ECARES

Smooth Cyclically Monotone Interpolation and Empirical Center-Outward Distribution Functions

Eustasio del Barrio
Departamento de Estadistica e Investigacion Operativa,
Facultad de Ciencias, Universidad de Valladolid, Spain

Juan Cuesta Albertos
Departamento de Matematicas, Facultad de Ciencias,
Universidad de Cantabria, Santander, Spain

Marc Hallin
ECARES, Université libre de Bruxelles, Belgium

Carlos Matran
Departamento de Estadistica e Investigacion Operativa,
Facultad de Ciencias, Universidad de Valladolid, Spain

May 2018

ECARES working paper 2018-15

ECARES
ULB - CP 114/04
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

Smooth Cyclically Monotone Interpolation and Empirical Center-Outward Distribution Functions

Eustasio del Barrio,[1] **Juan Cuesta Albertos,**[2] **Marc Hallin,**[3] **and Carlos Matrán**[4]

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

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

[3] *e-mail:* mhallin@ulb.ac.be *ECARES, Université libre de Bruxelles, Belgium*

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

Abstract: We consider the smooth interpolation problem under cyclical monotonicity constraint. More precisely, consider finite n-tuples $\mathcal{X} = \{x_1, \ldots, x_n\}$ and $\mathcal{Y} = \{y_1, \ldots, y_n\}$ of points in $\mathbb{R}^d$, and assume the existence of a unique bijection $T : \mathcal{X} \to \mathcal{Y}$ such that $\{(x, T(x)) : x \in \mathcal{X}\}$ is cyclically monotone: our goal is to define continuous, cyclically monotone maps $\bar{T} : \mathbb{R}^d \to \mathbb{R}^d$ such that $\bar{T}(x_i) = y_i$, $i = 1, \ldots, n$, extending a classical result by Rockafellar on the subdifferentials of convex functions. Our solutions $\bar{T}$ are Lipschitz, and we provide a sharp lower bound for the corresponding Lipschitz constants. The problem is motivated by, and the solution naturally applies to, the concept of empirical center-outward distribution function in $\mathbb{R}^d$ developed in Hallin (2018). Those empirical distribution functions indeed are defined at the observations only. Our interpolation provides a smooth extension, as well as a multivariate, outward-continuous, jump function version thereof (the latter naturally generalizes the traditional left-continuous univariate concept); both satisfy a Glivenko-Cantelli property as $n \to \infty$.

1. Introduction

1.1. Smooth interpolation under cyclical monotonicity constraint

A subset S of $\mathbb{R}^d \times \mathbb{R}^d$ is said to be *cyclically monotone* if, for any finite collection of points $\{(x_1, y_1), \ldots, (x_k, y_k)\} \subseteq S$, denoting by $\langle x, y\rangle$ the scalar product of x and y in $\mathbb{R}^d$,

$$\langle y_1, x_2 - x_1\rangle + \langle y_2, x_3 - x_2\rangle + \ldots + \langle y_k, x_1 - x_k\rangle \leq 0. \tag{1.1}$$

A mapping T from $\mathbb{R}^d$ to $\mathbb{R}^d$ is said to be *cyclically monotone* iff $\{(x, T(x)) \,|\, x \in \mathbb{R}^d\}$ is cyclically monotone.

Note that a finite subset $S = \{(x_1, y_1), \dots, (x_n, y_n)\}$ of $\mathbb{R}^d \times \mathbb{R}^d$ is cyclically monotone if and only if (1.1) holds for $k = n$—equivalently, iff, among all pairings of $\mathcal{X} := (x_1, \dots, x_n)$ and $\mathcal{Y} := (y_1, \dots, y_n)$, S maximizes $\sum_{i=1}^n \langle x_i, y_i \rangle$ (that is, maximizes an empirical correlation), or minimizes $\sum_{i=1}^n \|y_i - x_i\|^2$, where $\|x\|$ stands for the Euclidean norm of $x \in \mathbb{R}^d$.

Let $\mathcal{X} = \{x_1, \dots, x_n\}$ and $\mathcal{Y} = \{y_1, \dots, y_n\}$ denote two n-tuples of points in in $\mathbb{R}^d$. Assuming that there exists a *unique* bijection $T : \mathcal{X} \to \mathcal{Y}$ such that $\{(x, T(x)) | \ x \in \mathcal{X}\}$ is cyclically monotone, there is no loss of generality in relabeling $\mathcal{Y}$ so that $y_i = T(x_i)$. Accordingly, we throughout, are making the following assumption.

Assumption (A). The n-tuples $\mathcal{X}$ and $\mathcal{Y}$ are such that $T : x_i \mapsto T(x_i) = y_i$, $i = 1, \dots, n$ is the unique cyclically monotone bijective map from $\mathcal{X}$ to $\mathcal{Y}$.

We consider, under Assumption (A), the smooth interpolation problem under cyclical monotonicity constraint. More precisely, our goal is to construct a smooth (at least, continuous), cyclically monotone map $\bar{T} : \mathbb{R}^d \to \mathbb{R}^d$ such that $\bar{T}(x_i) = T(x_i) = y_i$ for $i = 1, \dots, n$.

It is well known that the subdifferential of a convex function Ψ from $\mathbb{R}^d$ to $\mathbb{R}$ enjoys cyclical monotonicity. A classical result by Rockafellar (1966) establishes the converse: any cyclically monotone subset $S = \{(x_i, y_i) | i = 1, \dots, n\}$ of $\mathbb{R}^d \times \mathbb{R}^d$ is contained 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 x) subgradient $\nabla\Psi(x)$ and subdifferential $\{(x, \nabla\Psi(x)) | x \in \mathbb{R}^d\}$. When Ψ is convex and differentiable, the mapping $x \mapsto \nabla\Psi(x)$ thus enjoys cyclical monotonicity. We show (Corollary 2.3) that, conversely, any cyclically monotone subset $S = \{(x_i, y_i) | i = 1, \dots, n\}$ of $\mathbb{R}^d \times \mathbb{R}^d$ is the subdifferential (at x_i, $i = 1, \dots, n$) of some (continuously) differentiable convex function Ψ.

Note that Assumption (A) holds if and only if identity is the unique minimizer of

$$\frac{1}{n} \sum_{i=1}^n \|x_i - y_{\sigma(i)}\|^2 \tag{1.2}$$

among the set of all permutations σ of $\{1, \dots, n\}$. The same condition can be recast in terms of uniqueness of the solution of the linear program

$$\begin{aligned} &\min_\pi \sum_{i=1}^n \sum_{j=1}^n c_{i,j} \pi_{i,j} \\ &\text{s.t. } \sum_{i=1}^n \pi_{i,j} = \sum_{j=1}^n \pi_{i,j} = \frac{1}{n}, \\ &\qquad \pi_{i,j} \geq 0, \ i, j = 1, \dots, n, \end{aligned} \tag{1.3}$$

with $c_{i,j} = \|x_i - y_j\|^2$: clearly, (1.2) holds if and only if $\pi_{i,i} = \frac{1}{n}$, $\pi_{i,j} = 0$, $j \neq i$ is the unique solution of (1.3).

1.2. Center-outward distribution functions in $\mathbb{R}^d$

This cyclical monotone interpolation problem naturally arises in relation with the measure transportation-based concept of *center-outward distribution function* considered in [5], which we now describe.

Denote by $\mathcal{P}_d$ the family of nonvanishing Lebesgue-absolutely continuous probability measures over $\mathbb{R}^d$, more precisely, define $\mathcal{P}_d$ as the set of probability distributions P with a density f such that, for all $D \in \mathbb{R}^+$, there exist $0 < \lambda_{D;f} \leq \Lambda_{D;f} < \infty$ such that $\lambda_{D;f} \leq f(x) \leq \Lambda_{D;f}$ for all x such that $\|x\| \leq D$. Building on [3] and [4], [5] defines the center-outward distribution function $F_\pm$ of $\mathrm{P} \in \mathcal{P}_d$ as the unique gradient of a convex function pushing P forward to the uniform U_d over the unit ball $\mathbb{S}_d$ in $\mathbb{R}^d$. By *uniform* here we mean the product measure of a uniform distribution over the directions (the unit sphere $\mathcal{S}_{d-1}$) and a uniform distribution over the distances to the origin (the unit interval $[0, 1]$); this reduces to Lebesgue-uniformity for $d = 1$, but not for $d \geq 2$. The center-outward distribution function $F_\pm$ of $\mathrm{P} \in \mathcal{P}_d$ is shown to be a continuous bijective cyclically monotone mapping from $\mathbb{R}^d$ to $\mathbb{S}_d$.

The empirical counterpart $F_\pm^{(n)}$ of $F_\pm$ is defined as a cyclically monotone (discrete) mapping from the (random) sample $X_1^{(n)}, \ldots, X_n^{(n)}$ to a (nonrandom) "regular" grid over $\mathbb{S}_d$. That grid is built as the intersection of the collection $\{0, \frac{1}{n_R+1}, \ldots, \frac{n_R}{n_R+1}\}\mathcal{S}_{d-1}$ of $(n_R + 1)$ nested hyperspheres with a collection of n_S radii from the origin, where n_R and n_S are such that $n_0 := n - n_R n_S < \min(n_R, n_S)$ (the origin, in that grid, is given multiplicity n_0; see Section 4.2 of [5] for details). Hence, $F_\pm^{(n)}$ is defined at the observed points $X_1^{(n)}, \ldots, X_n^{(n)}$ only. A Glivenko-Cantelli theorem, of the form

$$\max_{1 \leq i \leq n} \|F_\pm^{(n)}(X_i^{(n)}) - F_\pm(X_i^{(n)})\| \quad \text{a.s., as } n \to \infty \tag{1.4}$$

is established (without any moment assumptions) for i.i.d. samples with probability distribution $\mathrm{P} \in \mathcal{P}_d$ and center-outward distribution function $F_\pm$.

The empirical center-outward distribution function $F_\pm^{(n)}$, irrespective of d, carries the same information as the sample itself, and perfectly fulfills its statistical role as a sample summary. One may like, however, to define it as an object of the same nature—a smooth cyclically monotone mapping from $\mathbb{R}^d$ to $\mathbb{S}_d$—as its population counterpart $F_\pm$. This brings into the picture the problem of the existence and construction, within the class of gradients of convex functions, of a continuous extension $x \mapsto \bar{F}_\pm^{(n)}(x)$ of the discrete $F_\pm^{(n)}$, yielding a Glivenko-Cantelli theorem of the form

$$\sup_{x \in \mathbb{R}^d} \|\bar{F}_\pm^{(n)}(x) - \bar{F}_\pm(x)\| \to 0, \quad \text{a.s., as } n \to \infty. \tag{1.5}$$

That problem, which is left open in [5], reduces to the smooth interpolation problem considered here, and is discussed in Section 3.1.

Now, the traditional definition of an empirical distribution function in dimension one ($d = 1$) yields, in obvious standard notation, a right-continuous step function interpolation of the couples $\{(X_1^{(n)}, F^{(n)}(X_1^{(n)})), \ldots, (X_n, F^{(n)}(X_n^{(n)}))\}$ (although any smooth, monotone nondecreasing interpolation would do the same job—satisfying, in particular, a Glivenko-Cantelli result under sup form as in (1.5)). That traditional definition actually is mapping the real line $\mathbb{R}$ to a regular grid of $(n+1)$ points, $\{0, \frac{1}{n+1}, \ldots, \frac{n}{n+1}\}$ (the denominator $n+1$ is adopted for convenience, and without any loss of generality, in order for $F^{(n)}(X_i^{(n)})$, $i = 1, \ldots, n$ to take values in the open unit interval).

Due to the lack of a canonical ordering of $\mathbb{R}^d$ ($d \geq 2$), the problem of extending this step function version of $F^{(n)}$ to the center-outward d-dimensional situation is much less obvious, the very concept of a step function being unclear. In Section 3.2, we show how the smooth interpolation constructed in Section 3.1 allows for a natural definition of such a step function interpolation, yielding cyclically monotone outward-continuous mappings $\bar{F}_{\pm}^{(n)*}$ from $\mathbb{R}^d$ to the nested collection $\{0, \frac{1}{n_R+1}, \ldots, \frac{n_R}{n_R+1}\}\mathcal{S}_{d-1}$ of (n_R+1) hyperspheres characterizing $F_{\pm}^{(n)}$. Those mappings still enjoy a sup form of Glivenko-Cantelli (as in (1.5)); instead of *steps*, they yield *plateaux* (*hyperplateaux* for $d \geq 3$), the boundaries of which are the continuous *quantile contours* or *hypersurfaces*

$$\big(\bar{F}_{\pm}^{(n)}\big)^{-1}\Big(\frac{r}{n+1}\mathcal{S}_{d-1}\Big) = \big(\bar{F}_{\pm}^{(n)*}\big)^{-1}\Big(\frac{r}{n+1}\mathcal{S}_{d-1}\Big), \qquad r = 1, \ldots, n_R.$$

Contrary to the univariate situation, however, $\bar{F}_{\pm}^{(n)*}$ is not uniquely defined: distinct smooth interpolations $\bar{F}_{\pm}^{(n)}$ (as described in Section 3.1) may produce distinct versions of $\bar{F}_{\pm}^{(n)*}$, with distinct discontinuity contours; all of them of course coincide at the observed points.

2. Cyclically monotone interpolation

We now turn back to the smooth interpolation problem described in Section 1.1, where we refer to for notation and assumptions. Our solution is constructed in two steps. First (Step 1), we extend T to a piecewise constant cyclically monotone map defined on a set in $\mathbb{R}^d$ whose complementary has zero Lebesgue measure. Being piecewise constant, that map cannot be smooth. To fix this problem, we apply (Step 2) a regularization procedure yielding the required smoothness while keeping the interpolation feature. For Step 1, we rely on the following result.

Proposition 2.1. *Assume that $x_1, \ldots, x_n \in \mathbb{R}^d$ and $y_1, \ldots, y_n \in \mathbb{R}^d$ are such that $i \neq j$ implies $x_i \neq x_j$ and $y_i \neq y_j$. Then,*
(i) the map $T(x_i) = y_i$, $i = 1, \ldots, n$ is cyclically monototone if and only if there exist real numbers $\psi_1, \ldots, \psi_n$ such that

$$\langle x_i, y_i \rangle - \psi_i = \max_{j=1,\ldots,n} (\langle x_i, y_j \rangle - \psi_j), \quad i = 1, \ldots, n;$$

(ii) furthermore, T is the unique cyclically monototone map from $\{x_1, \ldots, x_n\}$ to $\{y_1, \ldots, y_n\}$ if and only if there exist real numbers $\psi_1, \ldots, \psi_n$ such that

$$\langle x_i, y_i \rangle - \psi_i > \max_{j=1,\ldots,n, j\neq i} (\langle x_i, y_j \rangle - \psi_j), \quad i = 1, \ldots, n. \tag{2.1}$$

Proof. Duality yields, for the linear program (1.3),

$$\begin{aligned} &\min_{\pi} \sum_{i=1}^{n}\sum_{j=1}^{n} c_{i,j}\pi_{i,j} && = \quad \max_{a,b} \frac{1}{n}\sum_{i=1}^{n} a_i + \frac{1}{n}\sum_{j=1}^{n} b_j \\ &\text{s.t. } \sum_{i=1}^{n}\pi_{i,j} = \sum_{j=1}^{n}\pi_{i,j} = \frac{1}{n}, && \qquad \text{s.t. } a_i + b_j \le c_{i,j},\ i,j = 1,\dots,n. \\ &\quad \pi_{i,j} \ge 0,\ i,j = 1,\dots,n \end{aligned} \tag{2.2}$$

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

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

With the change of variables $a_i =: \|x_i\|^2 - 2\varphi_i$, $b_j =: \|y_j\|^2 - 2\psi_j$, the dual programs (2.2) take the form

$$\begin{aligned} &\max_{\pi} \sum_{i=1}^{n}\sum_{j=1}^{n} \pi_{i,j}\langle x_i, y_j\rangle && = \quad \min_{\varphi,\psi} \frac{1}{n}\sum_{i=1}^{n} \varphi_i + \frac{1}{n}\sum_{j=1}^{n} \psi_j \\ &\text{s.t. } \sum_{i=1}^{n}\pi_{i,j} = \sum_{j=1}^{n}\pi_{i,j} = \frac{1}{n}, && \qquad \text{s.t. } \varphi_i + \psi_j \ge \langle x_i, y_j\rangle,\ i,j = 1,\dots,n \\ &\quad \pi_{i,j} \ge 0,\ i,j = 1,\dots,n \end{aligned} \tag{2.3}$$

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

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \pi_{i,j}\langle x_i, y_j\rangle = \frac{1}{n}\sum_{i=1}^{n} \varphi_i + \frac{1}{n}\sum_{j=1}^{n} \psi_j.$$

Let (φ,ψ) be a minimizer for the right-hand side program in (2.3). Then, replacing φ_i with $\tilde{\varphi}_i := \max_{j=1,\dots,n}(\langle x_i, y_j\rangle - \psi_j)$ yields a new feasible solution $(\tilde{\varphi},\psi)$ satisfying $\varphi_i \ge \tilde{\varphi}_i$. Optimality of (φ,ψ) thus implies that $\varphi_i = \tilde{\varphi}_i$, so that, at optimality,

$$\varphi_i = \max_{j=1,\dots,n}(\langle x_i, y_j\rangle - \psi_j), \quad i = 1,\dots,n. \tag{2.4}$$

Now, if condition (1.2) holds, then $\pi_{i,i} = 1/n$, $\pi_{i,j} = 0$, $j \ne i$ is the unique maximizer in the left-hand side linear program in (2.3). Therefore, (φ,ψ) is a minimizer for the right-hand side program if and only if

$$\frac{1}{n}\sum_{i=1}^{n}(\varphi_i + \psi_i - \langle x_i, y_i\rangle) = 0.$$

In view of (2.4) this implies that

$$\langle x_i, y_i\rangle - \psi_i = \max_{j=1,\dots,n}(\langle x_i, y_j\rangle - \psi_j), \quad i = 1, \dots, n. \tag{2.5}$$

Conversely, assume that the weights $\psi_1, \dots, \psi_n$ are such that (2.5) holds. Then, letting $\varphi_i = \max_{j=1,\dots,n}(\langle x_i, y_j\rangle - \psi_j)$, we have that (φ, ψ) is a feasible solution for which

$$\frac{1}{n}\sum_{i=1}^{n}(\varphi_i + \psi_i - \langle x_i, y_i\rangle) = 0,$$

which, in view of the discussion above, implies that the map $T : x_i \mapsto T(x_i) = y_i$ is cyclically monotone. This completes the proof of Part *(i)* of the lemma.

As for Part *(ii)* of the proposition, note that T is the unique cyclically monotone map from $\{x_1, \dots, x_n\}$ to $\{y_1, \dots, y_n\}$ if and only if, for every choice of indices $\{i_0, i_1, \dots, i_m\} \subseteq \{1, \dots, n\}$, we have

$$\langle x_{i_0}, y_{i_0} - y_{i_1}\rangle + \langle x_{i_1}, y_{i_0} - y_{i_2}\rangle + \cdots + \langle x_{i_m}, y_{i_m} - y_{i_0}\rangle > 0, \tag{2.6}$$

while (2.1) holds if and only if there exist real numbers $\psi_1, \dots, \psi_n$ such that

$$\langle x_i, y_i - y_j\rangle > \psi_i - \psi_j \quad \text{ for all } i \neq j.$$

On the other hand, defining $f_{i,j}(\psi) := \psi_i - \psi_j - \langle x_i, y_i - y_j\rangle$ for $i \neq j$, we can apply Farkas' Lemma (see, e.g., Theorem 21.1. in [7]) to see that either there exists $\psi \in \mathbb{R}^n$ such that $f_{i,j}(\psi) < 0$ for all $i \neq j$ (equivalently, (2.1) holds), or there exist nonnegative weights $\lambda_{i,j}$, not all zero, such that

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

Consider the graph with vertices $\{1, \dots, n\}$ and (directed) edges corresponding to those pairs (i, j) for which $\lambda_{i,j} > 0$. There cannot be a vertex of degree one in the graph since, in that case, $\sum_{i\neq j} \lambda_{i,j} f_{i,j}(\psi)$ could not be bounded from below. Hence, the graph contains at least a cycle, that is, there exist $i_0, i_1, \dots, i_m$ such that λ_{i_0,i_1}, $\lambda_{i_1,i_2}, \dots$, and λ_{i_m,i_0} all are strictly positive. Part *(i)* of the lemma then implies the existence of $\bar\psi_1, \dots, \bar\psi_n$ such that $f_{i,j}(\bar\psi) \leq 0$ for all $i \neq j$. But then

$$0 \leq \sum_{i\neq j} \lambda_{i,j} f_{i,j}(\bar\psi) \leq 0,$$

which implies that $f_{i,j}(\bar\psi) = 0$ for each pair i, j with $\lambda_{i,j} > 0$, so that

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

This in turn entails (observe that the sum $\bar\psi_i - \bar\psi_j$ along a cycle $i_0, i_1, \dots, i_m, i_0$ vanishes)

$$\langle x_{i_0}, y_{i_0} - y_{i_1}\rangle + \langle x_{i_1}, y_{i_1} - y_{i_2}\rangle + \cdots + \langle x_{i_m}, y_{i_m} - y_{i_0}\rangle = 0. \tag{2.7}$$

But (2.7) contradicts (2.6), which implies that if T is the unique cyclically monototone map from $\{x_1, \dots, x_n\}$ to $\{y_1, \dots, y_n\}$ then (2.1) holds. Conversely, if (2.1) holds, then, for every cycle $i_0, i_1, \dots, i_m, i_0$, we have

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

and T is the unique cyclically monototone map from $\{x_1, \dots, x_n\}$ to $\{y_1, \dots, y_n\}$. This completes the proof. □

Remark 2.1.1. While in Proposition 2.1 we have assumed that $y_1, \dots, y_n$ are n distinct points of $\mathbb{R}^d$, similar versions of this lemma can be proved if we relax this condition. For the applications in Section 3.1, it will be useful to consider the case when $y_1 = \cdots = y_{n_0}$ and y_1 and $y_1 \neq y_i$ for $i > n_0$. In this case, the proof above is easily adapted to show that the map $T(x_i) = y_i$, $i = 1, \dots, n$ is cyclically monotone if and only if there exist real numbers $\psi_1, \psi_{n_0+1}, \dots, \psi_n$ such that, setting $\psi_i = \psi_1$, $i = 2, \dots, n_0$,

$$\langle x_i, y_i\rangle - \psi_i = \max_{j=1,\dots,n}(\langle x_i, y_j\rangle - \psi_j), \quad i = 1, \dots, n.$$

Similarly, the map $T(x_i) = y_i$, $i = 1, \dots, n$ is the unique cyclically monototone map from $\{x_1, \dots, x_n\}$ to $\{y_1, y_{n_0+1} \dots, y_n\}$ mapping n_0 points in $\{x_1, \dots, x_n\}$ to y_1 if and only if there exist real numbers $\psi_1, \psi_{n_0+1}, \dots, \psi_n$ such that

$$\langle x_i, y_1\rangle - \psi_1 > \langle x_i, y_j\rangle - \psi_j, \quad i = 1, \dots, n_0,\ j = n_0 + 1, \dots, n,$$

$$\langle x_i, y_i\rangle - \psi_i > \langle x_i, y_j\rangle - \psi_j, \quad i = n_0 + 1, \dots, n,\ j = 1, n_0 + 1, \dots, n,\ j \neq i.$$

This can be proved with straightforward changes to the proof of Proposition 2.1. We omit details.

As a consequence of Proposition 2.1, 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 $\psi_1, \dots, \psi_n$ such that (2.1) holds. Consider the convex map

$$x \mapsto \varphi(x) := \max_{1\leq j\leq n}(\langle x, y_j\rangle - \psi_j). \tag{2.8}$$

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

$$\bar{T}(x) := \nabla\varphi(x), \quad x \in \bigcup_{i=1}^n C_i.$$

By construction, $x_i \in C_i$, hence $\bar{T}$ is an extension of T. Rockafellar's Theorem (Theorem 12.15 in [8]) implies that $\bar{T}$ is cyclically monotone. We could (in case $\bigcup_{i=1}^n C_i \subsetneq \mathbb{R}^d$) extend $\bar{T}$ from $\bigcup_{i=1}^n C_i$ to $\mathbb{R}^d$ while preserving cyclical monotonicity, but such extension of $\bar{T}$ 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

$$\varphi_\varepsilon(x) := \inf_{y\in\mathbb{R}^d}\Big[\varphi(y) + \frac{1}{2\varepsilon}\|y-x\|^2\Big], \quad x\in\mathbb{R}^d,\ \varepsilon>0 \tag{2.9}$$

of φ (as defined in (2.8)): see, e.g., [8]. The following theorem shows that, for sufficiently small $\varepsilon > 0$, $\nabla\varphi_\varepsilon$—the so-called Yosida regularization of $\nabla\varphi$ (see [9])—provides the desired continuous, cyclically monotone interpolation of $(x_1, y_1), \ldots, (x_n, y_n)$.

Summing up, we have the following theorem.

Theorem 2.2. *Let Assumption (A) hold, and consider φ as in (2.8), with $\psi_1, \ldots, \psi_n$ satisfying (2.1). Let φ_ε as in (2.9). Then, there exists $e > 0$ such that for every $\varepsilon \le e$ the map φ_ε is continuously differentiable and $T_\varepsilon := \nabla\varphi_\varepsilon$ is a continuous, cyclically monotone map such that*

$$T_\varepsilon(x_i) = y_i, \quad i = 1, \ldots, n$$

and $\|T_\varepsilon(x)\| \le \max_{i=1,\ldots,n}\|y_i\|$ for all $x \in \mathbb{R}^d$.

We note that the main conclusion of Theorem 2.2 remains true in the setup of Remark 2.1.1 and we still can guarantee that there exists a convex, continously differentiable φ such that $\nabla\varphi(x_i) = y_1$, $i = 1, \ldots, n_0$, $\nabla\varphi(x_i) = y_i$, $i = n_0 + 1, \ldots, n$ in that case. More generally, the following corollary, which heuristically can be interpreted as a discrete version of the fact that a smooth convex function has a positive semi-definite second-order differential, is an immediate consequence.

Corollary 2.3. *Any cyclically monotone subset $S = \{(x_i, y_i)|i = 1, \ldots, n\}$ of $\mathbb{R}^d \times \mathbb{R}^d$ with $x_i \ne x_j$ if $i \ne j$ is in the subdifferential (at x_i, $i = 1, \ldots, n$) of some (continuously) differentiable convex function Ψ.*

Proof of Theorem 2.2. The map φ_ε is convex and continuously differentiable since φ is convex (see, e.g., Theorem 2.26 in [8]). Hence $T_\varepsilon := \nabla\varphi_\varepsilon$ is a cyclically monotone, continuous map for every $\varepsilon > 0$. Setting

$$\tilde{\varepsilon}_0 = \min_{1\le i\le n}\Big((\langle x_i, y_i\rangle - \psi_i) - \max_{j\ne i}(\langle x_i, y_j\rangle - \psi_j)\Big),$$

let $\varepsilon_0 = \frac{1}{2}\tilde{\varepsilon}_0 \min(1, 1/\max_{1\le i\le n}\|y_i\|)$; note that , by (2.1), $\tilde{\varepsilon}_0$ is strictly positive, hence also ε_0. If x lies in the ε_0-ball $B(x_i, \varepsilon_0)$ centered at x_i, then, if $j \ne i$,

$$\begin{aligned}\langle x, y_i\rangle - \psi_i &= \langle x_i, y_i\rangle - \psi_i + \langle x - x_i, y_i\rangle > \langle x_i, y_j\rangle - \psi_j + \tilde{\varepsilon}_0 - \varepsilon_0\|y_i\| \\ &\ge \langle x_i, y_j\rangle - \psi_j + \frac{1}{2}\tilde{\varepsilon}_0 \ge \langle x, y_j\rangle - \psi_j.\end{aligned}$$

This shows that $B(x_i, \varepsilon_0) \subset C_i$ and $\varphi(x) = \langle x, y_i\rangle - \psi_i$ in $B(x_i, \varepsilon_0)$.

Assume now that $0 < \varepsilon \leq \frac{1}{2}\varepsilon_0 \min(1, 1/\max_{1\leq i\leq n} \|y_i\|)$, and take $x \in B(x_i, \varepsilon)$. The map $y \mapsto \langle y, y_i\rangle - \psi_i + \frac{1}{2\varepsilon}\|y-x\|^2$ attains its global minimum at $y = x - \varepsilon y_i \in B(x_i, \varepsilon_0)$. For any y, we have

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

This proves that

$$\varphi_\varepsilon(x) = \langle x, y_i\rangle - \psi_i - \frac{\varepsilon}{2}\|y_i\|^2, \quad x \in B(x_i, \varepsilon).$$

In particular, we conclude that $T_\varepsilon(x_i) = y_i$. For the last claim, we note that

$$T_\varepsilon(x) = \frac{1}{\varepsilon}(x - y_0),$$

where y_0 is the unique minimizer of $y \mapsto \varphi(y) + \frac{\|y-x\|^2}{2\varepsilon}$ (again by Theorem 2.26 in [8])). But y_0 is such a minimizer if and only if $0 \in \partial\varphi(y_0) + \frac{1}{\varepsilon}(y_0 - x)$, that is, if and only if $T_\varepsilon(x) \in \partial\varphi(y_0)$, where $\partial\varphi(y_0)$ denotes the subdifferential of φ at y_0. Now, for every $x \in \mathbb{R}^d$, $\partial\varphi(x)$ equals the closure of the convex hull of the set of limit points of sequences of the type $\nabla\varphi(x_n)$ with $x_n \to x$ (this is Theorem 25.6 in [7]). The map φ is differentiable in the regions C_i, with gradient y_i. Hence, for every x, $T_\varepsilon(x)$ belongs to the convex hull of $\{y_1, \ldots, y_n\}$. This completes the proof, □

Remark 2.3.1. It is important to note that, in spite of what intuition may suggest, and except in the one-dimensional case ($d = 1$), linear interpolation does not work in this problem. Assume that $n \geq d+1$ and that $\{x_1, \ldots, x_n\}$ are in general position. Denoting by $\mathcal{C}$ the convex hull of $\{x_1, \ldots, x_n\}$, there exists a partition of $\mathcal{C}$ into d-dimensional simplices determined by points in $\{x_1, \ldots, x_n\}$: every point in $\mathcal{C}$ thus can be written in a unique way as a linear convex combination of the points determining the simplex it belongs to (with obvious modification for boundary points). Therefore, if $x \in \mathcal{C}$, there exist uniquely defined coefficients $\lambda_i^x \in [0, 1]$, $i = 1, \ldots, n$, with $\sum_i \lambda_i^x = 1$ and $\#\{i|\lambda_i^x \neq 0\} \leq d + 1$, such that $x = \sum_{i=1}^k \lambda_i^x x_i$. A "natural" linear interpolation of T on $\mathcal{C}$ would be

$$x \mapsto \sum_{i=1}^k \lambda_i^x y_i,\ x \in \mathcal{C}.$$

For $d = 1$, this map is trivially monotone increasing, hence cyclically monotone. Starting with $d = 2$, however, this is no longer true, as the following counterexample shows. Let (for $d = 2$)

$$x_1 = (0, 0),\ x_2 = (0, 1),\ x_3 = (1, 1) \quad \text{and} \quad y_1 = (-.5, -.01),\ y_2 = (.5, .01),\ y_3 = (1, 0).$$

It is easily checked that the map $x_i \mapsto y_i$, $i = 1, 2, 3$ is the only cyclically monotone one pairing those points. Now, let us consider the points

$$x_0 = .8x_1 + .1x_2 + .1x_3 \quad \text{and} \quad y_0 = .8y_1 + .1y_2 + .1y_3.$$

The computation of all possible 24 pairings shows that the only cyclically monotone mapping between the sets $\{x_0, \dots, x_3\}$ and $\{y_0, \dots, y_3\}$ is

$$x_i \mapsto \begin{cases} y_i & \text{if } i = 1, 3 \\ y_0 & \text{if } i = 2 \\ y_2 & \text{if } i = 0 \end{cases}$$

where obviously x_0 is not paired with y_0 (nor x_2 with y_2).

Remark 2.3.2. It is worth remarking at this point that the interpolating function T_ε given by the proof of Theorem 2.2 is not only continuous but, in fact, Lipschitz with constant $1/\varepsilon$ (see, e.g., Exercise 12.23 in [8]). 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 $\|y_i\| \le 1$, $i = 1, \dots, n$ (note that this does not imply any loss of generality; we could adequately normalize the data to get this satisfied, then backtransform the interpolating function). Set

$$\varepsilon_0 := \frac{1}{2} \min_{1 \le i \le n} \Big((\langle x_i, y_i \rangle - \psi_i) - \max_{j \ne i} (\langle x_i, y_j \rangle - \psi_j) \Big). \tag{2.10}$$

Then, arguing as in the proof of Theorem 2.2, we see that $B(x_i, \varepsilon_0) \subset C_i$. Let $\varepsilon > 0$ and $\delta > 0$ be such that $\varepsilon + \delta < \varepsilon_0$. Then, for $x \in B(x_i, \delta)$, we have $x - \varepsilon y_i \in B(x_i, \varepsilon_0)$, and we can mimic the argument in the proof above to conclude that, for $x \in B(x_i, \delta)$, we have $\varphi_\varepsilon(x) = \langle x, y_i \rangle - \psi_i - \frac{\varepsilon}{2}\|y_i\|^2$, and, consequently, $T_\varepsilon(x_i) = y_i$ for every $\varepsilon < \varepsilon_0$ with ε_0 given by (2.10). By continuity of the Yosida regularization (see Theorem 2.26 in [8]), we conclude that $T_{\varepsilon_0}(x_i) = y_i$, $i = 1, \dots, n$. We summarize our findings in the following result.

Corollary 2.4. *Let Assumption (A) hold. Assume further that $\|y_i\| \le 1$, $i = 1, \dots, n$. Let $\varphi(x) := \max_{1 \le j \le n}(\langle x, y_j \rangle - \psi_j)$ with $\psi_1, \dots, \psi_n$ as in (2.1), φ_ε as in (2.8), and ε_0 as in (2.10). Then $T_{\varepsilon_0} := \nabla \varphi_{\varepsilon_0}$ is a Lipschitz continuous, cyclically monotone map, with Lipschitz constant $1/\varepsilon_0$, such that $T_{\varepsilon_0}(x_i) = y_i$, $i = 1, \dots, n$ and $\|T_{\varepsilon_0}(x)\| \le 1$ for every $x \in \mathbb{R}^d$.*

To conclude, let us turn to the choice of the weights ψ_i that satisfy condition (2.1), as required by our construction. In view of Corollary (2.4) and the discussion in Remark 2.3.2, choosing the weights that maximize ε_0 in (2.10) results in smoother interpolations. This maximization problem can be recast as the linear program

$$\begin{aligned} &\max_{\psi,\varepsilon} \quad \varepsilon/2 \\ &\text{s.t.} \quad \langle x_i, y_i - y_j\rangle \geq \psi_i - \psi_j + \varepsilon, \quad i,j\in\{1,\dots,n\},\, i\neq j, \end{aligned} \tag{2.11}$$

the dual of which is

$$\begin{aligned} \min_{z_{i,j}, i\neq j} \ & \frac{1}{2} \sum_{i,j=1,\dots,n;\, i\neq j} z_{i,j}\langle x_i, y_i - y_j\rangle \\ \text{s.t.} \ & \sum_{j=1,\dots,n;\, j\neq i} (z_{i,j} - z_{j,i}) = 0, \quad i = 1,\dots,n \\ & \sum_{i,j=1,\dots,n;\, i\neq j} z_{i,j} = 1, \quad z_{i,j}\geq 0. \end{aligned} \tag{2.12}$$

Note that (2.12) is a variation of a constrained transportation problem in which the constraint is that the first and second marginals of the joint distribution with probability weights $z_{i,j}$ are the same. In the particular case $n = 2$, we see that the optimum in (2.12) (hence in (2.11)) is $\varepsilon_0 = \frac{1}{4}\langle x_1 - x_2, y_1 - y_2\rangle > 0$. The optimal weights can be chosen as $\psi_i = \frac{1}{2}\langle (x_1 + x_2), y_i\rangle$, $i = 1, 2$. 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 then yields

$$\begin{aligned} T_\varepsilon(x) &= y_1 \quad \text{if} \quad \frac{1}{\varepsilon}\Big(x - \frac{x_1+x_2}{2}\Big) \leq y_1, \\ T_\varepsilon(x) &= y_2 \quad \text{if} \quad \frac{1}{\varepsilon}\Big(x - \frac{x_1+x_2}{2}\Big) \geq y_2, \end{aligned}$$

while

$$T_\varepsilon(x) = \frac{1}{\varepsilon}\Big(x - \frac{x_1+x_2}{2}\Big) \quad \text{if} \quad y_1 \leq \frac{1}{\varepsilon}\Big(x - \frac{x_1+x_2}{2}\Big) \leq y_2.$$

We see that T_ε is an extension of the map $x_i \mapsto y_i$, $i = 1, 2$ 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 $\varepsilon \leq \frac{x_2 - x_1}{y_2 - y_1}$ or, equivalently, $\frac{1}{\varepsilon} \geq \frac{y_2-y_1}{x_2-x_1}$ (note that $\frac{y_2-y_1}{x_2-x_1}$ is the minimal Lipschitz constant of any Lipschitz extension of the map $x_i \mapsto y_i$). For $y_1 = -1$, $y_2 = 1$, we get $\varepsilon_0 = \frac{x_2-x_1}{2} = \frac{y_2-y_1}{x_2-x_1}$ and we see that the interpolating function T_{ε_0} is the Lipschitz extension of $x_i \mapsto y_i$ with minimal Lipschitz constant.

3. Application to empirical center-outward distribution functions

3.1. Continuous center-outward empirical distribution functions

In this section, we deal with the smooth extension of the empirical center-outward distribution functions introduced in [5]. We briefly recall the setup.

Factorizing the sample size into $n = n_R n_S + n_0$ with $n_R, n_S, n_0 \in \mathbb{N}$ such that $0 \leq n_0 < \min(n_R, n_S)$, assume that $n_R \to \infty$ and $n_S \to \infty$ as $n \to \infty$. Consider a regular n_S-tuple of unit vectors (meaning that the sequence of uniform discrete measures on $\{u_1, \dots, u_{n_S}\}$ converges weakly as $n \to \infty$ to the uniform measure over the unit sphere in $\mathbb{R}^d$). Finally, consider the grid of $n_R n_S$ points that can be obtained intersecting the rays starting at the origin with directions in $\{u_1, \dots, u_{n_S}\}$ with the hyperspheres centered at the origin with radii $\frac{1}{n_R+1}, \dots, \frac{n_R}{n_R+1}$ plus n_0 copies of the origin. Then, given a (random) sample $Z_1^{(n)}, \dots, Z_n^{(n)}$, the empirical center-outward distribution function $F_\pm^{(n)}$ is defined as a cyclically monotone mapping from $Z_1^{(n)}, \dots, Z_n^{(n)}$ to the above constructed grid. If $Z_1^{(n)}, \dots, Z_n^{(n)}$ are i.i.d. realizations from $P \in \mathcal{P}_d$ (as in subsection 1.2, we write $\mathcal{P}_d$ for the family of probability measures with a nonvanishing density over $\mathbb{R}^d$), then with probability one this empirical center-outward distribution function is uniquely defined. Now, using Theorem 2.2 (and subsequent comments for the case $n_0 > 0$), we can extend $F_\pm^{(n)}$ to a Lipschitz continuous gradient of convex function over $\mathbb{R}^d$, which we denote by $\bar{F}_\pm^{(n)}$.

Our main result in this section is an extension of the Glivenko-Cantelli result (1.4) to those empirical center-outward distribution functions $\bar{F}_\pm^{(n)}$ defined over $\mathbb{R}^d$. We recall from Subsection 1.2 that, for $P \in \mathcal{P}_d$, $F_\pm$ denotes the center-outward distribution function of P, that is, the unique continuous gradient of a convex function pushing P forward to U_d, (the product measure of a uniform distribution over the unit sphere $\mathcal{S}_{d-1}$ and a uniform distribution over the unit interval $[0, 1]$); it follows from [4] that $F_\pm$ is a homeomorphism from $\mathbb{R}^d \backslash F_\pm^{-1}(\{0\})$ to $\mathbb{S}_d \backslash \{0\}$, where $F_\pm^{-1}(\{0\})$ has Lebesgue measure zero.

Theorem 3.1. *(Glivenko-Cantelli for center-outward distribution functions) With the above notation,*

$$\sup_{x \in \mathbb{R}^d} \|\bar{F}_\pm^{(n)}(x) - F_\pm(x)\| \to 0$$

with probability one.

Proof. Denote by $\mathrm{U}_d^{(n)}$ the discrete probability measure that gives mass $\frac{n_0}{n}$ to the origin and $\frac{1}{n}$ to the remaining points in the regular grid used for the definition of $F_\pm^{(n)}$, and note that $\mathrm{U}_d^{(n)}$ converges weakly to U_d. Also write $P^{(n)}$ for the empirical measure on $Z_1^{(n)}, \dots, Z_n^{(n)}$. Over a probability one set Ω_0, say, $P^{(n)}$ converges weakly to P. In the remainder of this proof, we assume, without loss of generality, that Ω_0 is the whole space. We note that $\bar{F}_\pm^{(n)} = \nabla\varphi_n$, $F_\pm = \nabla\varphi$ for some φ_n, φ convex and continuously differentiable over $\mathbb{R}^d$, and recall that by construction $\nabla\varphi_n$ maps $P^{(n)}$ to $\mathbf{U}_d^{(n)}$. Now, the convex potential φ is uniquely defined up to an additive constant. By Theorem 2.8 in [1], there exist constants a_n such that, if $\tilde{\varphi}_n = \varphi_n - a_n$, then $\tilde{\varphi}_n(x) \to \varphi(x)$ for every $x \in \mathbb{R}^d$ (we note that, while the statement of the cited result assumes convergence in transportation cost metric rather than weak convergence, the proof depends only on the fact that, in that case, $\pi_n = (Id \times \nabla\varphi_n)\sharp P^{(n)}$ converges weakly to $\pi = (Id \times \nabla\varphi)\sharp P$, which holds in the setup considered here, see, Lemma 8.5 in [5]). But then (see Theorem 25.7 in [7]), $\bar{F}_{\pm}^{(n)}(x) = \nabla\varphi_n(x) \to \nabla\varphi(x) = F_{\pm}(x)$ (uniformly over compact sets). It only remains to show that uniform convergence holds over $\mathbb{R}^d$.

For this, it suffices to show that, for every $w \in \mathbb{R}^d$,

$$\sup_{x\in\mathbb{R}^d} \left|\langle\big(\bar{F}_{\pm}^{(n)}(x) - F_{\pm}(x)\big), w\rangle\right| \to 0. \tag{3.1}$$

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

$$\left|\langle\big(\nabla\varphi_n(x_n) - \nabla\varphi(x_n)\big), w\rangle\right| > \varepsilon \tag{3.2}$$

for all n. The sequence x_n must be unbounded (otherwise (3.2) cannot hold). Hence, using compactness of the unit sphere and taking subsequences if necessary, we can assume that $x_n = \lambda_n u_n$ with $0 < \lambda_n \to \infty$, $\|u_n\| = 1$ and $u_n \to u$ for some u with $\|u\| = 1$. Again by compactness, we can assume that $\nabla\varphi(x_n) \to y$ and $\nabla\varphi_n(x_n) \to z$. By Lemma 3.2 below, we have that $y = u$. On the other hand, by monotonicity, for every $x \in \mathbb{R}^d$,

$$\langle\nabla\varphi_n(x_n) - \nabla\varphi_n(x), x_n - x\rangle \geq 0.$$

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

$$\langle\nabla\varphi_n(x_n) - \nabla\varphi_n(\tau u_n), u_n\rangle \geq 0.$$

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

$$\langle z - \nabla\varphi(\tau u), u\rangle \geq 0.$$

Now, we can take $\tau_n \to \infty$ with $\nabla\varphi(\tau_n u)$ converging to some limit. By Lemma 3.2, the limit must be u and, from the last inequality, we see that $\langle z - u, u\rangle \geq 0$, that is, $\langle z, u\rangle \geq \|u\|^2 = 1$. But this implies that $z = u = y$ and contradicts (3.2), which completes the proof. □

Lemma 3.2. *Assume $\varphi : \mathbb{R}^d \to \mathbb{R}$ is a differentiable convex function such that $\nabla\varphi$ is a homeomorphism from $\mathbb{R}^d$ to the open unit ball. If we take $x_n = \lambda_n u_n$ with $0 < \lambda_n \to \infty$, $\|u_n\| = 1$ and $u_n \to u$, then, $\nabla\varphi(x_n) \to u$.*

Proof. By monotonicity we have that

$$\langle x_n - x, \nabla\varphi(x_n) - \nabla\varphi(x)\rangle \geq 0$$

for every $x \in \mathbb{R}^d$ or, equivalently,

$$\langle x_n - (\nabla\varphi)^{-1}(w)), \nabla\varphi(x_n) - w\rangle \geq 0$$

for every w with $\|w\| < 1$. But this means that

$$\langle u_n - \tfrac{1}{\lambda_n}(\nabla\varphi)^{-1}(w), \nabla\varphi(x_n) - w\rangle \geq 0$$

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

3.2. A step-function version of center-outward empirical distribution functions

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

With the same notation as in the previous section, let $\bar{F}^{(n)}_{\pm}$ be some smooth interpolation of $F^{(n)}_{\pm}$. For any $r \in [0, 1]$ and u on the unit sphere $\mathcal{S}_{d-1}$, define

$$\lfloor ru \rfloor_{n_R} := \frac{\lfloor (n_R + 1)r \rfloor}{n_R + 1} u :$$

$ru \mapsto \lfloor ru \rfloor_{n_R}$ maps a outward-open, inward-closed spherical annulus comprised in between two hyperspheres of the grid onto its inner boundary sphere while preserving directions. Then, a "multivariate step function" version of the empirical center-outward distribution function $F^{(n)}_{\pm}$, continuous from outward, can be defined as

$$\bar{F}^{(n)*}_{\pm} := \lfloor \bar{F}^{(n)}_{\pm} \rfloor_{n_R}. \tag{3.3}$$

Instead of *steps*, those functions yield *plateaux* or *hyperplateaux*, the boundaries (equivalently, the discontinuity points) of which are the continuous *quantile contours* or *hypersurfaces* characterized by $\bar{F}^{(n)}_{\pm}$. Those "quantile contours" present an obvious statistical interest.

In contrast with the univariate case, this "step function version" (3.2) of the empirical center-outward distribution function $F^{(n)}_{\pm}$, however, is not unique, and depends on the smooth interpolation $\bar{F}^{(n)}_{\pm}$ adopted. However, all its versions enjoy cyclical monotonicity and obviously satisfy the sup form of Glivenko-Cantelli: with probability one,

$$\sup_{x \in \mathbb{R}^d} \|\bar{F}^{(n)*}_{\pm}(x) - F_{\pm}(x)\| \to 0.$$

4. Some numerical results

In this section, we provide some two-dimensional numerical illustrations of the results we established in this paper. The codes we used were written in R, and can handle sample sizes as high as $n = 2000$ (with $n_R = 50$ and $n_S = 40$) on a computer with 16Gb RAM. The algorithm consists of three main steps:

(step 1) Determine the optimal assignment between the sample points and the regular grid. For this, we used the cubic implementation of the Hungarian algorithm included in the 'clue' R package (for a detailed account of the Hungarian algorithm and the complexity of different implementations, see, e.g., Chapter 4 in Burkhard et al. (2009)). Faster algorithms are available, as for instance, the 'assignment' function in the 'adagio' package, but they apply only to integer-valued cost matrices.

(step 2) Compute the optimal value ε_0 of the regularization parameter. This is achieved by solving a linear program via the simplex method; it is the slowest part of our algorithm, and, most likely, much computational time can be saved here via sophisticated linear programming methods.

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

For higher sample sizes (n = 5000, for instance), our code is in trouble unless larger memory space (at least 64 Gb RAM) is available. The problem originates in step 2 from the fact that the simplex implementation that we use in R does not allow to take advantage of the sparsity of the constraint matrix in the linear program (2.11); that issue can be overcome by using a commercial solver like xpress, which we did for the large sample cases ($n = 5000$) below.

The sections below investigate the convergence of our method (Section 4.1), and its ability to recover the "shape" of a distribution (Sections 4.2 and 4.3). For obvious graphical reasons, we only consider bivariate observations.

4.1. Convergence

In this section, we illustrate the convergence (as formulated by the Glivenko-Cantelli result of Theorem 3.1), of empirical contours to their population counterparts as the sample size increases. The problem is that analytical expressions for the population contours are not easily derived, except for spherical distributions. We therefore investigate the case of i.i.d. observations with bivariate $\mathcal{N}(0, \mathrm{Id})$ distributions, and increasing samples sizes n =100, 200, 500, 1000, 2000, 5000.

Inspection of Figure 1 clearly shows the expected consistency. The empirical contours are nicely nested, as they are supposed to be. For sample sizes as big as $n = 500$, and despite the fact that the underlying distribution is light-tailed, the .90 empirical contour still exhibits significant "spikes" out and in the theoretical circular contour; those spikes, however, rapidly and uniformly disappear from $n = 1000$ on.

4.2. Gaussian mixtures

Gaussian mixtures generate a variety of alternative and possibly multimodal and non-convex empirical dataclouds. In Figure 2, we simulated $n = 2000$ observations from a symmetric mixture of two spherical Gaussians. Figure 2 clearly demonstrates the quantile contour nature of our interpolations, as opposed to

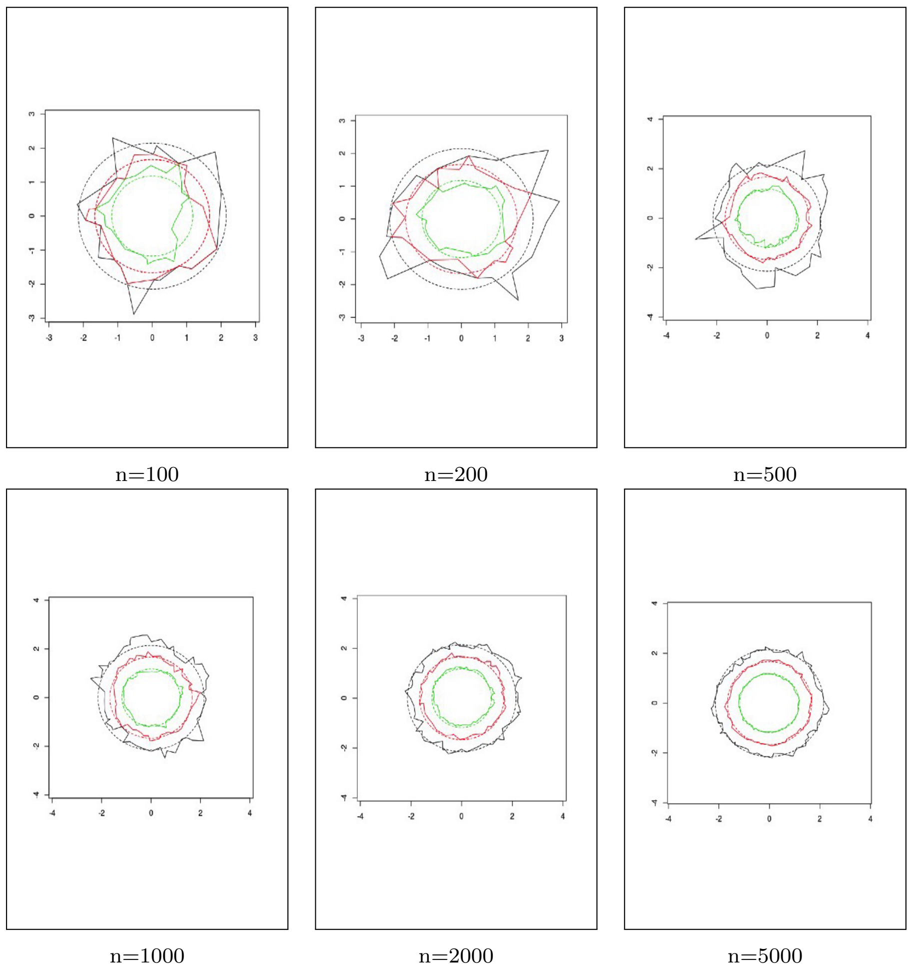
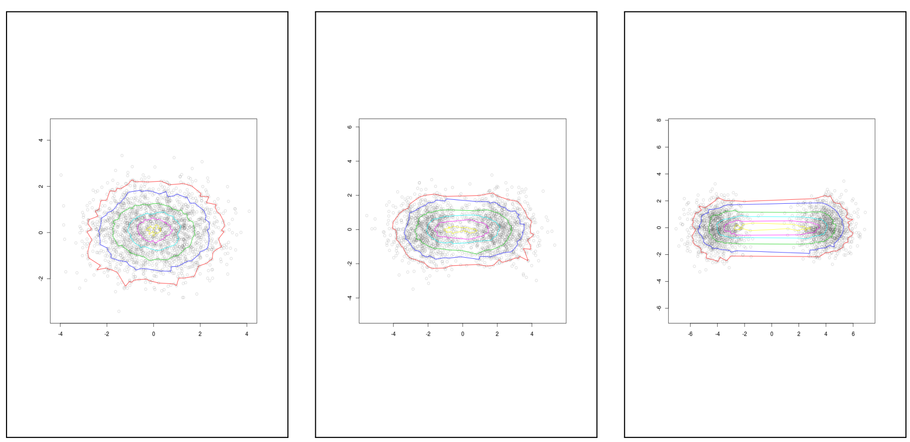


Fig 1. *Smoothed empirical center-outward quantile contours (probability contents .50 (green), .75 (red), .90 (black)) computed from* $n = 100$*, 200, 500, 1000, 2000, 5000 i.i.d. observations from a bivariate* $\mathcal{N}(0,\ Id)$ *distribution, along with their (spherical) theoretical counterparts.*

level contours. Level contours in the right-hand panel clearly would produce disconnected regions separating the two modes of the mixture. Here, the contours remain nested—a fundamental monotonicity property of quantiles. The low-probability region between the two component populations is characterized by a "flat profile" of empirical quantile contours: the whole region in between the two modes is "quite central", and one can move, for instance, from one mode to the other without crossing the .5 contour.

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

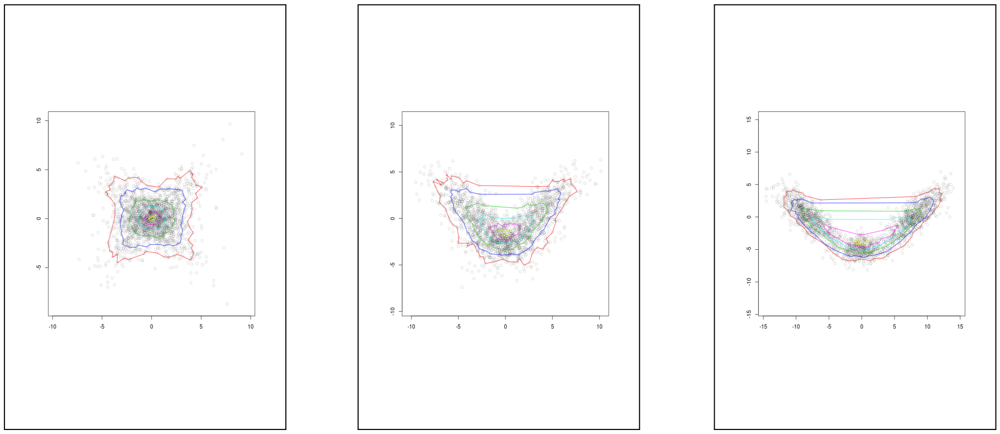


$\frac{1}{2}\mathcal{N}\left(\begin{pmatrix}-1\\0\end{pmatrix},\mathrm{Id}\right)+\frac{1}{2}\mathcal{N}\left(\begin{pmatrix}1\\0\end{pmatrix},\mathrm{Id}\right)$ $\frac{1}{2}\mathcal{N}\left(\begin{pmatrix}-2\\0\end{pmatrix},\mathrm{Id}\right)+\frac{1}{2}\mathcal{N}\left(\begin{pmatrix}2\\0\end{pmatrix},\mathrm{Id}\right)$ $\frac{1}{2}\mathcal{N}\left(\begin{pmatrix}-4\\0\end{pmatrix},\mathrm{Id}\right)+\frac{1}{2}\mathcal{N}\left(\begin{pmatrix}4\\0\end{pmatrix},\mathrm{Id}\right)$

FIG 2. *Smoothed empirical center-outward quantile contours (probability contents .02 (yellow), .20 (cyan), .25 (light blue) .50 (green), .75 (dark blue), .90 (red)) computed from* $n = 2000$ *i.i.d. observations from mixtures of two bivariate Gaussian distributions.*

4.3. Bounded supports

Although the results have been derived in the general context of nonvanishing densities, they also hold under absolutely continuous compactly supported



$$\tfrac{3}{8}\mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}5&-4\\-4&5\end{pmatrix}\right)+\tfrac{3}{8}\mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}5&4\\4&5\end{pmatrix}\right)+\tfrac{1}{4}\mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}4&0\\0&1\end{pmatrix}\right)$$

$$\tfrac{3}{8}\mathcal{N}\left(\begin{pmatrix}-3\\0\end{pmatrix},\begin{pmatrix}5&-4\\-4&5\end{pmatrix}\right)+\tfrac{3}{8}\mathcal{N}\left(\begin{pmatrix}3\\0\end{pmatrix},\begin{pmatrix}5&4\\4&5\end{pmatrix}\right)+\tfrac{1}{4}\mathcal{N}\left(\begin{pmatrix}0\\-\frac{5}{2}\end{pmatrix},\begin{pmatrix}4&0\\0&1\end{pmatrix}\right)$$

$$\tfrac{3}{8}\mathcal{N}\left(\begin{pmatrix}-8\\0\end{pmatrix},\begin{pmatrix}5&-4\\-4&5\end{pmatrix}\right)+\tfrac{3}{8}\mathcal{N}\left(\begin{pmatrix}8\\0\end{pmatrix},\begin{pmatrix}5&4\\4&5\end{pmatrix}\right)+\tfrac{1}{4}\mathcal{N}\left(\begin{pmatrix}0\\-5\end{pmatrix},\begin{pmatrix}4&0\\0&1\end{pmatrix}\right)$$

FIG 3. *Smoothed empirical center-outward quantile contours (probability contents .02 (yellow), .20 (cyan), .25 (light blue) .50 (green), .75 (dark blue), .90 (red)) computed from* $n = 2000$ *i.i.d. observations from mixtures of three bivariate Gaussian distributions.*

distributions—the assumption made in [3]. Figure 3 provides simulations for uniforms with triangular and squared supports (sample size $n = 2000$, with $n_R = 50$, $n_S = 40$), and shows how the contours evolve from nested circles in the center to nested triangles and squares in the vicinity of support boundaries.

References

[1] del Barrio, E. and Loubes, J.M. (2018). Central Limit Theorem for empirical transportation cost in general dimension, *Annals of Probability*, to appear.

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

[3] Chernozhukov, V., Galichon, A., Hallin, M. and Henry, M. (2017). Monge-Kantorovich depth, quantiles, ranks, and signs, *Annals of Statistics* **45**, 223–256.

[4] Figalli, A. (2017). On the continuity of center-outward distribution and quantile functions, arXiv:1805.04946.

[5] Hallin, M. (2018). Distribution and quantile functions, ranks, and signs in $\mathbb{R}^d$: a measure transportation approach, ECARES Working Paper.

[6] Rockafellar, R.T. (1966). Characterization of the subdifferential of convex functions, *Pacific Journal of Mathematics* **17**, 497–510.

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

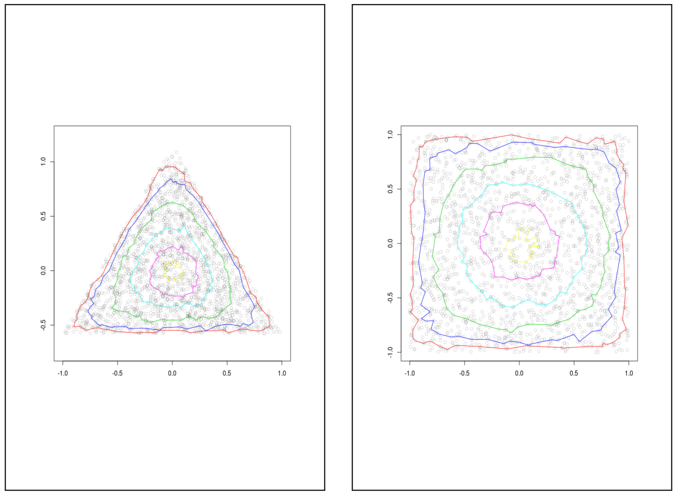


Fig 4. *Smoothed empirical center-outward quantile contours (probability contents .50 (green), .75 (red), .90 (black)) computed from* $n = 2000$ *i.i.d. observations from Lebesgue-uniform distributions over the triangle and the square, respectively.*

[8] Rockafellar, R.T. and Wets, R.J.B. (1998). *Variational Analysis*, Springer.
[9] Yosida, K. (1964). *Functional Analysis*, Springer Verlag.