The Annals of Statistics
2021, Vol. 49, No. 2, 1139–1165
https://doi.org/10.1214/20-AOS1996

DISTRIBUTION AND QUANTILE FUNCTIONS, RANKS AND SIGNS IN DIMENSION d: A MEASURE TRANSPORTATION APPROACH

BY MARC HALLIN[1], EUSTASIO DEL BARRIO[2,*], JUAN CUESTA-ALBERTOS[3] AND CARLOS MATRÁN[2,†]

[1]*ECARES and Département de Mathématique, Université libre de Bruxelles, Belgium, mhallin@ulb.ac.be*
[2]*IMUVA and Departamento de Estadística e Investigación Operativa, Universidad de Valladolid, Spain, *tasio@eio.uva.es; †carlos.matran@uva.es*
[3]*Departamento de Matemáticas, Estadística y Computación, Facultad de Ciencias, Universidad de Cantabria, Spain, juan.cuesta@unican.es*

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 and signs introduced in Chernozhukov, Galichon, Hallin and Henry (*Ann. Statist.* **45** (2017) 223–256) 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 (*Duke Math. J.* **80** (1995) 309–323) and our results 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 (*Sankhyā* **21** (1959) 247–256) 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 of the concept, is provided. A numerical study of the resulting empirical quantile contours is conducted.

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

Received May 2019; revised June 2020.
MSC2020 subject classifications. Primary 62G30; secondary 62B05.
Key words and phrases. Multivariate distribution function, multivariate quantiles, multivariate ranks, multivariate signs, Glivenko–Cantelli theorem, Basu theorem, distribution-freeness, ancillarity, cyclical monotonicity.

(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 in dimension one 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 $\mathrm{P}^{(n)}_{\boldsymbol{\theta},f}$ 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, $\mathbf{X} \sim \mathrm{P}^{(n)}_{\boldsymbol{\theta},f}$ 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 null hypothesis $H_0^{(n)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ (i.e., $\mathrm{P}^{(n)}_{\boldsymbol{\theta},f} \in \mathcal{P}^{(n)}_{\boldsymbol{\theta}_0} := \{\mathrm{P}^{(n)}_{\boldsymbol{\theta}_0,f} | f \in \mathcal{F}_1\}$) 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 $\mathcal{P}^{(n)}_{\boldsymbol{\theta}}$ in the sense of Basu (1959) (see, e.g., Example 7 in Lehmann and Scholz (1992)),*

[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, for example, Hallin and Werker (1999)), etc.

[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, for instance, Example 7 in Lehmann and Scholz (1992).

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

In families satisfying the condition (Koehn and Thomas (1975)) of nonexistence 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 $\mathbf{Z}^{(n)}_{(\cdot)}(\boldsymbol{\theta})$."[4]

Intuitively, thus, and leaving aside the required mathematical precautions, the order statistic $\mathbf{Z}^{(n)}_{(\cdot)}(\boldsymbol{\theta})$, being minimal sufficient for $\mathcal{P}^{(n)}_{\boldsymbol{\theta}}$, is carrying all the information about the nuisance f and nothing but that information, while the ranks, being (essentially) "maximal independent of $\mathbf{Z}^{(n)}_{(\cdot)}(\boldsymbol{\theta})$," 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* $\mathrm{P}^{(n)}_{\boldsymbol{\theta},f}$*, via rank-based procedures (here, 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 property of ranks under a group of order-preserving transformations of $\mathbb{R}^n$ *generating* the fixed-$\boldsymbol{\theta}$ submodel (i.e., yielding a unique orbit in the family $\mathcal{P}^{(n)}_{\boldsymbol{\theta}}$ 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, 1969), Sen and Puri (1967). 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 (2014), etc. to quote only a very few. Componentwise ranks actually are intimately related to copula transforms, of which they constitute the empirical version:

[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.

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, Oja and Paindaveine (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, Oja and Tienari (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 Chakraborty 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$^+$).

(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, Paindaveine and Šiman (2010) for the related concept of quantile, López-Pintado and Romo (2009) 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, 2002c)), 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, 2005, 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 nonelliptical 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

[6] Biswas, Mukhopadhyay and Ghosh (2014) is an exception, but fails on (DF$^+$).

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, Spokoiny and Suvorikova (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 comprises a main text and an Online Appendix (Hallin et al. (2020)) consisting of eight sections, labeled A, B, ..., H. 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 these 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 the 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 transportation with the usual distribution function F 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 Appendix D 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 $n = 20{,}000$ in any dimension (the complexity of the algorithms in $\mathbb{R}^d$ only depends 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. Our *center-outward distribution and quantile functions* actually are particular cases of what Chernozhukov et al. (2017) define as *Monge-Kantorovich vector ranks and quantiles*; as our approach is focusing on the properties of distribution functions and no longer relies on Monge–Kantorovich optimization ideas, however, we consistently adopt the terminology *center-outward* 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 $\mathcal{F}_d := \{f := \mathrm{dP}/\mathrm{d}\mu_d, \mathrm{P} \in \mathcal{P}_d\}$ the corresponding family of densities, by $\mathcal{B}_d^n$ the n-fold product $\mathcal{B}_d \times \cdots \times \mathcal{B}_d$, 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, $\mathrm{spt}(\mathrm{P})$ for its interior. Finally, let $\mathcal{S}_{d-1}$, $\mathbb{S}_d$ and $\overline{\mathbb{S}}_d$ denote 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 (1995). 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.[8] That gradient, moreover, is a homeomorphism between $\mathbb{R}^d \setminus \nabla\phi^{-1}(\{\mathbf{0}\})$ and the punctured unit ball $\mathbb{S}_d \setminus \{\mathbf{0}\}$ (Figalli (2018)) and its inverse naturally qualifies as a quantile function—a very simple and intuitively clear characterization.

Things are slightly 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 less direct and slightly more elaborate definitions developed here, starting, as in Chernozhukov et al. (2017) and Ghosal and Sen (2019), with the quantile function $\mathbf{Q}_\pm$ (Definition 2.1), then defining the distribution function $\mathbf{F}_\pm$ via the Legendre transform (2.3). The two approaches, however, coincide in case P has a nonvanishing 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 its domain $\mathrm{dom}(\psi) := \{\mathbf{x} : \psi(\mathbf{x}) < \infty\}$ and (b) is 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.

THEOREM 2.1 (McCann (1995)). *Let* P_1 *and* P_2 *denote two distributions in* $\mathcal{P}_d$. *Then,*

(i) *the class of functions*

$$\nabla\Psi_{\mathrm{P}_1;\mathrm{P}_2} := \big\{\nabla\psi \,|\, \psi : \mathbb{R}^d \to \mathbb{R} \text{ convex, lower semicontinuous, and such that } \nabla\psi\#\mathrm{P}_1 = \mathrm{P}_2\big\} \tag{2.1}$$

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.

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

[8] 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 $X \sim \mathrm{P}_1$—we say that T is *pushing* P_1 *forward* 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 $\mathbf{x} \in \mathrm{spt}(\mathrm{P}_1)$.

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 $\mathbb{S}_d$ (i.e., on a set with U_d-probability zero), let us further impose

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

This extension—a standard construction in convex analysis (see, e.g., (A.18) in Figalli (2017))—yields a function ψ which is convex and lower semicontinuous over $\mathbb{R}^d$, with 1-Lipschitz Legendre transform (2.3). The domain of ψ, viz. $\mathrm{dom}(\psi) := \{\mathbf{u} | \psi(\mathbf{u}) < \infty\}$, is such that $\mathbb{S}_d \subset \mathrm{dom}(\psi) \subset \bar{\mathbb{S}}_d$. 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 Appendix B.3, we propose the following definition 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 forward to P, hence fully characterizes P. Such a.e. uniqueness, in probability and statistics, is not uncommon: densities, conditional expectations, likelihoods, MLEs, etc. 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

$$\phi(\mathbf{x}) := \psi^*(\mathbf{x}) := \sup_{\mathbf{u} \in \mathbb{S}_d} \big(\langle \mathbf{u}, \mathbf{x} \rangle - \psi(\mathbf{u})\big), \quad \mathbf{x} \in \mathbb{R}^d \tag{2.3}$$

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 family of 1-Lipschitz functions, 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]

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

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

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

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 $\mathrm{spt}(\mathrm{P})$ and $\mathbb{S}_d$, respectively, such that (a) $\mathrm{P}(\breve{\mathrm{spt}}(\mathrm{P})) = 1 = \mathrm{U}_d(\breve{\mathbb{S}}_d)$, (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 bijective, and (c) those restrictions are the inverse of each other. Accordingly, $\mathbf{F}_\pm$ qualifies as a center-outward distribution function.

[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, $\nabla\phi(\mathbf{x})$.

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* $[0, 1]$, $\mathbf{S}(\mathbf{Z}) := \mathbf{F}_\pm(\mathbf{Z})/\|\mathbf{F}_\pm(\mathbf{Z})\|$ *is 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 the *center-outward quantile contour* of order q as

$$\begin{aligned} \mathbb{C}(q) &:= \mathbf{Q}_\pm(q\bar{\mathbb{S}}_d) = \{\mathbf{z} | \|\mathbf{F}_\pm(\mathbf{z})\| \le q\} \quad \text{and} \\ \mathcal{C}(q) &:= \mathbf{Q}_\pm(q\mathcal{S}_{d-1}) = \{\mathbf{z} | \|\mathbf{F}_\pm(\mathbf{z})\| = q\}, \end{aligned} \tag{2.6}$$

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* $\inf\{x | F(x) \ge (1 + u)/2)\}$, $u \in (-1, 1)$, *and* $\mathbb{C}(q)$, *for* $q \in (0, 1)$, *with the interval* (F *the traditional distribution function*)[13]

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

The modulus $\|\mathbf{F}_\pm(\mathbf{x})\|$ thus is the order of the quantile contour and the P-probability content of the smallest 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 a quantile concept to be meaningful. The terminology *quantile region* and *quantile contour*, indeed, calls for 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

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

[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).

[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).

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 also $\mathbf{F}_\pm$, 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 $\mathbf{0}$, we have $\mathbf{Q}_\pm(V(\mathbf{u})) \cap \bigcap_{0<q<1} \mathbb{C}(q) = \varnothing$ and, consequently, $\partial\psi(\mathbf{0}) = \bigcap_{0<q<1} \mathbb{C}(q)$. Being 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 $\mathbf{Q}_\pm$ enjoying continuity over the punctured ball $\mathbb{S}_d \setminus \{\mathbf{0}\}$ but not over $\mathbb{S}_d$.

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

$$0 < \lambda_{D;f} \leq f(\mathbf{x}) \leq \Lambda_{D;f} < \infty \quad \text{for all } \mathbf{x} \in (D\mathbb{S}_d) \cap \overline{\text{spt}}(\mathrm{P}_f).$$

That class includes the class $\mathcal{P}_d^+$ of distributions with support $\text{spt}(\mathrm{P}) = \mathbb{R}^d$ considered by Hallin (2017) and Figalli (2018).

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

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

(i) $\mathbf{F}_\pm$ *and the center-outward quantile function* $\mathbf{Q}_\pm = \nabla\psi$ *are homeomorphisms between* $\mathbb{S}_d \setminus \{\mathbf{0}\}$ *and* $\text{spt}(\mathrm{P}) \setminus K$, *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* $\text{spt}(\mathrm{P})$;

[15] That convex support is not necessarily bounded.

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

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

(iiib) $f(\mathbf{z}) = c_d^{-1} \det[\mathbf{H}_\psi(\nabla\phi(\mathbf{z}))] \|\nabla\phi(\mathbf{z})\|^{1-d} I[\mathbf{z} \in \text{spt}(\text{P}_f) \setminus \mathbb{C}(0)]$ *where* c_d *denotes the area* $2\pi^{d/2}/\Gamma(d/2)$ *of the unit sphere* $\mathcal{S}_{d-1}$ *and* $\mathbf{H}_\psi(\mathbf{u})$ *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 $\text{P} = \nabla\Upsilon\#\text{U}_d$ where Υ is convex and $\nabla\Upsilon$ a homeomorphism from $\mathbb{S}_d \setminus \{\mathbf{0}\}$ to $\nabla\Upsilon(\mathbb{S}_d \setminus \{\mathbf{0}\})$ such that $\nabla\Upsilon(\{\mathbf{0}\})$ is a compact convex set of Lebesgue measure zero. By construction, such $\text{P} \in \mathcal{P}_d^\pm$ has center-outward quantile function $\mathbf{Q}_\pm = \nabla\Upsilon$ and center-outward distribution function $\mathbf{F}_\pm = (\nabla\Upsilon)^{-1}$ for $\mathbf{x}$ in the range of $\nabla\Upsilon$; $\|\mathbf{F}_\pm(\mathbf{x})\| = 1$ outside that range, and $\mathbf{F}_\pm$ satisfies Proposition 2.3. The latter statement actually can be rephrased as $\mathcal{P}_d^{\text{conv}} \subset \mathcal{P}_d^\pm$, with the following immediate corollary in terms of quantile regions and contours.

COROLLARY 2.1. *For any* $\text{P} \in \mathcal{P}_d^\pm$ (*hence, any* $\text{P} \in \mathcal{P}_d^{\text{conv}}$) *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 $\text{P} \in \mathcal{P}_d^\pm$, $\mathbf{F}_\pm$ 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, though: cfr the possible discontinuities at the origin) for $\text{P} \in \mathcal{P}_d^\pm$.

2.2. *Center-outward ranks and signs in* $\mathbb{R}^d$. Turning to the sample situation, denote by $\mathbf{Z}^{(n)} := (\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)})$ 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

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

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" over 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 $\frac{j}{n_R+1}$, $j = 1, \ldots, n_R$,

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

along with n_0 copies of the origin whenever $n_0 > 0$. In theory, by a "regular" n_S-tuple $\mathfrak{S}^{(n_S)}$, 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, Williams and Notz (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 $\mathbf{F}_{\pm}^{(n)} : (\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}) \mapsto (\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_1^{(n)}), \ldots, \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_n^{(n)})) =: \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}^{(n)})$ satisfying

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

or, equivalently,

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

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* $\mathbf{Z}_{(\cdot)}^{(n)}$ 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 $\mathbf{Z}_{(\cdot)}^{(n)} := (\mathbf{Z}_{(1)}^{(n)}, \ldots, \mathbf{Z}_{(n)}^{(n)})$, 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 $\mathbf{Z}_{(\cdot)}^{(n)}$-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 $\mathbb{R}^{nd} \setminus N$: each of the minimizers $\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}^{(n)})$ of (2.10) indeed is such that the n-tuple

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

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 $\mathbf{z} \mapsto (\mathbf{z}_{(\cdot)}, \mathbf{F}_{\pm}^{(n)}(\mathbf{z}))$ then is injective over $\mathbb{R}^{nd} \setminus N$. 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 $\mathbb{R}^{nd} \setminus N$ of $\mathbf{z} \mapsto (\mathbf{z}_{(\cdot)}, \mathbf{F}_{\pm}^{(n)}(\mathbf{z}))$.

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 $\{(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_k, \mathbf{y}_k)\} \subseteq S$,

$$\langle \mathbf{y}_1, \mathbf{x}_2 - \mathbf{x}_1 \rangle + \langle \mathbf{y}_2, \mathbf{x}_3 - \mathbf{x}_2 \rangle + \cdots + \langle \mathbf{y}_k, \mathbf{x}_1 - \mathbf{x}_k \rangle \leq 0. \tag{2.12}$$

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 semidefinite 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, equivalently, minimizes $\sum_{i=1}^n \|\mathbf{y}_i - \mathbf{x}_i\|^2$ (a squared 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 costs $\|\mathbf{y}_i - \mathbf{x}_i\|^2$. 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*

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

  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* $\mathbf{S}_{\pm,i}^{(n)} := \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)}) I[\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)}) \neq \mathbf{0}]/\|\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)})\|$, 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*

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

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

Proposition 2.4 considerably reinforces, under more general assumptions (no compact supports, no second-order moments), an early strong consistency result by Cuesta-Albertos, Matrán and Tuero-Díaz (1997) as well as the Glivenko–Cantelli result of Boeckel, Spokoiny and Suvorikova (2018). It readily follows from the more general Proposition 3.3 below, which establishes (2.13) under sup form for 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) *$\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}^{(n)}) := (\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_1^{(n)}), \ldots, \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_n^{(n)}))$ 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 $(R_{\pm,1}^{(n)}, \ldots, R_{\pm,n}^{(n)})$ and $(\mathbf{S}_{\pm,1}^{(n)}, \ldots, \mathbf{S}_{\pm,n}^{(n)})$ of center-outward ranks and signs are mutually independent; for $n_0 > 0$, the same independence holds for the $(n_R n_S)$-tuples of ranks and signs associated with the (random) set $\{i_1 < \cdots < i_{n_R n_S}\}$ of indices for which $\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_{i_j}^{(n)}) \neq \mathbf{0}$*;

(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*;

(v) *for $n_0 \le 1$ or after adequate tie-breaking (see the 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 $\mathbf{z} \mapsto (\mathbf{z}_{(\cdot)}, \mathbf{F}_{\pm}^{(n)}(\mathbf{z}))$ is injective over $\mathbb{R}^{nd} \setminus N$ (N the Lebesgue measure-zero random set on which (2.10) admits two solutions or more; see page 13). As previously explained, injectivity is easily restored via a simple tie-breaking device: (v) then holds irrespective of the value 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 $\mathcal{B}_{\pm}^{(n)} \cap (\mathbb{R}^{nd} \setminus M) = \tilde{\mathcal{B}}_{\pm}^{(n)} \cap (\mathbb{R}^{nd} \setminus M)$.*

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 1 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, $\sup_{\mathbf{x}\in\mathbb{R}^d} \|\overline{\mathbf{F}}_{\pm}^{(n)}(\mathbf{x}) - \mathbf{F}_{\pm}(\mathbf{x})\| \to 0$

[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.

a.s. as $n \to \infty$—rather than the $\max_{1\le i\le n}$ form given 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 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 $T : \mathbf{x}_i \mapsto T(\mathbf{x}_i) = \mathbf{y}_i$ for $i = 1, \ldots, n$ 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 $\overline{T} : \mathbb{R}^d \to \mathbb{R}^d$ such that $\overline{T}(\mathbf{x}_i) = T(\mathbf{x}_i) = \mathbf{y}_i$ for $i = 1, \ldots, n$.

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 $S = \{(\mathbf{x}_i, \mathbf{y}_i)|i = 1, \ldots, n\}$ of $\mathbb{R}^d \times \mathbb{R}^d$ enjoying cyclical monotonicity is the subdifferential (at $\mathbf{x}_1, \ldots, \mathbf{x}_n$) of some (continuously) differentiable convex function ψ.

Note that Assumption (A) holds if and only if identity is the unique minimizer, among the set of all permutations σ of $\{1, \ldots, n\}$, of $\sum_{i=1}^n \|\mathbf{x}_i - \mathbf{y}_{\sigma(i)}\|^2$. 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

$$\min_{\pi} \sum_{i=1}^n \sum_{j=1}^n c_{i,j}\pi_{i,j} \quad \text{s.t.} \quad \sum_{i=1}^n \pi_{i,j} = \sum_{j=1}^n \pi_{i,j} = \frac{1}{n}, \quad \pi_{i,j} \ge 0, i, j = 1, \ldots, n. \tag{3.1}$$

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 $j \ne i$ 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 $\mathbb{R}^d$ the complement of which 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 preserving the interpolation feature. For Step 1, we rely on the following result (see Appendix F.1 for the proof).

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

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

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

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

$$\langle \mathbf{x}_i, \mathbf{y}_i \rangle - \psi_i > \max_{j=1,\ldots,n, j\ne i} \big(\langle \mathbf{x}_i, \mathbf{y}_j \rangle - \psi_j\big), \quad i = 1, \ldots, n. \tag{3.2}$$

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 $T(\mathbf{x}_i) = \mathbf{y}_i$ for $i = 1, \ldots, n$ is cyclically monotone if and only if there exist real numbers $\psi_{n_0+1}, \ldots, \psi_n$ and ψ_1 such that, setting $\psi_i = \psi_1$ for $i = 2, \ldots, n_0$,

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

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 $\psi_{n_0+1}, \ldots, \psi_n$ and ψ_1 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

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

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, for example, 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 (1965))—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 semidefinite 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 Rockafellar and Wets (1998)). 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 it satisfied, then backtransform the interpolating function). Set

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

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 that proof to conclude that, for $\mathbf{x} \in B(\mathbf{x}_i, \delta)$, it holds that $\varphi_\varepsilon(\mathbf{x}) = \langle \mathbf{x}, \mathbf{y}_i \rangle - \psi_i - \frac{\varepsilon}{2} \|\mathbf{y}_i\|^2$, which entails $T_\varepsilon(\mathbf{x}_i) = \mathbf{y}_i$ 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

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

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

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

$$\text{s.t.} \sum_{j=1,\ldots,n; j \neq i} (z_{i,j} - z_{j,i}) = 0, \tag{3.8}$$

$$\sum_{i,j=1,\ldots,n; i \neq j} z_{i,j} = 1, \quad z_{i,j} \geq 0, i, j = 1, \ldots, n.$$

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, Magnanti and Orlin (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 W in the graph, $\delta_{ij}(W) = 1$ if the arc connecting i and j belongs to W ($\delta_{ij}(W) = 0$ otherwise), and $f(W) \geq 0$ is the flow along 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

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

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 length of the path $(i_1, i_2, \ldots, i_k)$ 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 initial values $d_{0,1} = 0$, $d_{0,i} = \infty$ for $i \neq 1$ and letting $d_{k+1,i} = \min_j (d_{k,j} + c_{j,i})$ 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; letting $\tilde{d}_i := \min_{0 \leq k \leq n-1} \tilde{d}_{k,i}$, set $\psi_i := -\tilde{d}_i$. 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 $O(n^3)$ computing time.

For $n = 2$, it is easily seen that the optimum in the dual (3.7) (hence in the primal problem (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 an i.i.d. sample of observations with distribution $\mathrm{P} \in \mathcal{P}_d^{\pm}$ and center-outward distribution function $\mathbf{F}_{\pm}$. Then*

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

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 $(\mathbf{Z}_i^{(n)}, \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)}))$, $i = 1, \ldots, n$. 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 $\overline{\mathbf{Q}}_{\pm}^{(n)} := (\overline{\mathbf{F}}_{\pm}^{(n)})^{-1}$—of the empirical quantile function $\mathbf{Q}_{\pm}^{(n)}$. For $n_0 > 1$, the restriction of $\overline{\mathbf{F}}_{\pm}^{(n)}$ to $\mathbb{R}^d \setminus (\mathbf{F}_{\pm}^{(n)})^{-1}(\mathbf{0})$ (a restriction that has Lebesgue measure one) can be considered instead. In all cases, strong consistency still holds for $\overline{\mathbf{Q}}_{\pm}^{(n)}$; its uniformity is lost, however, unless spt(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 $n = 20{,}000$ (with $n_R = 100$ and $n_S = 200$, for instance) on a computer with 32 Gb RAM. The algorithm consists of three main steps.

(Step 1) Determine the optimal pairing 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 Burkard, Dell'Amico and Martello (2009)). Faster algorithms are available, though, as Bertsekas' *auction algorithm* or its variant, the *forward/reverse auction algorithm* (Chapter 4 in Bertsekas (1991)), 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

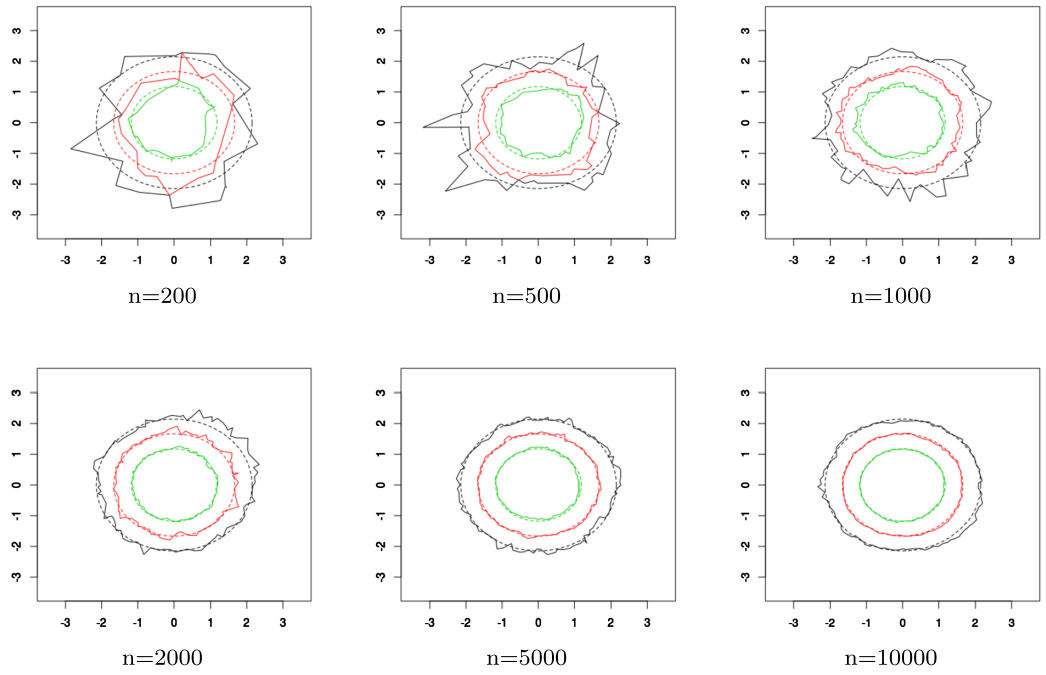

FIG. 1. *Smoothed empirical center-outward quantile contours* (*probability contents* 0.50 (*green*), 0.75 (*red*), *and* 0.90 (*black*)) *computed from* $n = 200, 500, 1000, 2000, 5000,$ *and* 10,000 *i.i.d. observations from a bivariate* $\mathcal{N}(\mathbf{0}, \mathbf{I})$ *distribution*, *along with their theoretical counterparts* (*dotted lines*).

of the assignment problem was not optimal: return to Step 1 with a smaller value of ϵ. If not, 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 $\mathcal{N}(\mathbf{0}, \mathbf{I})$ distributions, and increasing samples sizes $n = 200, \ldots, 10{,}000$. Figure 1 clearly shows the expected consistency. Empirical contours are nicely nested, as they are supposed to be. For sample sizes as big as $n = 1000$ (with $n_R = 25$ and $n_S = 40$), and despite the fact that the underlying distribution is light-tailed, the 0.90 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 0.90 based on $n_R = 25$ observations; they rapidly and uniformly disappear from $n = 2000$ on.

Figures 2–4 consider various Gaussian mixtures. Gaussian mixtures generate a variety of possibly multimodal and nonconvex empirical dataclouds. In Figure 2, we simulated $n = 2000$ observations from a symmetric mixture of two spherical Gaussians. It clearly demonstrates the quantile contour nature of our interpolations, as opposed to level contours. Level contours clearly would produce (mostly, in the right-hand panel) 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 the empirical quantile contours

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

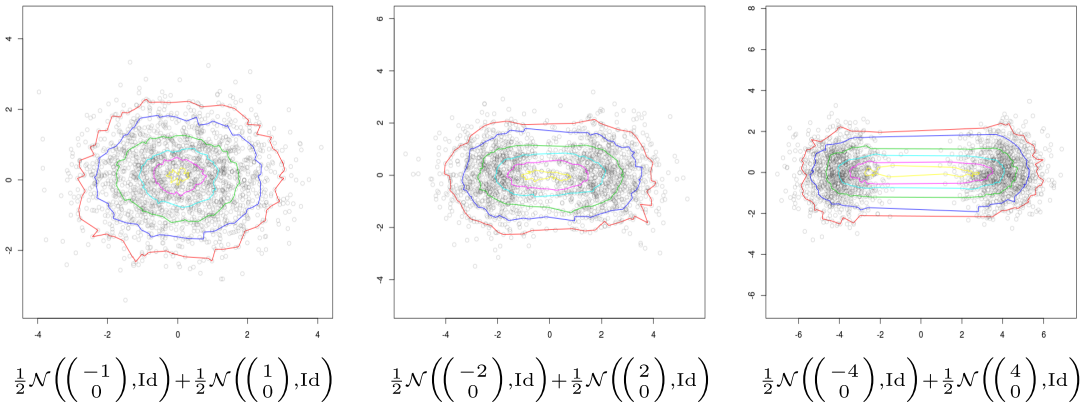

FIG. 2. *Smoothed empirical center-outward quantile contours* (*probability contents* 0.02 (*yellow*), 0.20 (*cyan*), 0.25 (*light blue*) 0.50 (*green*), 0.75 (*dark blue*), 0.90 (*red*)) *computed from* $n = 2000$ *i.i.d. observations from mixtures of two bivariate Gaussian distributions*.

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 0.90 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, these sign curves are quite uninformative and we did not plot them in Figures 1 and 2. In the highly nonspherical Gaussian mixture of Figure 3, sign curves are conveying an essential information.

Figure 4 is providing the full picture for $n = 20{,}000$ (see also Figure 3 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

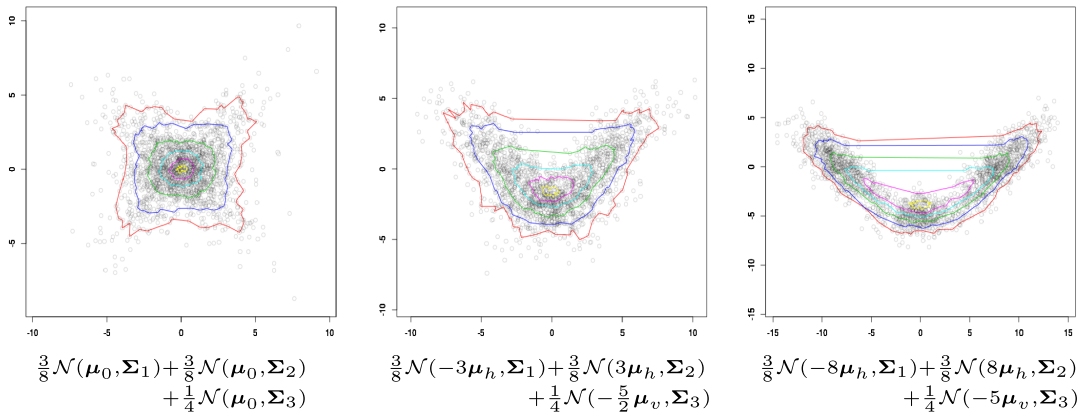

FIG. 3. *Smoothed empirical center-outward quantile contours* (*probability contents* 0.02 (*yellow*), 0.20 (*cyan*), 0.25 (*light blue*) 0.50 (*green*), 0.75 (*dark blue*), 0.90 (*red*)) *computed from* $n = 2000$ *i.i.d. observations from mixtures of three bivariate Gaussian distributions, with* $\boldsymbol{\mu}_0 = \binom{0}{0}$, $\boldsymbol{\mu}_h = \binom{1}{0}$, $\boldsymbol{\mu}_v = \binom{0}{1}$, $\boldsymbol{\Sigma}_1 = \left(\begin{smallmatrix}5 & -4\\ -4 & 5\end{smallmatrix}\right)$, $\boldsymbol{\Sigma}_2 = \left(\begin{smallmatrix}5 & 4\\ 4 & 5\end{smallmatrix}\right)$, *and* $\boldsymbol{\Sigma}_3 = \left(\begin{smallmatrix}4 & 0\\ 0 & 1\end{smallmatrix}\right)$.

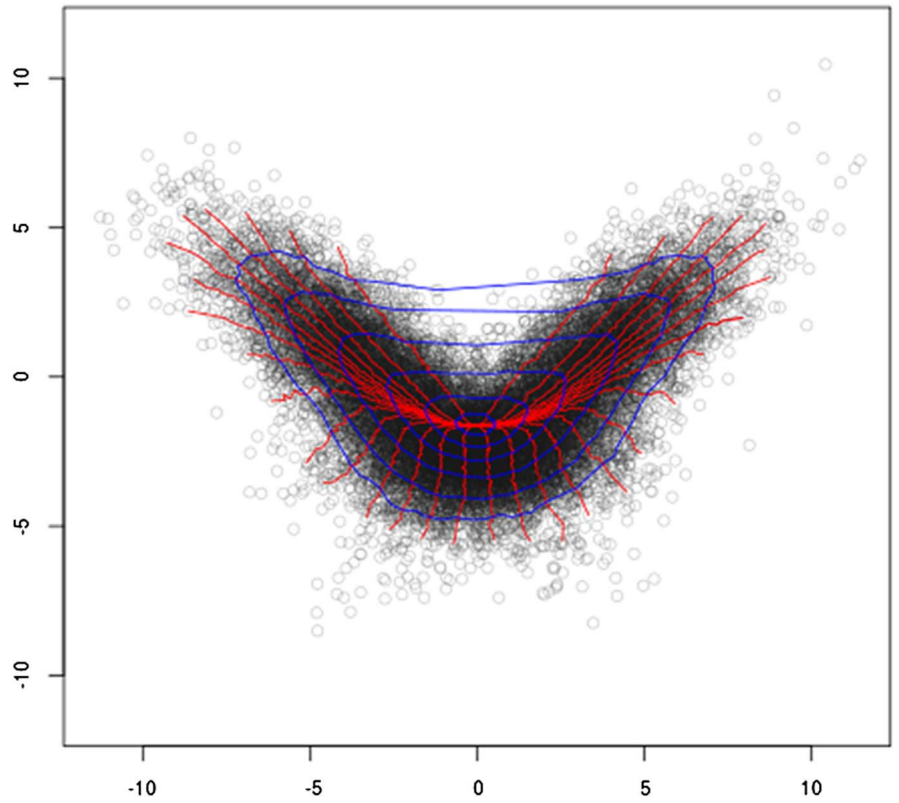

FIG. 4. *Center-outward quantile contours and sign curves for the same Gaussian mixture as in the middle panel of Figure 3, with* $n = 20{,}000$.

curves roughly has the same probability contents, Figure 4 provides graphical evidence of a very low density in the central concavity bridged by the quantile 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, Drton and Han (2019), Deb and Sen (2019), Ghosal and Sen (2019), Hallin, La Vecchia and Liu (2020), Hallin, La Vecchia and Liu (2020), Hallin, Hlubinka and Hudecová (2020), etc. A number of questions remain open,[19] 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?

(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 Maha-

[19] Some of them have been solved or partially solved since these lines were written (added in proof).

lanobis ranks and signs with the center-outward ones. Can we similarly construct one-step R-estimators? This, which requires Hájek-type asymptotic representation results,[20] 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, for example, on Kolmogorov–Smirnov or Cramér–von Mises distances between center-outward distribution functions?[21]

(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 use them in the construction of multivariate median or sign tests? can we, on the model of Carlier, Chernozhukov and Galichon (2016) or Hallin, Paindaveine and Šiman (2010), Hallin, Lu, Paindaveine and Šiman (2015), perform parametric and nonparametric multiple-output quantile regression? construct multivariate growth charts (as in McKeague et al. (2011))?

(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, values at risk or expected shortfalls; see Gushchin and Borzykh (2017), Beirlant et al. (2020). 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.

Acknowledgments. This paper results from the merging of Hallin (2017) and del Barrio, Cuesta-Albertos, Hallin and Matrán (2018). Eustasio del Barrio, Juan Cuesta-Albertos and Carlos Matrán are supported in part by FEDER, Spanish Ministerio de Economía y Competitividad, grant MTM2017-86061-C2; Eustasio del Barrio and Carlos Matrán also acknowledge the support of the Junta de Castilla y León, grants VA005P17 and VA002G18. Marc Hallin thanks Marc Henry for guiding his first steps into the subtleties of measure transportation.

SUPPLEMENTARY MATERIAL

Supplement to "Distribution and quantile functions, ranks, and signs in dimension d: A measure transportation approach" (DOI: 10.1214/20-AOS1996SUPP; .pdf). Supplementary material and technical proofs can be found in an Online Appendix (Hallin et al. (2020)) consisting of eight sections labeled A, B, ..., H, along with additional references.

REFERENCES

AHUJA, R. K., MAGNANTI, T. L. and ORLIN, J. B. (1993). *Network Flows*: *Theory*, *Algorithms*, *and Applications*. Prentice Hall, Englewood Cliffs, NJ. MR1205775

BASU, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhyā* **15** 377–380. MR0074745 https://doi.org/10.1007/978-1-4419-5825-9_14

BASU, D. (1959). The family of ancillary statistics. *Sankhyā* **21** 247–256. MR0110115 https://doi.org/10.1007/978-1-4419-5825-9_18

[20] Such results have been obtained in Hallin, Hlubinka and Hudecová (2020) for multiple-output regression, in Hallin, La Vecchia and Liu (2020) for VAR models (added in proof).

[21] See Hallin, Mordant and Segers (2020) for an answer (added in proof).

BEIRLANT, J., DEL BARRIO, E., BUITENDAG, S., HALLIN, M. and KAMPER, FR. (2020). Center-outward quantiles and the measurement of multivariate risk. *Insurance Math. Econom.* **95** 79–100.

BELLONI, A. and WINKLER, R. L. (2011). On multivariate quantiles under partial orders. *Ann. Statist.* **39** 1125–1179. MR2816350 https://doi.org/10.1214/10-AOS863

BERTSEKAS, D. P. (1991). *Linear Network Optimization*: *Algorithms and Codes*. MIT Press, Cambridge, MA. MR1201582

BICKEL, P. J. (1965). On some asymptotically nonparametric competitors of Hotelling's T^2. *Ann. Math. Stat.* **36** 160–173; correction, ibid. 1583. MR0181052

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. MR3286925 https://doi.org/10.1093/biomet/asu045

BOECKEL, M., SPOKOINY, V. and SUVORIKOVA, A. (2018). Multivariate Brenier cumulative distribution functions and their application to nonparametric testing. Available at arXiv:1809.04090.

BURKARD, R., DELL'AMICO, M. and MARTELLO, S. (2009). *Assignment Problems*. SIAM, Philadelphia, PA. MR2488749 https://doi.org/10.1137/1.9780898717754

CHAKRABORTY, B. and CHAUDHURI, P. (1996). On a transformation and re-transformation technique for constructing an affine equivariant multivariate median. *Proc. Amer. Math. Soc.* **124** 2539–2547. MR1363452 https://doi.org/10.1090/S0002-9939-96-03657-X

CHAKRABORTY, B. and CHAUDHURI, P. (1998). On an adaptive transformation-retransformation estimate of multivariate location. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 145–157. MR1625644 https://doi.org/10.1111/1467-9868.00114

CHAKRABORTY, A. and CHAUDHURI, P. (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *Ann. Statist.* **42** 1203–1231. MR3224286 https://doi.org/10.1214/14-AOS1226

CARLIER, G., CHERNOZHUKOV, V. and GALICHON, A. (2016). Vector quantile regression: an optimal transport approach. *Ann. Statist.* **44** 1165–1192.

CHAKRABORTY, A. and CHAUDHURI, P. (2017). Tests for high-dimensional data based on means, spatial signs and spatial ranks. *Ann. Statist.* **45** 771–799. MR3650400 https://doi.org/10.1214/16-AOS1467

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

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. MR1245371

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

CHOI, K. and MARDEN, J. (1997). An approach to multivariate rank tests in multivariate analysis of variance. *J. Amer. Statist. Assoc.* **92** 1581–1590. MR1615267 https://doi.org/10.2307/2965429

CUESTA-ALBERTOS, J. A., MATRÁN, C. and TUERO-DÍAZ, A. (1997). Optimal transportation plans and convergence in distribution. *J. Multivariate Anal.* **60** 72–83. MR1441460 https://doi.org/10.1006/jmva.1996.1627

DE VALK, C. and SEGERS, J. (2018). Stability and tail limits of transport-based quantile contours. Available at arXiv:1811.12061.

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

DEL BARRIO, E., GONZÁLEZ-SANZ, A. and HALLIN, M. (2020). A note on the regularity of center-outward distribution and quantile functions. *J. Multivariate Anal.* **180** S0047259X20302529.

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

DICK, J. and PILLICHSHAMMER, F. (2014). Discrepancy theory and quasi-Monte Carlo integration. In *A Panorama of Discrepancy Theory* (W. Chen, A. Srivastava and G. Travaglini, eds.). *Lecture Notes in Math.* **2107** 539–619. Springer, Cham. MR3330331 https://doi.org/10.1007/978-3-319-04696-9_9

FIGALLI, A. (2017). *The Monge–Ampère Equation and Its Applications*. *Zurich Lectures in Advanced Mathematics*. European Mathematical Society (EMS), Zürich. MR3617963 https://doi.org/10.4171/170

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

GENEST, C. and RIVEST, L.-P. (2001). On the multivariate probability integral transformation. *Statist. Probab. Lett.* **53** 391–399. MR1856163 https://doi.org/10.1016/S0167-7152(01)00047-5

GHOSAL, P. and SEN, B. (2019). Multivariate ranks and quantiles using optimal transportation and applications to goodness-of-fit testing. Available at arXiv:1905.05340.

GUSHCHIN, A. A. and BORZYKH, D. A. (2017). Integrated quantile functions: Properties and applications. *Mod. Stoch. Theory Appl.* **4** 285–314. MR3739012 https://doi.org/10.15559/17-vmsta88

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.

DEL BARRIO, E., CUESTA-ALBERTOS, J., HALLIN, M. and MATRÁN, C. (2018). Smooth cyclically monotone interpolation and empirical center-outward distribution functions. Available at arXiv:1806.01238.

HALLIN, M., HLUBINKA, D. and HUDECOVÁ, S. (2020). Fully distribution-free center-outward rank tests for multiple-output regression and MANOVA. Available at arXiv:2007.15496.

HALLIN, M., INGENBLEEK, J.-F. and PURI, M. L. (1989). Asymptotically most powerful rank tests for multivariate randomness against serial dependence. *J. Multivariate Anal.* **30** 34–71. MR1003708 https://doi.org/10.1016/0047-259X(89)90087-0

HALLIN, M., MORDANT, G. and SEGERS, J. (2019). Multivariate goodness-of-fit tests based on Wasserstein distance. Available at arXiv:2003.06684.

HALLIN, M., LA VECCHIA, D. and LIU, H. (2020). Center-outward R-estimation for semiparametric VARMA models. *J. Amer. Statist. Assoc.* https://doi.org/10.1080/01621459.2020.1832501

HALLIN, M., LA VECCHIA, D. and LIU, H. (2020). Rank-based testing for semiparametric VAR models: a measure transportation approach. Available at arXiv:2011.06062.

HALLIN, M., LU, Z., PAINDAVEINE, D. and ŠIMAN, M. (2015). Local bilinear multiple-output quantile regression. *Bernoulli* **21** 1435–1466.

HALLIN, M. and MEHTA, C. (2015). *R*-estimation for asymmetric independent component analysis. *J. Amer. Statist. Assoc.* **110** 218–232. MR3338498 https://doi.org/10.1080/01621459.2014.909316

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. MR2329466 https://doi.org/10.1214/009053606000000948

HALLIN, M. and PAINDAVEINE, D. (2002a). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.* **30** 1103–1133. MR1926170 https://doi.org/10.1214/aos/1031689019

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. MR1963662

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 Methods* (*Neuchâtel*, 2002). *Stat. Ind. Technol.* 271–282. Birkhäuser, Basel. MR2001322

HALLIN, M. and PAINDAVEINE, D. (2004a). Rank-based optimal tests of the adequacy of an elliptic VARMA model. *Ann. Statist.* **32** 2642–2678. MR2153998 https://doi.org/10.1214/009053604000000724

HALLIN, M. and PAINDAVEINE, D. (2004b). Multivariate signed-rank tests in vector autoregressive order identification. *Statist. Sci.* **19** 697–711. MR2185591 https://doi.org/10.1214/088342304000000602

HALLIN, M. and PAINDAVEINE, D. (2005). Affine-invariant aligned rank tests for the multivariate general linear model with VARMA errors. *J. Multivariate Anal.* **93** 122–163. MR2119768 https://doi.org/10.1016/j.jmva.2004.01.005

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. MR2329465 https://doi.org/10.1214/009053606000000731

HALLIN, M. and PAINDAVEINE, D. (2006b). Parametric and semiparametric inference for shape: The role of the scale functional. *Statist. Decisions* **24** 327–350. MR2305111 https://doi.org/10.1524/stnd.2006.24.3.327

HALLIN, M. and PAINDAVEINE, D. (2008). Optimal rank-based tests for homogeneity of scatter. *Ann. Statist.* **36** 1261–1298. MR2418657 https://doi.org/10.1214/07-AOS508

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

HALLIN, M., PAINDAVEINE, D. and VERDEBOUT, T. (2010). Optimal rank-based testing for principal components. *Ann. Statist.* **38** 3245–3299. MR2766852 https://doi.org/10.1214/10-AOS810

HALLIN, M., PAINDAVEINE, D. and VERDEBOUT, T. (2013). Optimal rank-based tests for common principal components. *Bernoulli* **19** 2524–2556. MR3160563 https://doi.org/10.3150/12-BEJ461

HALLIN, M., PAINDAVEINE, D. and VERDEBOUT, T. (2014). Efficient R-estimation of principal and common principal components. *J. Amer. Statist. Assoc.* **109** 1071–1083. MR3265681 https://doi.org/10.1080/01621459.2014.880057

HALLIN, M. and WERKER, B. J. M. (1999). Optimal testing for semi-parametric AR models—From Gaussian Lagrange multipliers to autoregression rank scores and adaptive tests. In *Asymptotics, Nonparametrics, and Time Series* (S. Ghosh, ed.). *Statist. Textbooks Monogr.* **158** 295–350. Dekker, New York. MR1724702

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

HALLIN, M., DEL BARRIO, E., CUESTA-ALBERTOS, J. A. and MATRÁN, C. (2021). Supplement to "Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach." https://doi.org/10.1214/20-AOS1996SUPP

HAMEL, A. H. and KOSTNER, D. (2018). Cone distribution functions and quantiles for multivariate random variables. *J. Multivariate Anal.* **167** 97–113. MR3830636 https://doi.org/10.1016/j.jmva.2018.04.004

HE, X. and WANG, G. (1997). Convergence of depth contours for multivariate datasets. *Ann. Statist.* **25** 495–504. MR1439311 https://doi.org/10.1214/aos/1031833661

HODGES, J. L. JR. (1955). A bivariate sign test. *Ann. Math. Stat.* **26** 523–527. MR0070921 https://doi.org/10.1214/aoms/1177728498

ILMONEN, P. and PAINDAVEINE, D. (2011). Semiparametrically efficient inference based on signed ranks in symmetric independent component models. *Ann. Statist.* **39** 2448–2476. MR2906874 https://doi.org/10.1214/11-AOS906

JUDD, K. L. (1998). *Numerical Methods in Economics*. MIT Press, Cambridge, MA. MR1646331

KARP, R. M. (1978). A characterization of the minimum cycle mean in a digraph. *Discrete Math.* **23** 309–311. MR0523080 https://doi.org/10.1016/0012-365X(78)90011-0

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

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

LEHMANN, E. L. and SCHOLZ, F.-W. (1992). Ancillarity. In *Current Issues in Statistical Inference*: *Essays in Honor of D. Basu. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **17** 32–51. IMS, Hayward, CA. MR1194408 https://doi.org/10.1214/lnms/1215458837

LIU, R. Y. (1992). Data depth and multivariate rank tests. In L_1-*Statistical Analysis and Related Methods* (*Neuchâtel*, 1992) 279–294. North-Holland, Amsterdam. MR1214839

LIU, R. Y. and SINGH, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.* **88** 252–260. MR1212489

LÓPEZ-PINTADO, S. and ROMO, J. (2009). On the concept of depth for functional data. *J. Amer. Statist. Assoc.* **104** 718–734. MR2541590 https://doi.org/10.1198/jasa.2009.0108

MARDEN, J. I. (1999). Multivariate rank tests. In *Multivariate Analysis*, *Design of Experiments*, *and Survey Sampling*. *Statist. Textbooks Monogr.* **159** 401–432. Dekker, New York. MR1719102

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

MCKEAGUE, I. W., LÓPEZ-PINTADO, S., HALLIN, M. and ŠIMAN, M. (2011). Analyzing growth trajectories. *J. Dev. Orig. Health Dis.* **2** 322–329.

MÖTTÖNEN, J. and OJA, H. (1995). Multivariate spatial sign and rank methods. *J. Nonparametr. Stat.* **5** 201–213. MR1346895 https://doi.org/10.1080/10485259508832643

MÖTTÖNEN, J., OJA, H. and TIENARI, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *Ann. Statist.* **25** 542–552. MR1439313 https://doi.org/10.1214/aos/1031833663

NIEDERREITER, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. *CBMS–NSF Regional Conference Series in Applied Mathematics* **63**. SIAM, Philadelphia, PA. MR1172997 https://doi.org/10.1137/1.9781611970081

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. MR2498716 https://doi.org/10.1016/j.jmva.2008.08.004

OJA, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates: A review. *Scand. J. Stat.* **26** 319–343. MR1712063 https://doi.org/10.1111/1467-9469.00152

OJA, H. (2010). *Multivariate Nonparametric Methods with R*: *An Approach Based on Spatial Signs and Ranks*. *Lecture Notes in Statistics* **199**. Springer, New York. MR2598854 https://doi.org/10.1007/978-1-4419-0468-3

OJA, H. and PAINDAVEINE, D. (2005). Optimal signed-rank tests based on hyperplanes. *J. Statist. Plann. Inference* **135** 300–323. MR2200471 https://doi.org/10.1016/j.jspi.2004.04.022

OJA, H. and RANDLES, R. H. (2004). Multivariate nonparametric tests. *Statist. Sci.* **19** 598–605. MR2185581 https://doi.org/10.1214/088342304000000558

PATEIRO-LÓPEZ, B. and RODRÍGUEZ-CASAL, A. (2010). Generalizing the convex hull of a sample: The R package α-hull. *J. Stat. Softw.* **34**.

PEYRÉ, G. and CUTURI, M. (2019). Computational optimal transport. *Found. Trends Mach. Learn.* **11** 355–607.

PURI, M. L. and SEN, P. K. (1966). On a class of multivariate multisample rank-order tests. *Sankhyā Ser. A* **28** 353–376. MR0231488

PURI, M. L. and SEN, P. K. (1969). A class of rank order tests for a general linear hypothesis. *Ann. Math. Stat.* **40** 1325–1343. MR0245155 https://doi.org/10.1214/aoms/1177697505

PURI, M. L. and SEN, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley, New York. MR0298844

RANDLES, R. H. (1989). A distribution-free multivariate sign test based on interdirections. *J. Amer. Statist. Assoc.* **84** 1045–1050. MR1134492

ROCKAFELLAR, R. T. (1966). Characterization of the subdifferentials of convex functions. *Pacific J. Math*. **17** 497–510. MR0193549

ROCKAFELLAR, R. T. and WETS, R. J.-B. (1998). *Variational Analysis*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **317**. Springer, Berlin. MR1491362 https://doi.org/10.1007/978-3-642-02431-3

SANTNER, T. J., WILLIAMS, B. J. and NOTZ, W. I. (2003). *The Design and Analysis of Computer Experiments*. *Springer Series in Statistics*. Springer, New York. MR2160708 https://doi.org/10.1007/978-1-4757-3799-8

SEGERS, J., VAN DEN AKKER, R. and WERKER, B. J. M. (2014). Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation. *Ann. Statist*. **42** 1911–1940. MR3262472 https://doi.org/10.1214/14-AOS1244

SEN, P. K. and PURI, M. L. (1967). On the theory of rank order tests for location in the multivariate one sample problem. *Ann. Math. Stat*. **38** 1216–1228. MR0212954 https://doi.org/10.1214/aoms/1177698790

SERFLING, R. (2002). Quantile functions for multivariate analysis: Approaches and applications. *Stat. Neerl*. **56** 214–232. MR1916321 https://doi.org/10.1111/1467-9574.00195

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

UM, Y. and RANDLES, R. H. (1998). Nonparametric tests for the multivariate multi-sample location problem. *Statist. Sinica* **8** 801–812. MR1651509

YOSIDA, K. (1965). *Functional Analysis*. *Die Grundlehren der Mathematischen Wissenschaften* **123**. Academic Press, New York; Springer, Berlin. MR0180824

ZUO, Y. (2018). On general notions of depth for regression. Available at arXiv:1805.02046v1.

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. MR2329471 https://doi.org/10.1214/009053606000000876

ZUO, Y. and SERFLING, R. (2000). General notions of statistical depth function. *Ann. Statist*. **28** 461–482. MR1790005 https://doi.org/10.1214/aos/1016218226

Further references are provided in the Online Appendix.