla\psi\#\mu = \nu$, that is, $\gamma = (\text{ identity} \times \nabla\psi)\#\mu$, where $(\text{ identity} \times \nabla\psi)\mathbf{x} := (\mathbf{x}, \nabla\psi(\mathbf{x}))$.*

Lemma F.4 (Rockafellar 1966, Theorem 1). *The subdifferential $\partial\psi$ of a convex function ψ on $\mathbb{R}^d$ enjoys cyclical monotonicity. Conversely, any cyclically monotone set S of $\mathbb{R}^d \times \mathbb{R}^d$ is contained in the subdifferential $\partial\psi$ of some convex function ψ on $\mathbb{R}^d$.*

This implies the existence of a gradient of convex function running through any n-tuple of cyclically monotone couples $((\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_n, \mathbf{y}_n)) \in \mathbb{R}^d \times \mathbb{R}^d$.

We now turn to the proof of Propositions F.1 and F.2.

Proof of Proposition F.1. Denote by $(\Omega, \mathcal{A}, \mathrm{P})$ the (unimportant) probability space underlying the observation of the sequence of $\mathbf{Z}_i^{(n)}$'s, $n \in \mathbb{N}$, by $\gamma^{(n)} = (\text{identity} \times \mathbf{F}_{\pm}^{(n)})\#\mu^{(n)}$ the empirical distribution, with marginals $\mu^{(n)}$ and $\mathrm{U}^{(n)}$, of the couples $(\mathbf{Z}_i^{(n)}, \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)}))$, and by $\gamma = (\text{identity} \times \mathbf{F}_{\pm})\#\mathrm{P}$ (with marginals P, U_d) the joint distribution of $(\mathbf{Z}, \mathbf{F}_{\pm}(\mathbf{Z}))$. Here, $\mu^{(n)}$, hence also $\gamma^{(n)}$, are random measures, with realizations $\mu_\omega^{(n)}$ and $\gamma_\omega^{(n)}$.

A sequence $\gamma_\omega^{(n)}$, $n \in \mathbb{N}$, is P-a.s. asymptotically tight since $\mu_\omega^{(n)}$ converges weakly to P with probability one and $\mathrm{U}^{(n)}$ has uniformly bounded support. By Prohorov's theorem, subsequences $\gamma_\omega^{(n_k)}$ can be extracted that converge weakly (to some γ_ω^∞'s). Those $\gamma_\omega^{(n_k)}$'s by construction have cyclically monotone supports, and their marginals $\mu_\omega^{(n_k)}$ and $\mathrm{U}^{(n_k)}$ converge weakly to P and U_d. Hence, by Lemma F.2, all limiting γ_ω^∞'s have cyclically monotone supports and marginals P and U_d, respectively.

In view of Lemma F.1, there exists only one γ with cyclically monotone support and marginals P and U_d. Hence, irrespective of the choice of the weakly converging subsequence $\gamma_\omega^{(n_k)}$, all limiting γ_ω^∞'s coincide with γ, which implies that the original sequence is converging weakly to γ. Moreover, that limit is the same for any ω in some $\Omega_1 \subseteq \Omega$ such that $\mathrm{P}(\Omega_1) = 1$.

Rockafellar's Theorem (Lemma F.4) provides a convex function ψ the subgradient of which contains the support of γ. Lemma F.3 and the definition of $\mathbf{F}_\pm$ conclude that $\gamma = (\text{identity} \times \nabla\psi)\#P = (\text{identity} \times \mathbf{F}_\pm)\#P$. $\square$

Proof of Proposition F.2. Being the gradient $\nabla\phi$ of a convex function, $\mathbf{F}_\pm = \nabla\phi$ is a monotone function (see, e.g., Rockafellar and Wets (1998)). The sequence $\nabla\phi(\mathbf{x}_n)$ is bounded. Taking subsequences if necessary, we can assume that $\nabla\phi(\mathbf{x}_n) \to \mathbf{y}$ for some $\mathbf{y}$ with $\|\mathbf{y}\| \leq 1$. By monotonicity, we have that

$$\langle \mathbf{x}_n - \mathbf{x}, \nabla\phi(\mathbf{x}_n) - \nabla\phi(\mathbf{x}) \rangle \geq 0$$

for every $\mathbf{x} \in \mathbb{R}^d$. In particular,

$$\langle \mathbf{x}_n - (\nabla\phi)^{-1}(\mathbf{w})), \nabla\phi(\mathbf{x}_n) - \mathbf{w} \rangle \geq 0$$

for every $\mathbf{w}$ with $0 < \|\mathbf{w}\| < 1$. But this means that

$$\langle \mathbf{u}_n - \tfrac{1}{\lambda_n}(\nabla\phi)^{-1}(\mathbf{w}), \nabla\phi(\mathbf{x}_n) - \mathbf{w} \rangle \geq 0$$

and, taking limits, that $\langle \mathbf{u}, \mathbf{y} - \mathbf{w} \rangle \geq 0$ for every $\mathbf{w}$ with $\|\mathbf{w}\| \leq 1$. From this we conclude that $\langle \mathbf{u}, \mathbf{y} \rangle \geq \|\mathbf{u}\|$. But, since $\|\mathbf{y}\| \leq 1$, this only can happen if $\mathbf{y} = \mathbf{u}$. $\square$

We now can proceed with the proof of Proposition 3.3.

Proof of Proposition 3.3. Denote by $\mathrm{U}_d^{(n)}$ the discrete probability measure assigning mass n_0/n to the origin and mass $1/n$ to the remaining points in the regular grid used for the definition of $\mathbf{F}_\pm^{(n)}$, and note that $\mathrm{U}_d^{(n)}$ converges weakly to U_d. Also write $\mathrm{P}^{(n)}$ for the empirical measure on $Z_1^{(n)}, \ldots, Z_n^{(n)}$. Over a probability one set Ω_0, say, of the underlying probability space Ω, the sequence $\mathrm{P}^{(n)}$ converges weakly to P. In the remainder of this proof, we tacitly assume that $\omega \in \Omega_0$. Note that $\overline{\mathbf{F}}_\pm^{(n)} = \nabla\phi_n$ for some convex ϕ_n, and that $\mathbf{F}_\pm = \nabla\phi$ with ϕ convex and continuously differentiable over $\mathbb{R}^d$. Recall, moreover, that $\nabla\varphi_n$, by construction, maps $\mathrm{P}^{(n)}$ to $\mathrm{U}_d^{(n)}$. By Theorem 2.8 in del Barrio and Loubes (2019), after subtracting centering constants if necessary, we can assume that $\phi_n(\mathbf{x}) \to \phi(\mathbf{x})$ for every $\mathbf{x} \in \mathrm{spt}(\mathrm{P})$. Actually, the statement of that result assumes convergence in transportation cost metric rather than weak convergence; the proof, however, only depends on the fact that, in that case, the sequence $\pi_n = (\text{identity} \times \nabla\varphi_n)\#\mathrm{P}^{(n)}$ converges weakly to $\pi = (\text{identity} \times \nabla\varphi)\#\mathrm{P}$, which, in view of Proposition F.1, holds here. We claim that, in fact, $\phi_n(\mathbf{x}) \to \phi(\mathbf{x})$ for every $\mathbf{x} \in \mathbb{R}^d$. To see this, first note that, by Proposition 3.2, $\|\overline{\mathbf{F}}_\pm^{(n)}(\mathbf{x})\| = \|\nabla\phi_n(\mathbf{x})\| \leq 1$ for every $\mathbf{x} \in \mathbf{R}^d$, which implies that the sequence $\{\phi_n\}$ is uniformly 1-Lipschitz, hence uniformly equicontinuous on $\mathbb{R}^d$. Also, since ϕ_n is pointwise convergent in $\mathrm{spt}(\mathrm{P})$, we can apply the Arzelà-Ascoli Theorem to conclude that we can extract a uniformly convergent subsequence over any compact subset of $\mathbb{R}^d$. By extracting a further subsequence, we can assume that $\phi_n \to \rho$ pointwise on all of $\mathbf{R}^d$ for some function ρ. This function must be convex and 1-Lipschitz (in particular, finite over

all $\mathbb{R}^d$) and, obviously, $\rho(\mathbf{x}) = \phi(\mathbf{x})$ for every $\mathbf{x} \in \mathrm{spt}(\mathrm{P})$. We note also that for every $\mathbf{u} \in \mathbb{S}_d$ there exists some $\mathbf{x} \in \mathrm{spt}(\mathrm{P})$ with $\mathbf{u} = \nabla\phi(\mathbf{x}) = \nabla\rho(\mathbf{x})$. Hence, for every $\mathbf{z}$, $\rho(\mathbf{z}) - \rho(\mathbf{x}) \geq \langle \mathbf{u}, \mathbf{z} - \mathbf{x} \rangle$. By duality,

$$\rho(\mathbf{x}) = \phi(\mathbf{x}) = \langle \mathbf{u}, \mathbf{x} \rangle - \phi^*(\mathbf{u}) = \langle \mathbf{u}, \mathbf{x} \rangle - \psi(\mathbf{u}).$$

This shows that

$$\rho(\mathbf{z}) \geq \sup_{u \in \mathbb{S}_d} \left(\langle \mathbf{u}, \mathbf{z} \rangle - \Psi(\mathbf{u}) \right) = \phi(\mathbf{z}), \quad \mathbf{z} \in \mathbb{R}^d. \tag{F.14}$$

To get an upper bound we note that, for every $\mathbf{x}$, $\mathbf{u}_n := \nabla\phi_n(\mathbf{x}) \in \mathbb{S}_d$. Since $\langle \mathbf{x}, \mathbf{u}_n \rangle = \phi_n(\mathbf{x}) + \phi_n^*(\mathbf{u}_n)$ we obtain that

$$\phi_n(\mathbf{x}) = \langle \mathbf{x}, \mathbf{u}_n \rangle - \phi_n^*(\mathbf{u}_n) \leq \sup_{\mathbf{u} \in \mathbb{S}_d} \left(\langle \mathbf{x}, \mathbf{u} \rangle - \phi_n^*(\mathbf{u}) \right) =: \tilde{\phi}_n(\mathbf{x}). \tag{F.15}$$

Now, $\tilde{\phi}_n$ is the convex conjugate of the function

$$\mathbf{u} \mapsto \tilde{\psi}_n(\mathbf{u}) = \begin{cases} \phi_n^*(\mathbf{u}) & \mathbf{u} \in \bar{\mathbb{S}}_d \\ \infty & \mathbf{u} \notin \bar{\mathbb{S}}_d. \end{cases}$$

Using Theorem 2.8 in del Barrio and Loubes (2019) again, we obtain that, for every $\mathbf{u} \in \bar{\mathbb{S}}_d$, $\phi_n^*(\mathbf{u}) \to \psi(\mathbf{u})$ and, consequently, for every $\mathbf{u} \in \mathbb{R}^d \setminus \mathcal{S}_{d-1}$, that $\tilde{\psi}_n(\mathbf{u}) \to \psi(\mathbf{u})$. Combining this with Theorems 7.17 and 11.34 in Rockafellar and Wets (1998), we conclude that $\tilde{\phi}_n(\varphi(\mathbf{x}) \to \psi^*(\mathbf{x}) = \phi(\mathbf{x})$ for every $\mathbf{x} \in \mathbf{R}^d$. Combined with (F.15), this shows that $\rho(\mathbf{x}) \leq \phi(\mathbf{x})$ which, along with (F.14), yields $\phi_n(\mathbf{x}) \to \phi(\mathbf{x})$. But then, Theorem 25.7 in Rockafellar (1970) implies that

$$\overline{\mathbf{F}}_{\pm}^{(n)}(\mathbf{x}) = \nabla\phi_n(\mathbf{x}) \to \nabla\phi(\mathbf{x}) = \mathbf{F}_{\pm}(\mathbf{x}), \quad \mathbf{x} \in \mathbf{R}^d,$$

uniformly over compact sets.

It only remains to show that uniform convergence holds over $\mathbb{R}^d$. For this, it suffices to show that, for every $\mathbf{w} \in \mathbb{R}^d$,

$$\sup_{\mathbf{x} \in \mathbb{R}^d} \left| \langle \left(\overline{\mathbf{F}}_{\pm}^{(n)}(\mathbf{x}) - \mathbf{F}_{\pm}(\mathbf{x}) \right), \mathbf{w} \rangle \right| \to 0. \tag{F.16}$$

Let us assume that, on the contrary, there exist $\varepsilon > 0$, $\mathbf{w} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$, and $\mathbf{x}_n \in \mathbb{R}^d$ such that

$$\left| \langle \left(\overline{\mathbf{F}}_{\pm}^{(n)}(\mathbf{x}_n) - \mathbf{F}_{\pm}(\mathbf{x}_n) \right), \mathbf{w} \rangle \right| > \varepsilon \tag{F.17}$$

for all n. The sequence $\mathbf{x}_n$ must be unbounded (otherwise (F.17) cannot hold). Hence, using compactness of the unit sphere and taking subsequences if necessary, we can assume that $\mathbf{x}_n = \lambda_n \mathbf{u}_n$ with $0 < \lambda_n \to \infty$, $\|\mathbf{u}_n\| = 1$, and $\mathbf{u}_n \to \mathbf{u}$ for some $\mathbf{u}$ with $\|\mathbf{u}\| = 1$. Again by compactness, we can assume that $\mathbf{F}_{\pm}(\mathbf{x}_n) \to \mathbf{y}$ and $\overline{\mathbf{F}}_{\pm}^{(n)}(\mathbf{x}_n) \to \mathbf{z}$. By Proposition F.2, we have that $\mathbf{y} = \mathbf{u}$. On the other hand, by monotonicity, for every $\mathbf{x} \in \mathbb{R}^d$,

$$\langle \overline{\mathbf{F}}_{\pm}^{(n)}(\mathbf{x}_n) - \overline{\mathbf{F}}_{\pm}^{(n)}(\mathbf{x}), \mathbf{x}_n - \mathbf{x} \rangle \geq 0.$$

Taking $\tau > 0$ and $\mathbf{x} = \tau \mathbf{u}_n$, we obtain that, if n is large enough (to ensure $\lambda_n > \tau$), then

$$\langle \overline{\mathbf{F}}_{\pm}^{(n)}(\mathbf{x}_n) - \overline{\mathbf{F}}_{\pm}^{(n)}(\tau \mathbf{u}_n), \mathbf{u}_n \rangle \geq 0.$$

We conclude that, for every $\tau > 0$

$$\langle \mathbf{z} - \mathbf{F}_{\pm}(\tau \mathbf{u}), \mathbf{u} \rangle \geq 0.$$

Now, we can take $\tau_n \to \infty$ and use Proposition F.2 to obtain that $\langle \mathbf{z} - \mathbf{u}, \mathbf{u} \rangle \geq 0$, that is, $\langle \mathbf{z}, \mathbf{u} \rangle \geq \|\mathbf{u}\|^2 = 1$. This, however, implies that $\mathbf{z} = \mathbf{u} = \mathbf{y}$, which contradicts (F.17), hence completes the proof. $\qquad \square$

Appendix G: A "multivariate step function" version of $\mathbf{F}_{\pm}^{(n)}$

Although, for $d = 1$, a smooth monotone increasing interpolation of the n-tuple $(X_i^{(n)}, F^{(n)}(X_i^{(n)}))$ in general provides a better approximation of F, empirical distribution functions are traditionally defined as right-continuous step functions—the exact opposite of smooth functions. Such step function interpolation yields some interpretational advantages in terms of the empirical measure of regions of the form $(-\infty, x]$, $x \in \mathbb{R}$. Still for $d = 1$, an outward-continuous center-outward counterpart can be defined in a very natural way, with interpretation in terms of the empirical measure of central regions of the form $[x-, x^+]$ where $[x^-, x_{1/2}^{(n)})$ and $(x_{1/2}^{(n)}, x^+]$ ($x_{1/2}^{(n)}$ an empirical median) contain the same number of observations: see Figure 5 in Appendix B.

A similar solution can be constructed for $d \geq 2$. Let $\overline{\mathbf{F}}_{\pm}^{(n)}$ be some smooth interpolation of $\mathbf{F}_{\pm}^{(n)}$. For any $r \in [0, 1]$ and $\mathbf{u}$ on the unit sphere $\mathcal{S}_{d-1}$, define $\lfloor r\mathbf{u} \rfloor_{n_R} := \lfloor (n_R + 1)r \rfloor \mathbf{u}/n_R + 1$. Then, $r\mathbf{u} \mapsto \lfloor r\mathbf{u} \rfloor_{n_R}$ maps an outward-open, inward-closed spherical annulus comprised in between two hyperspheres of the grid onto its inner boundary sphere while preserving directions. A "multivariate step function" version of the empirical center-outward distribution function $\mathbf{F}_{\pm}^{(n)}$, continuous from outward, can be defined as $\overline{\mathbf{F}}_{\pm}^{(n)*} := \lfloor \overline{\mathbf{F}}_{\pm}^{(n)} \rfloor_{n_R}$.

Instead of *steps*, those functions yield *plateaux* or *hyperplateaux*, the boundaries (equivalently, the discontinuity points) of which are the continuous *quantile contours* or *hypersurfaces* characterized by $\overline{\mathbf{F}}_{\pm}^{(n)}$. Those "quantile contours" present an obvious statistical interest. In contrast with the univariate case, this "step function version" of the empirical center-outward distribution function $\mathbf{F}_{\pm}^{(n)}$, for $d > 1$, is not unique, and depends on the smooth interpolation $\overline{\mathbf{F}}_{\pm}^{(n)}$ adopted. However, all its versions enjoy cyclical monotonicity and obviously satisfy the sup form of Glivenko-Cantelli: $\sup_{\mathbf{x} \in \mathbb{R}^d} \|\overline{\mathbf{F}}_{\pm}^{(n)*}(\mathbf{x}) - \mathbf{F}_{\pm}(\mathbf{x})\| \to 0$ a.s. as $n \to \infty$.

Appendix H: Further numerical results

H.1. *Center-outward quantiles and Tukey depth*

Statistical depth and our measure transportation approach are sharing the same ultimate objective of defining a concept of multivariate quantile. Some comparisons thus are quite natural—although not entirely straightforward, as we shall see, as the two concepts are of a different nature. The discussion below is restricted to Tukey depth, but similar conclusions hold for other depth concepts.
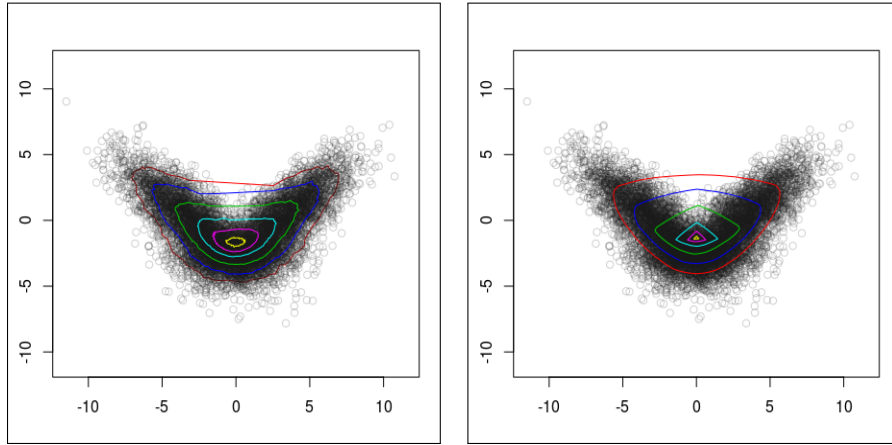

Fig 6: Center-outward quantile contours (left) and Tukey contours (right) for the same Gaussian mixture as in the middle panel of Figure 3, with $n = 10000$.

Whether theoretical or empirical, center-outward quantile functions and Tukey depth produce contours—in short, quantile contours and Tukey contours. For a spherical distribution with center $\boldsymbol{\mu}$, the family of population quantiles (indexed by their probability contents) and the family of population Tukey contours (indexed by depth) coincide (see Section 2.4 of Chernozhukov et al (2017)) with the family of (hyper)spheres centered at $\boldsymbol{\mu}$. Empirical quantile and empirical Tukey contours both consistently reconstruct those (hyper)spheres. As a rule, the empirical Tukey contours are smoother than the empirical quantile ones—although pairwise comparisons are difficult (or meaningless), as the probability contents of a Tukey contour (indexed by depth), unlike that of a quantile contour, depends on the underlying distribution. If smooth estimation of the *family* of population quantile contours were the objective, Tukey contours thus are doing a better job here. This is somewhat misleading, though—the (deterministic) family of (hyper)spheres centered at $\boldsymbol{\mu}$ is doing even better! But the objective here is not the smooth reconstruction of the *family* of quantile contours: we want something consistent that for finite n has the nature of an empirical quantile

function, which requires (cyclical) monotonocity properties that Tukey depth, even under spherical distributions, does not satisfy. And, of course, things only get worse under non-spherical densities.
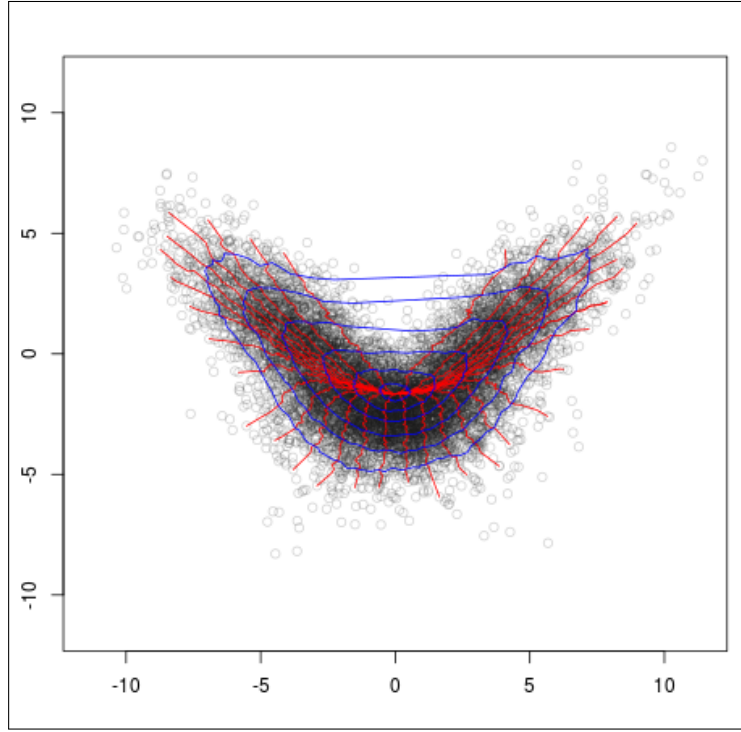

Fig 7: Center-outward quantile contours and sign curves for the same Gaussian mixture as in Figure 6, with $n = 20000$.

Now, let us have a closer look at the banana-shaped Gaussian mixtures of Figure 3. Figure 6 is providing, side by side, a plot of some quantile and Tukey contours for $n = 10000$. The concave shape of the distribution is only partially picked up by the outer quantile contours (left-hand panel)—despite of proven asymptotic concavity. The same concavity is not picked up at all (not even asymptotically so) by the Tukey contours (right-hand panel), which are inherently convex. Even worse, the inner Tukey contours display a misleading spike pointing upwards to the empty region. Despite of this, and although the theoretical weakness (lack of cyclical monotonicity) of Tukey contours as multivariate quantiles remain the same as previously discussed, one may feel that Tukey depth, as a descriptive tool, is doing almost as well, with less computational efforts, as empirical center-outward quantiles. As mentioned in Section 4, this is neglecting directional information contained in the empirical sign curves. Tukey depth, which is scalar-valued, has nothing equivalent to offer.

Figure 7 is providing the full picture for $n = 20000$. The sign curves to the

left and to the right of the vertical direction are very neatly combed to the left and to the right parts of the contours. Since each curvilinear sector comprised between two consecutive sign curves roughly has the same probability contents, Figure 7 provides evidence of a very low density in the central concavity bridged by the contours, thus producing a clear visualization of the banana shape of the dataset. Such figures, rather than contours alone, are the descriptive plots associated with empirical center-outward quantile functions.

Irrespective of the point of view adopted—be it inferential or data-analytical— center-outward quantile plots, thus, are carrying an information that Tukey depth plots cannot provide, which is well worth the additional computational effort.

H.2. Compact convex supports

All simulations in Section 4 have been conducted under $\mathbb{R}^d$-supported distributions. In this section, we consider two simple compactly supported cases.
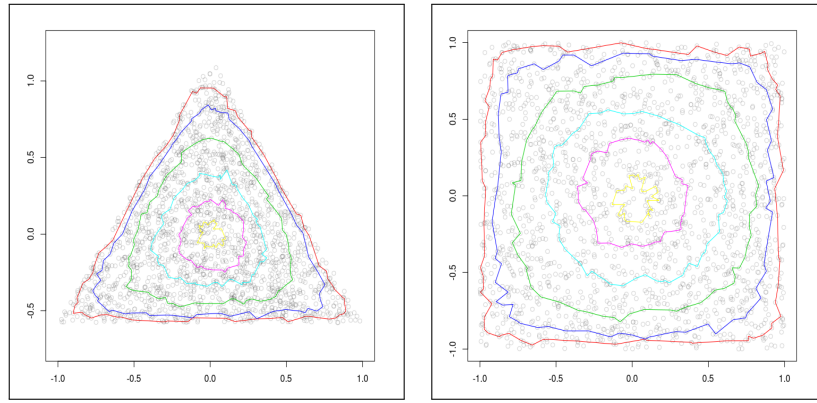
Fig 8: Smoothed empirical center-outward quantile contours (probability contents .50 (green), .75 (red), .90 (black)) computed from $n = 2000$ i.i.d. observations from Lebesgue-uniform distributions over the triangle and the square, respectively.

Figure 8 provides simulations for Lebesgue-uniforms with triangular and squared supports (sample size $n = 2000$, with $n_R = 50$ and $n_S = 40$), and shows how the contours evolve from nested circles in the center, where boundary effects are weak or absent, to nested triangles and squares as the boundary effects become dominant.

H.3. Disconnected supports

Figure 9 provides two independent simulations (sample size $n = 10000$) from Lebesgue-uniforms supported on two disconnected half-balls. Although the as-

sumptions for consistency are not satisfied, the contours and sign curves provide a very good description of the dataset, demonstrating, as in Figure 7, their complementarity: while the contours alone fail to disclose disconnectedness, that crucial feature of the dataset is fully revealed by the sign curves—an information that no depth concept can provide. Note also the (unsurprising) instability of the median set in such case; the same instability would occur in dimension one with a density supported on two disjoint intervals of equal probability 1/2, due to the lack of injectivity of the distribution function.
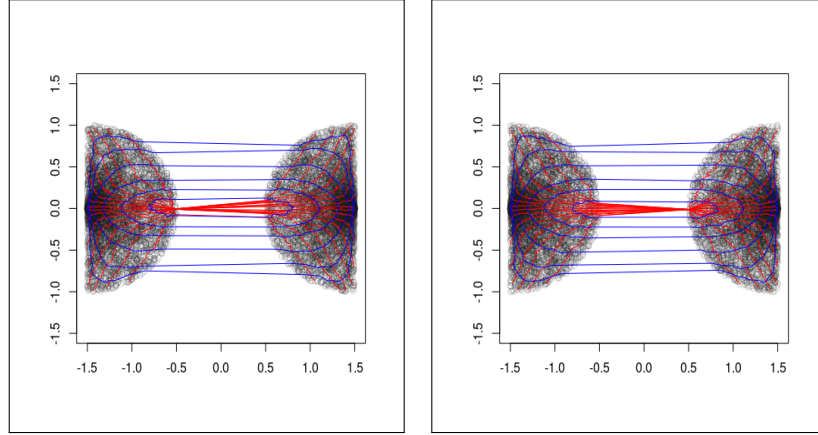
Fig 9: Smoothed empirical center-outward quantile contours and sign curves computed from $n = 10000$ i.i.d. observations from a Lebesgue-uniform distribution over two half balls (two independent simulations, showing the instability of the median set).

H.4. A non-connected α-hull contour

In this section, we provide an example of the dangers attached with the so-called α-hull interpolation, as considered in Chernozhukov et al. (2017).

Consider the six points

$$\mathbf{x}_1 = \begin{pmatrix} -2 \\ -1/2 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} -1 \\ -1/2 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} -3/2 \\ \frac{\sqrt{3}}{2} - 1/2 \end{pmatrix},$$

$$\mathbf{x}_4 = \begin{pmatrix} 1 \\ -1/2 \end{pmatrix}, \quad \mathbf{x}_5 = \begin{pmatrix} 2 \\ -1/2 \end{pmatrix}, \quad \mathbf{x}_6 = \begin{pmatrix} 3/2 \\ \frac{\sqrt{3}}{2} - 1/2 \end{pmatrix}.$$

Note that $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ and $\mathbf{x}_4$, $\mathbf{x}_5$, $\mathbf{x}_6$ are the vertices of two equilateral triangles with sides of length one; denote them as $\mathcal{A}$ and $\mathcal{B}$, respectively.

The complement of the α-hull of the set $\mathcal{X} := \{\mathbf{x_1}, \ldots, \mathbf{x_6}\}$ is defined as the union of all open balls of radius α that have empty intersection with $\mathcal{X}$.

Put $\alpha = 3/2$. In order for its intersection with $\mathcal{X}$ to be empty, a ball of radius α must be centered at distance at least α from each point in $\mathcal{X}$. Clearly, any point outside the triangles $\mathcal{A}$ and $\mathcal{B}$ belongs to some ball of radius α that does not intersect $\mathcal{X}$; hence, the α-hull of $\mathcal{X}$ is contained in $\mathcal{A} \cup \mathcal{B}$.
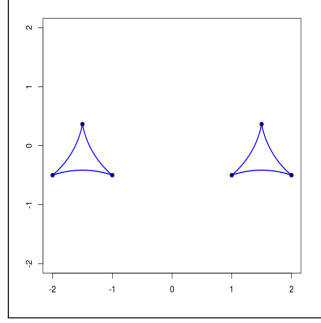

Fig 10: A disconnected α-hull contour. The picture has been produced with the `alphahull` R-package.

Some balls of radius α that do not intersect with $\mathcal{X}$ nevertheless intersect with $\mathcal{A}$ or $\mathcal{B}$. The "worst" case, that is, the balls of radius α that do not intersect with $\mathcal{X}$ while having largest intersection with $\mathcal{A}$ and $\mathcal{B}$ are those centered at $\mathbf{c}_1, \ldots, \mathbf{c}_6$ where $\mathbf{c}_1$, for instance, is maximizing, among all points at distance α from $\mathbf{x}_1$ and $\mathbf{x}_2$, the distance from $\mathbf{x}_3$; similarly, $\mathbf{c}_2$, say, is maximizing, among all points at distance α from $\mathbf{x}_2$ and $\mathbf{x}_3$, the distance from $\mathbf{x}_1$, etc. As a consequence, the α-hull of $\mathcal{X}$, for $\alpha = 3/2$, is the union of the two curvilinear triangles shown in Figure 10—obviously not a connected contour.

References

[1] Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. (1993). *Network Flows*, Prentice-Hall.

[2] Álvarez-Esteban, P.C., del Barrio, E., Cuesta-Albertos, J.A., and Matrán, C. (2018). Wide consensus aggregation in the Wasserstein space. Application to location-scatter families, *Bernoulli* **24**, 3147–3179.

[3] del Barrio, E., Beirlant, J., Buitendag, S., and Hallin, M. (2019). Center-outward quantiles and the measurement of multivariate risk, https://arxiv.org/abs/1912.04924

[4] del Barrio, E., Cuesta-Albertos, J.A., , Hallin, M., and Matrán, C. (2018). Smooth cyclically monotone interpolation and empirical center-outward distribution functions, arXiv:1806.01238v1.

[5] del Barrio, E., Cuesta-Albertos, J.A., Matrán, C., and Mayo-Íscar. (2018). Robust clustering tools based on optimal transportation, *Stat. Comput.*, to appear. https://doi.org/10.1007/s11222-018-9800-z

[6] del Barrio, E., González-Sanz, A., and Hallin, M. (2019). A note on the regularity of center-outward distribution and quantile functions, arXiv:1912.10719.

[7] del Barrio, E. and Loubes, J.M. (2018). Central Limit Theorem for empirical transportation cost in general dimension, *Ann. Probab.* **47**, 926–951.

[8] Basu, D. (1955). On statistics independent of a complete sufficient statistic, *Sankhyā Ser. A* **15**, 377–380.

[9] Basu, D. (1958). On statistics independent of sufficient statistics, *Sankhyā Ser. A* **20**, 223–226.

[10] Basu, D. (1959). The family of ancillary statistics, *Sankhyā Ser. A* **21**, 247–256.

[11] Bell, C.B., Blackwell, D., and Breiman, L. (1960). On the completeness of order statistics, *Ann. Math. Statist.* **31**, 794–797.

[12] Belloni, A. and Winkler, R.L. (2011). On multivariate quantiles under partial orders. *Ann. Statist.* **39**, 1125–1179.

[13] Bertsekas, D. P. (1991). *Linear Network Optimization. Algorithms and Codes.* The MIT Press.

[14] Bickel, P.J. (1965). On some asymptotically nonparametric competitors of Hotelling's T^2, *Ann. Math. Statist.* **36**, 160–173.

[15] Bickel, P.J., Klaassen, Chr., Ritov, Y., and Wellner, J. (1993). *Efficient and Adaptive Inference in Semiparametric Models*, Johns Hopkins University Press, Baltimore.

[16] Biswas, M., Mukhopadhyay, M., and Ghosh, A.K. (2014). A distribution-free two-sample run test applicable to high-dimensional data, *Biometrika* **101**, 913–926.

[17] Boeckel, M., Spokoiny, V., and Suvorikova, A. (2018). Multivariate Brenier cumulative distribution functions and their application to nonparametric testing, arXiv:1809.04090.

[18] Brenier, Y. (1987). Décomposition polaire et réarrangement monotone des champs de vecteurs, *C.R. Acad. Sc. Paris Ser.I* **305 19**, 805–808.

[19] Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions, *Comm. Pure Appl. Math.* **44**, 375–417.

[20] Burkhard, R., Dell'Amico, M., and Martello, S. (2009). *Assignment Problems*, SIAM.

[21] Carlier, G., Chernozhukov, V., and Galichon, A. (2016). Vector quantile regression, *Ann. Statist.* **44**, 1165–1192.

[22] Chakraborty, A. and Chaudhuri, P. (1996). On a transformation-retransformation technique for constructing an affine equivariant multivariate median, *Proc. Amer. Math. Soc.* **124**, 2539–2547.

[23] Chakraborty, A. and Chaudhuri, P. (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths, *Ann. Statist.* **42**, 1203–1231.

[24] Chakraborty, A. and Chaudhuri, P. (2017). Tests for high-dimensional data based on means, spatial signs and spatial ranks, *Ann. Statist.* **45**, 771–799.

[25] Chakraborty, B. and Chaudhuri, P. (1996). On a transformation and re-transformation technique for constructing affine equivariant multivariate median, *Proc. Amer. Math. Soc.* **124**, 2539–2547.

[26] Chakraborty, B. and Chaudhuri, P. (1998). On an adaptive transformation-retransformation estimate of multivariate location, *J. R. Stat. Soc. Ser. B* **60**, 145–157.

[27] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data, *J. Amer. Statist. Assoc.* **91**, 862–872.

[28] Chaudhuri, P. and Sengupta, D. (1993). Sign tests in multidimension : inference based on the geometry of the data cloud, *J. Amer. Statist. Assoc.* **88**, 1363–1370.

[29] Chernoff, H. and Savage, I.R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics, *Ann. Math. Statist.* **29**, 972–994.

[30] Chernozhukov, V., Galichon, A., Hallin, M. and Henry, M. (2017). Monge-Kantorovich depth, quantiles, ranks, and signs, *Ann. Statist.* **45**, 223–256.

[31] Choi, K. and Marden, J. (1997). An approach to multivariate rank tests in multivariate analysis of variance, *J. Amer. Statist. Assoc.* **92**, 1581–1590.

[32] Cuesta-Albertos, J.A. and Matrán, C. (1989). Notes on the Wasserstein metric in Hilbert spaces. *Ann. Probab.* **17**, 1264–1276.

[33] Cuesta-Albertos, J.A., Matrán, C., and Tuero-Diaz, A. (1997). Optimal transportation plans and convergence in distribution, *J. Multivariate Anal.* **60**, 72–83.

[34] De Valk, C. and Segers, J. (2018). Stability and tail limits of transport-based quantile contours, arXiv:1811.12061.

[35] Deb, N. and Sen, B. (2019). Multivariate rank-based distribution-free nonparametric testing using measure transportation, arXiv:1909.08733.

[36] Dick, J. and Pillichshammer, F. (2014). Discrepancy theory and quasi-Monte Carlo integration, in W. Chen, A. Srivastava and G. Travaglini, Eds, *A Panorama of Discrepancy Theory*, 539–620, Springer, New York.

[37] Ekeland, I., Galichon, A. and Henry, M. (2012). Comonotonic measures of multivariate risks, *Math. Finance* **22**, 109–132.

[38] Faugeras, 0. and Rüschendorf, L. (2017). Markov morphisms: a combined

copula and mass transportation approach to multivariate quantiles, *Mathematica Applicanda* **45** 2017, 3–206.

[39] Figalli, A. (2017). *The Monge-Ampère Equation and its Applications*, Zurich Lectures in Advanced Mathematics, EMS, Zurich.

[40] Figalli, A. (2018). On the continuity of center-outward distribution and quantile functions, *Nonlin. Anal.: Theory, Methods & Appl.* **177**, 413-421.

[41] Galichon, A. (2016). *Optimal Transport Methods in Economics*, Princeton University Press, Princeton, N.J.

[42] Genest, Chr. and Rivest, P.L. (2001). On the multivariate probability-integral transformation, *Statist. Probab. Lett.* **53**, 391–399.

[43] Ghosal, P. and Sen, B. (2019). Multivariate ranks and quantiles using optimal transportation and applications to goodness-of-fit testing, arXiv:1905.05340.

[44] Gushchin, A.A. and Borzykh, D.A. (2017). Integrated quantile functions: properties and applications, *Mod. Stoch. Theory Appl.* **4**, 285–314.

[45] Hallin, M. (2017). On distribution and quantile functions, ranks, and signs in $\mathbb{R}^d$, available at https://ideas.repec.org/p/eca/wpaper/2013-258262.html.

[46] Hallin, M., Hlubinka, D., and Hudecová, S. (2020). Efficient center-outward rank tests for multiple-output regression, unpublished manuscript.

[47] Hallin, M., Ingenbleek, J-Fr., and Puri, M.L. (1989). Asymptotically most powerful rank tests for multivariate randomness against serial dependence. *J. Multivariate Anal.* **30**, 34–71.

[48] Hallin, M., La Vecchia, D., and Liu, H. (2019). Center-outward R-estimation for semiparametric VARMA models, arxiv.org/abs/1910.08442.

[49] Hallin, M., Lu, Z., Paindaveine, D., and Šiman, M. (2015). Local bilinear multiple-output quantile regression, *Bernoulli* **21**, 1435–1466.

[50] Hallin, M. and Mehta, Ch. (2015). R-Estimation for asymmetric independent component analysis. *J. Amer. Statist. Assoc.* **110**, 218–232.

[51] Hallin, M., Oja, H., and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape II. Optimal R-estimation of shape, *Ann. Statist.* **34**, 2757–2789.

[52] Hallin, M. and Paindaveine, D. (2002a). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.* **30**, 1103–1133.

[53] Hallin, M. and Paindaveine, D. (2002b). Optimal procedures based on interdirections and pseudo-Mahalanobis ranks for testing multivariate elliptic white noise against ARMA dependence. *Bernoulli* **8**, 787–815.

[54] Hallin, M. and Paindaveine, D. (2002c). Multivariate signed ranks : Randles' interdirections or Tyler's angles? In *Statistical Data Analysis Based on the L_1 Norm and Related Procedures*, Y. Dodge, Ed., Birkhäuser, 271–282.

[55] Hallin, M. and Paindaveine, D. (2004a). Rank-based optimal tests of the adequacy of an elliptic VARMA model. *Ann. Statist.* **32**, 2642–2678.

[56] Hallin, M. and Paindaveine, D. (2004b). Multivariate signed rank tests in vector autoregressive order identification. *Statist. Sci.* **19**, 697–711.

[57] Hallin, M. and Paindaveine, D. (2005). Affine-invariant aligned rank tests

for multivariate general linear models with ARMA errors, *J. Multivariate Anal.* **93**, 122–163.

[58] Hallin, M. and Paindaveine, D. (2006a). Semiparametrically efficient rank-based inference for shape: I Optimal rank-based tests for sphericity, *Ann. Statist.* **34**, 2707–2756.

[59] Hallin, M. and Paindaveine, D. (2006b). Parametric and semiparametric inference for shape: the role of the scale functional, *Statist. Decisions* **24**, 1001–1023.

[60] Hallin, M. and Paindaveine, D. (2008). Optimal rank-based tests for homogeneity of scatter, *Ann. Statist.* **36**, 1261-1298.

[61] Hallin, M., Paindaveine, D., and Šiman, M. (2010). Multivariate quantiles and multiple-output regression quantiles: from L^1 optimization to halfspace depth" (with discussion), *Ann. Statist.* **38**, 635–669.

[62] Hallin, M., Paindaveine, D., and Verdebout, Th. (2010). Optimal rank-based testing for principal components, *Ann. Statist.* **38**, 3245–3299.

[63] Hallin, M., Paindaveine, D., and Verdebout, Th. (2013). Optimal rank-based tests for common principal components, *Bernoulli* **19**, 2524–2556.

[64] Hallin, M., Paindaveine, D., and Verdebout, Th. (2014). Efficient R-estimation of principal and common principal components, *J. Amer. Statist. Assoc.* **109**, 1071–1083.

[65] Hallin, M. and Werker, B.J.M. (1998). Optimal testing for semiparametric autoregressive models: from Gaussian Lagrange multipliers to regression rank scores and adaptive tests. In S. Ghosh, Ed., *Asymptotics, Nonparametrics, and Time Series*. M. Dekker, New York, 295–358.

[66] Hallin, M. and Werker, B.J.M. (2003). Semiparametric efficiency, distribution-freeness, and invariance, *Bernoulli* **9**, 137–165.

[67] Hamel, A. and Kostner, D. (2018). Cone distribution functions and quantiles for multivariate random variables, *J. Multivariate Anal.* **167**, 97–113.

[68] He, X. and Wang, G. (1997). Convergence of depth contours for multivariate datasets, *Ann. Statist.* **25**, 495–504.

[69] Hodges, J.L. (1955). A bivariate sign test, *Ann. Math. Statist.* **26**, 523-527.

[70] Ilmonen, P. and Paindaveine, D. (2011). Semiparametrically efficient inference based on signed ranks in symmetric independent component models, *Ann. Statist.* **39**, 2448–2476.

[71] Judd, K.L. (1998). *Numerical Methods in Economics*, MIT Press, Cambridge, MA.

[72] Karp, R. M. (1978). A characterization of the minimum cycle mean in a digraph. *Discrete Math.* **23**, 309–311.

[73] Koehn, U. and Thomas, D.L. (1975). On statistics independent of a sufficient statistic: Basu's Lemma. *The Amer. Statist.* **29**, 40–42.

[74] Koltchinskii, V. (1997). M-Estimation, convexity and quantiles, *Ann. Statist.* **25**, 435–477.

[75] Lehmann, E.L. and Scholz, F.-W. (1992). Ancillarity. In M. Ghosh and P.K. Pathak, Eds *Current Issues in Statistical Inference: Essays in Honor of D. Basu*, Institute of Mathematical Statistics, Hayward, CA, 32–51.

[76] Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*,

Springer, New York.

[77] Liu, R. Y. (1992). Data depth and multivariate rank tests, in Y. Dodge, Ed., *L^1 Statistics and Related Methods*, 279–294, North-Holland, Amsterdam.

[78] Liu, R.Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests, *J. Amer. Statist. Assoc.* **88**, 257–260.

[79] Lòpez-Pintado, S., and Romo, J. (2012). On the concept of depth for functional data, *J. Amer. Statist. Assoc.* **104**, 718–734.

[80] Marden, J. (1999). Multivariate rank tests., in S. Ghosh, Ed., *Design of Experiments and Survey Sampling*, M. Dekker, New York, 401–432.

[81] McCann, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps, *Duke Math. J.* **80**, 309–324.

[82] McKeague, I.W., López-Pintado, S., Hallin, M., and Šiman, M. (2011). Analyzing growth trajectories. *J. Dev. Or. Health and Disease* **2**, 322-329.

[83] Monge, G. (1781). *Mémoire sur la Théorie des Déblais et des Remblais*, *Mémoires Acad. Royale Sci.* **1781**, 666–704.

[84] Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *J. Nonparametr. Statist.* **5**, 201-213.

[85] Möttönen, J., Oja, H., and Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests, *Ann. Statist.* **25**, 542–552.

[86] Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods.* CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, PA.

[87] Nordhausen, K., Oja, H., and Paindaveine, D. (2009). Signed-rank tests for location in the symmetric independent component model. *J. Multivariate Anal.* **100**, 821–834.

[88] Nordhausen, K., Oja, H., and Tyler, D.E. (2006). On the efficiency of invariant multivariate sign and rank tests, in Liski, E.P., Isotalo, J., Niemelä, J., Puntanen, S., and Styan, G.P.H., Eds, *Festschrift for Tarmo Pukkila on his 60th Birthday*, 217-231, University of Tampere, Tampere.

[89] Oja, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates: a review. *Scand. J. Statist.* **26**, 319–343.

[90] Oja, H. (2010). *Multivariate Nonparametric Methods with R: an approach based on spatial signs and ranks.* Springer, New York.

[91] Oja, H., and Paindaveine, D. (2005). Optimal signed-rank tests based on hyperplanes, *J. Statist. Plann. Inference* **135**, 300–323.

[92] Oja, H. and Randles, R.H. (2004). Multivariate nonparametric tests. *Statist. Sci.* **19**, 598–605.

[93] Panaretos, V. and Zemel, Y. (2016). Amplitude and phase variation of point processes. *Ann. Statist.* **44**, 771–812.

[94] Panaretos, V. and Zemel, Y. (2018). Fréchet means and Procrustes analysis in Wasserstein space. *Bernoulli*, to appear.

[95] Panaretos, V. and Zemel, Y. (2019). Statistical aspects of Wasserstein distances. *Annual Review of Statistics and its Application* **6**, 405-431.

[96] Pateiro-López, B., and Rodríguez-Casal, A. (2010). Generalizing the convex hull of a sample: the R package α-hull. *J. Statist. Software* **34**.

[97] Peyré, G. and Cuturi, M. (2019). Computational optimal transport. *Foun-

dations and Trends in Machine Learning **11**, 355–607.

[98] Pfanzagl, J. (2011). *Parametric Statistical Theory*, Walter de Gruyter, Berlin.

[99] Puri, M.L. and Sen, P.K. (1966). On a class of multivariate multisample rank-order tests (1966). *Sankhya A* **28**, 353–376.

[100] Puri, M.L. and Sen, P.K. (1967). On the theory of rank order tests for location in the multivariate one-sample problem. *Ann. Math. Statist.* **38**, 1216–1228.

[101] Puri, M.L. and Sen, P.K. (1969). A class of rank order tests for a general linear hypothesis. *Ann. Math. Statist.* **40**, 1325–1343.

[102] Puri, M.L. and Sen, P.K. (1971). *Nonparametric Methods in Multivariate Analysis*. John Wiley & Sons, New York.

[103] Rachev, S.T. and Rüschendorf, L. (1998a, b). *Mass Transportation Problems* I and II, Springer, New York.

[104] Randles, R.H. (1989). A distribution-free multivariate sign test based on interdirections, *J. Amer. Statist. Assoc.* **84**, 1045–1050.

[105] Randles, R.H. (2000). A simpler, affine equivariant multivariate, distribution-free sign test. *J. Amer. Statist. Assoc.* **95**, 1263–1268.

[106] Rockafellar, R.T. (1966). Characterization of the subdifferential of convex functions, *Pacific J. Math.* **17**, 497–510.

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

[108] Rockafellar, R.T. and Wets, R.J.B. (1998). *Variational Analysis*, Springer.

[109] Rüschendorf, L. and Rachev, S.T. (1990). A characterization of random variables with minimum L²-distance, *J. Multivariate Anal.* **32**, 48–54.

[110] Santner, T.J., Williams, B.J. and Notz, W.I. (2003). *The Design and Analysis of Computer Experiments*, Springer-Verlag, New York.

[111] Segers, J., van den Akker, R., and Werker, B.J.M. (2014). Semiparametric Gaussian copula models: geometry and efficient R-estimation, *Ann. Statist.* **42**, 1911–1940.

[112] Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications, *Stat. Neerl.* **56**, 214–232.

[113] Serfling, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardization, *J. Nonparametr. Stat.* **22**, 915–936.

[114] Shi, H., Drton, M., and Han, F. (2019). Distribution-free consistent independence tests via Hallin's multivariate ranks, arXiv:1909.10024.

[115] Um, Y. and Randles, R.H. (1998). Nonparametric tests for the multivariate multi-sample location problem. *Statist. Sinica* **8**, 801–812.

[116] Villani, C. (2003). *Topics in Optimal Transportation*, American Mathematical Society, Providence, RI.

[117] Villani, C. (2009). *Optimal Transport: Old and New*, Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, Heidelberg.

[118] Yosida, K. (1964). *Functional Analysis*, Springer Verlag.

[119] Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function, *Ann. Statist.* **28**, 461–482.

[120] Zuo, Y. (2018). On general notions of depth for regression,

arXiv:1805.02046v1.

[121] Zuo, Y. and He, X. (2006). On the limiting distributions of multivariate depth-based rank sum statistics and related tests, *Ann. Statist.* **34**, 2879–2896.

For convenience and ease of use, this list regroups the references for both the Appendix and the main text.