Working Document: Optimal transportation approaches to learning multivariate extreme quantiles

Old version not updated since September 2024
The project has been split in two different parts since then

Benoît Reber

Supervisors
Professor Anne SABOURIN, (MAP5, Université Paris Cité, France)
Professor Johan SEGERS (ISBA, Université Catholique de Louvain, Belgium)

Date : September 2, 2024

Notation

- We write P for a fixed probability measure throughout the text, and E for the associated expectation;
- $|\cdot|$ denotes the Euclidean distance in $\mathbb{R}^d$ or $\mathbb{R}^d \times \mathbb{R}^d$;
- $\psi_{\#}P$ denotes the push-forward of the measure P by the fonction ψ;
- For $\mu \in \mathrm{M}_0(\mathbb{R}^d)$, $\bar{\mu}$ denotes the equivalence class of μ and $\tilde{\mu}$ denotes the element in $\bar{\mu}$ putting no mass on $\{0\}$;
- For a Borel measure m on $\mathbb{R}^d$, we write $\operatorname{res} m$ for the restriction of m to $\mathcal{B}(\mathbb{R}^d \setminus \{0\})$ the usual Borel σ-algebra on $\mathbb{R}^d \setminus \{0\}$;
- The symbols P, Q and μ, ν are reserved for probability measures in $\mathcal{P}(\mathbb{R}^d)$ and measures in $\mathrm{M}_0(\mathbb{R}^d)$ respectively;
- $\Pi_{(cm)}(P, Q)$ and $\Gamma_{0,(cm)}(\mu, \nu)$ denote the families of (cyclically monotone) couplings between P, Q and of (cyclically monotone)zero-couplings between μ, ν respectively;
- In general we use the symbols $P(Q)$ for the first(second) marginals of a coupling π, and the symbols $\mu(\nu)$ for the first(second) zero-marginals of a zero-coupling γ;
- When they exist we write ψ and $\bar{\psi}$ for closed convex functions satisfying $\operatorname{spt} \pi \subset \partial\psi$ and $\operatorname{spt} \gamma \subset \partial\bar{\psi}$ respectively.
- The symbols π and γ are reserved for elements in $\Pi_{(cm)}(P, Q)$ and $\Gamma_{0,(cm)}(\mu, \nu)$ respectively;
- $\mathrm{B}_{x,r} = \{y \in \mathrm{S} : d(x, y) < r\}$ denotes the open ball centered at x with radius r, and $\overline{\mathrm{B}}_{x,r}$ denotes its closure $\operatorname{cl} \mathrm{B}_{x,r}$;
- $\mathbb{S}^{d-1}$ denotes the unit sphere in $\mathbb{R}^d$;
- $\mathbb{R}_+$ and $\mathbb{R}_-$ refer to $[0, +\infty)$ and $(-\infty, 0]$ respectively.

Contents

1 Preliminaries

This first section is devoted to introducing the concepts used throughout the text. It contains no new results, hence its content can be skipped to a large extent by the reader familiar with these concepts.

To begin with, in Subsection 1.1 we introduce the space M_0 of Borel measures on $\mathrm{S} \setminus \mathrm{C}$ that are finite on sets bounded away from a closed subset C of the Polish space S. This family of measures was first studied in Hult and Lindskog (2006) and further developed by Lindskog et al. (2014). We summarize some useful results about it including versions of the Portmanteau and Prohorov theorems for this space. Seeing it as a particular case of vague convergence as defined in Kallenberg (2021), we get simple criteria for the weak convergence of random measures in the M_0-topology.

Relying on the M_0 space, in Subsection 1.2 we introduce regular variation of measures, which is a cornerstone of extreme-value theory, and which motivated the introduction of the M_0 space in the very first article Hult and Lindskog (2006). We also provide the classical polar decomposition argument which yields the form of a limit measure of regular variation which is developed in Lindskog et al. (2014) in the M_0 framework. We conclude this subsection with an elementary decomposition which will be used in Section 4 to define quantile regions of probability content near 1.

In Subsection 1.3, we introduce multivalued maps which will be used in the subsequent sections to deal with the subdifferential $\partial\psi$ of some closed convex function ψ that will arise in optimal transport. Identifying them with their graph, we define the Painlevé–Kuratowski convergence following Rockafellar and Wets (1998). To deal with optimal transport between random measures, hence random $\partial\psi$, we endow the family of closed subsets of $\mathbb{E} \subset \mathbb{R}^d$ with the Fell topology following Molchanov (2017). It yields a compact Polish space for which results from usual probability theory can be used.

In Subsection 1.5, we present the Monge–Kantorovitch problem between probability measures P, Q when the cost is the squared Euclidean distance. Relying on Knott and Smith (1984) and McCann (1995), we recall that if P vanishes on sets of Hausdorff dimension at most $d-1$, there is a unique optimal transport plan and its support is cyclically monotone, hence included in the subdifferential of some closed convex function ψ. As a consequence we can see coupling with cyclically monotone support as a generalization of optimal transport. However, we show that the aforementioned results cannot be directly extended to measures in M_0 with infinite total mass that arise as limits in the definition of regular variation.

Finally, in Subsection 1.6 we introduce briefly the Monge–Kantorovitch depth proposed in Chernozhukov et al. (2017) and relying on optimal transport from a reference distribution P to the distribution of interest Q and on the half-space depth. Then we present the center-outward quantile contours and regions built on the previous concept in Hallin et al. (2021) with the choice of the uniform spherical distribution U as reference distribution. To emphasize that the latter concept is a good candidate for quantile regions in $\mathbb{R}^d$, we recall that it matches the usual one in well known cases and benefits from good statistical properties demonstrated in Hallin et al. (2021).

1.1 M_0, a suitable family of measures for extremes

Let (S, d) be a complete separable metric space. We write $\mathcal{S}$ for the Borel σ-algebra on S, $\mathrm{B}_{x,r} = \{y \in \mathrm{S} : d(x, y) < r\}$ for the open ball centered at x with radius r, and $\overline{\mathrm{B}}_{x,r}$ for its closure $\operatorname{cl} \mathrm{B}_{x,r}$. Let $\mathrm{S}_0 = \mathrm{S} \setminus \mathrm{C}$ where C is a closed subset of S. One often choose $\mathrm{C} = \{s_0\}$ for some fixed element s_0 called the origin. When not specified otherwise, we choose $\mathrm{C} = \{0\}$. Its induced Borel σ-algebra will be denoted by $\mathcal{S}_0 = \mathcal{S} \cap (\mathrm{S} \setminus \mathrm{C})$.

Following Hult and Lindskog (2006) and Lindskog et al. (2014), we define $\mathrm{C}^r = \{x \in \mathrm{S} : d(x, \mathrm{C}) < r\}$ and let $\mathrm{M}_0(\mathrm{S})$ denote the set of nonnegative Borel measures on S_0 which are finite on sets bounded away from C, i.e., whose restriction to $\mathrm{S} \setminus \mathrm{C}^r$ is finite for each $r > 0$, and let $\mathrm{C}^+_{b,0}(\mathrm{S})$ denote the set of nonnegative bounded continuous functions whose support is bounded away from C, i.e., there exists $r > 0$ such that $f(\mathrm{C}^r) = \{0\}$.

Remark 1.1. *For the sake of clarity, when* $\mathrm{C} = \{s_0\}$ *we will directly write* $\mathrm{B}_{s_0,r}$ *instead of* C^r*.*

We now introduce a notion of convergence in $\mathrm{M}_0(\mathrm{S})$.

Definition 1.2. *Let* $\mu_n, \mu \in \mathrm{M}_0(\mathrm{S}), n \geq 1$*. We say that* μ_n *converges to* μ *in* $\mathrm{M}_0(\mathrm{S})$ *if* $\mu_n(f) \to \mu(f)$ *for each* $f \in \mathrm{C}^+_{b,0}(\mathrm{S})$*. When no confusion is possible we sometimes write that* μ_n M_0*-converges to* μ *and use the symbol* $\mu_n \xrightarrow{0} \mu$ *to denote this convergence.*

Proposition 1.3 (Hult and Lindskog (2006))**.** *The aforementioned convergence induces a topology making* $\mathrm{M}_0(\mathrm{S})$ *a Polish space with associated distance defined for all* μ, ν *in* $\mathrm{M}_0(\mathrm{S})$ *as*

$$d_0(\mu, \nu) = \int_0^\infty e^{-r} p_r(\mu^{(r)}, \nu^{(r)})[1 + p_r(\mu^{(r)}, \nu^{(r)})]^{-1} \mathrm{d}r$$

where p_r *denotes the Prohorov metric and* $m^{(r)}$ *denotes the restriction of* m *to* $\mathrm{S} \setminus \mathrm{C}^r$ *for every* $m \in \mathrm{M}_0(\mathrm{S})$*.*

The topology induced by d_0 is the weakest topology that makes the evaluation maps $\mathrm{M}_0(\mathrm{S}) \to [0, \infty[: \mu \mapsto \mu(f)$ *continuous, where* $f \in \mathrm{C}^+_{b,0}(\mathrm{S})$*. A basis of this topology is formed by sets of the form*

$$\{\mu \in \mathrm{M}_0(\mathrm{S}) : \mu(f_i) \in]a_i, b_i[\,,\, i = 1, \dots, K\}$$

for every $K \geq 1,\ a_i < b_i, i \geq 1,\ f_i \in \mathrm{C}^+_{b,0}(\mathrm{S}), i \geq 1$.

In the following sections, it will be sometimes useful to consider measures in $\mathrm{M}_0(\mathrm{S})$ as Borel measures on S as a whole. We introduce here another family of measures that naturally leads to an equivalent definition of $\mathrm{M}_0(\mathrm{S})$ that we find useful when it comes to understand zero-couplings in Definition 2.4 but is of no other interest.

Definition 1.4. *Let* $\mathrm{M}^+_0(\mathrm{S})$ *denote the family of nonnegative Borel measures on S as a whole which are finite on sets bounded away from zero. We define an equivalence relation* $\sim$ *as follows:* $\forall \mu, \nu \in \mathrm{M}^+_0(S)$, $\mu \sim \nu$ *iff* $\mu|_{\mathrm{S}_0} = \nu|_{\mathrm{S}_0}$. *Let* $\tilde{\mathrm{M}}_0$ *denote the quotient set* $\mathrm{M}^+_0(S)/\sim$. *We equip this space with the metric* d_0*, making* $\tilde{\mathrm{M}}_0$ *and* M_0 *have the same topology and Borel sets. It is possible since the definition of* d_0 *only involves the restriction of the measures to sets bounded away from zero and one can notice that for every* $\mu, \nu \in \mathrm{M}^+_0$, $\mu \sim \nu$ *iff* $d_0(\mu, \nu) = 0$. *It is then clear that* M_0 *and* $\tilde{\mathrm{M}}_0$ *are isometric. A measure* $\mu \in \mathrm{M}_0$ *can be seen as the class* $\overline{\mu}$ *of Borel measures* m *on S (with* C *included) such that* $m|_{\mathrm{S}_0} = \mu$ *and* $m(\mathrm{C}) \in \mathbb{R}_+ \cup \{\infty\}$.

For μ in $\mathrm{M}_0(\mathrm{S})$, we write $\overline{\mu}$ for the equivalence class containing μ, while $\tilde{\mu}$ will always denote the element in $\overline{\mu}$ such that $\tilde{\mu}(\mathrm{C}) = 0$. When no confusion is possible we will sometimes write $\overline{\mu}$ for an arbitrary element in the equivalence class. We also introduce the restriction operator

$$\mathrm{res} : \mu \in \mathrm{M}^+_0(\mathrm{S}) \longmapsto \mu|_{\mathrm{S}_0} \in \mathrm{M}_0(\mathrm{S}). \tag{1}$$

It is clear that for each $\mu' \in \bar{\mu}$, we have $\mathrm{res}(\mu') = \mu$ and that res is continuous (and isometric) when considered from $\tilde{\mathrm{M}}_0(\mathrm{S})$ to $\mathrm{M}_0(\mathrm{S})$.

We recall here the general definition of the support of a measure to emphasize our choice of a specific definition consistent with the equivalence classes we have just introduced. Let $\mathcal{N}_x(\mathcal{T})$ denote the set of open neighborhood of x when considering the topology $\mathcal{T}$.

Definition 1.5 (Support of a measure)**.** *Let* $(X, \mathcal{T})$ *be a topological space and let* $\mathcal{B}(X)$ *denote the Borel* σ*-algebra on* X*. The support of a measure* μ *on* $(X, \mathcal{B}(X))$ *is the closed set consisting of all points* x *in* X *such that every open neighborhood of* x *receives a positive mass*

$$\mathrm{spt}\,\mu = \{x \in X \mid \forall \mathrm{N} \in \mathcal{N}_x(\mathcal{T}) : \mu(\mathrm{N}) > 0\}$$

Equivalently, we have

$$\operatorname{spt}\mu = \bigcap \{F : F^c \in \mathcal{T},\ \textit{and } \mu(X \setminus F) = 0\} ..$$

From this usual definition, it is clear that the support $\operatorname{spt}\mu$ of the Borel measure μ on S_0 is a subset of S_0 while $\operatorname{spt}\tilde{\mu}$ and $\operatorname{spt}\mu'$ for some $\mu' \in \bar{\mu}$ are included in S. Moreover, without further assumptions there is no reason for $\operatorname{spt} m = \operatorname{spt} m'$ to hold when $m, m' \in \bar{\mu}$. To avoid ambiguities we choose a specific definition of spt in order to ensure that all the above sets are equals.

Definition 1.6 (Support of a measure in M_0)**.** *For $\mu \in \mathrm{M}_0(\mathrm{S})$ the expressions $\operatorname{spt}\mu$, $\operatorname{spt}\overline{\mu}$ and $\operatorname{spt}\tilde{\mu}$ all stand for the same set $\operatorname{spt}\tilde{\mu}$.*

Remark 1.7. *Assume $\mathrm{C} = \{s_0\}$. Let μ in M_0, then $\{s_0\}$ belongs to $\operatorname{spt}\mu$ iff for every $\mathrm{N} \in \mathcal{N}_{s_0}(\mathcal{T})$—$\mathcal{T}$ denotes the topology induced on S by the distance d—we have $\mu(\mathrm{N} \setminus \{s_0\}) > 0$. In particular, $\operatorname{spt}\mu$ cannot be equal to $\{s_0\}$.*

The space $\mathrm{M}_0(\mathrm{S})$ enjoys useful properties similar to Portmanteau theorem, Prohorov theorem, or continuous mapping theorem for probability measures which will be of constant use. We refer to Hult and Lindskog (2006) and Lindskog et al. (2014) for details and give here some results.

Theorem 1.8 (Portmanteau Theorem 2.4 in Hult and Lindskog (2006))**.** *Let $\mu, \mu_n \in \mathrm{M}_0(\mathrm{S}), n \geq 1$. The following statements are equivalent.*

(i) $\mu_n \xrightarrow{0} \mu$ *in* $\mathrm{M}_0(\mathrm{S})$ *as* $n \to \infty$,

(ii) $\mu_n(A) \to \mu(A)$ *as* $n \to \infty$ *for every* $A \in \mathcal{S}_0$ *satisfying* $\mu(\partial A) = 0$ *and* $\mathrm{cl}(A) \cap \mathrm{C} = \varnothing$ *(closure in* S *as a whole),*

(iii) $\mu(\mathrm{int}(A)) \leq \liminf \mu_n(\mathrm{int}(A) \leq \limsup \mu_n(\mathrm{cl}(A)) \leq \mu(\mathrm{cl}(A))$ *for every* $A \in \mathrm{S}_0$ *such that* $\mathrm{cl}(A) \cap \mathrm{C} = \varnothing$ *(closure in* S *as a whole).*

Theorem 1.9 (Prohorov Theorem 2.7 in Hult and Lindskog (2006))**.** *The set $M \subset \mathrm{M}_0(\mathrm{S})$ is relatively compact if and only if there exists a sequence $r_i > 0, i \geq 1$ satisfying $r_i \downarrow 0$ as $i \to \infty$ such that for each $i \geq 1$ we have*

$$\sup_{\mu \in M} \mu(\mathrm{S} \setminus \mathrm{C}^{r_i}) < \infty$$

and for each $\eta > 0$ there exists a compact set $K_i \subset \mathrm{S} \setminus \mathrm{C}^{r_i}$ such that

$$\sup_{\mu \in M} \mu(\mathrm{S} \setminus (\mathrm{C}^{r_i} \cup K_i)) < \eta.$$

This result gives a simple way to prove the M_0-convergence of μ_n to μ by showing (1) that $\mu_n, n \geq 1$ is relatively compact and (2) that all accumulation points are equals to μ.

To deal with empirical measures, one needs to consider random measures. Let $\stackrel{0d}{\longrightarrow}$ denote the convergence in distribution in the M_0-topology—i.e., weak convergence of probability measures on the metric space $\mathrm{M}_0(S)$. To benefit from general results in probability theory for random measures, we check that the topology we have endowed $\mathrm{M}_0(S)$ with is a vague topology in the sense of Kallenberg (2021). We introduce the latter notion relying on pages 15 and 142 of the latter reference.

Definition 1.10 (Vague convergence)**.** *A sequence* $\mathrm{S}_n \subset \mathcal{S}, n \geq 1$ *is called a localizing sequence of* S *if* $\mathrm{S}_n \subset \mathrm{S}_{n+1}$ *for every* $n \geq 1$ *and* $\bigcup_{n\geq 1} \mathrm{S}_n = \mathrm{S}$*. We say that a set* $B \subset \mathrm{S}$ *is bounded if* $B \subset \mathrm{S}_n$ *for some* $n \geq 1$*. The family of bounded sets of* $\mathcal{S}$ *will be denoted* $\hat{\mathcal{S}}$*. We write* $\hat{C}_{\mathrm{S}}$ *for the set of bounded continuous non-negative functions on* $\mathbb{R}^d$ *with bounded support.*

Let $\mathcal{M}$ *denote the set of locally finite measures on* $\mathbb{R}^d$*, i.e., the measures* μ *on* $\mathcal{S}$ *satisfying* $\mu(B) < \infty$ *for every* $B \in \hat{\mathcal{S}}$*. We endow it with the topology generated by the evaluation maps* $\pi_f : \mu \mapsto \mu(f)$ *for all* $f \in \hat{C}_{\mathrm{S}}$*. A sequence* $\mu_n \in \mathcal{M}, n \geq 1$ *converges vaguely to* $\mu \in \mathcal{M}$ *if and only if* $\mu_n(f) \to \mu(f)$ *as* $n \to \infty$ *for all* $f \in \hat{C}_{\mathrm{S}}$*.*

Note that both the set $\mathcal{M}$ of locally finite measures and the vague topology defined above depend on the choice of the localizing sequence. If ρ is a metric generating $\mathcal{S} = \mathcal{B}(\mathrm{S})$, we may choose $\hat{\mathcal{S}}$ to consist of all sets B of $\mathcal{S}$ which are bounded for the metric ρ.

In Basrak and Planinić (2019) two particular cases are presented as Examples 2.4 and 2.5.
Following Example 2.4, if S is locally compact, one can choose $\hat{\mathcal{S}}$ to be the Borel sets of S that are relatively compact. When S is Polish, this yields the vague topology on $\mathrm{M}_+(\mathrm{S})$ the set of positive Radon measures described in Resnick (2008) and used to develop extreme value theory and widely used in the litterature.
In Example 2.5, it is shown that choosing $\mathrm{S}_n = \{s \in \mathrm{S} : d(x, \mathrm{C}) > 1/n\}, n \geq 1$ for localizing sequence of S_0 yields $\mathcal{M} = \mathrm{M}_0(\mathrm{S})$ and the vague topology matches the M_0 topology defined earlier following Hult and Lindskog (2006) and Lindskog et al. (2014).

As a consequence, the two most common frameworks used in extreme value theory are special cases of the general concept of vague topology as defined in Kallenberg (2017). In particular, this simple fact allows one to adapt proofs from one framework to the other one and will be widely used in Section 4 when dealing with empirical measures. Working with the vague topology on $\mathrm{M}_+(\mathrm{S})$ mentioned above, Resnick

(2008) gives a simple criterion involving the Laplace functional of which he makes extensive use. This quantity is linked to the Laplace transform and introduced below.

Definition 1.11 (Laplace functionals)**.** *We call Laplace functional of the random measure μ in $\mathrm{M}_0(\mathrm{S})$ the map*

$$\begin{aligned}\Psi_{\mu_n} : (\mathcal{S}_0/\mathcal{B}(\mathbb{R}))_+ &\longrightarrow \mathbb{R}_+ \\ f &\longmapsto \mathbb{E}(\exp^{-\mu(f)})\end{aligned}$$

where $(\mathcal{S}_0/\mathcal{B}(\mathbb{R}))_+$ denote the collection of $\mathcal{S}_0/\mathcal{B}(\mathbb{R})$ measurable functions that are non-negative.

We can state Theorem 4.11 from Kallenberg (2017) in our particular case to get a result similar to Proposition 3.19 in Resnick (2008).

Theorem 1.12. *Let $\mu_n, n \geq 0$ and μ be random measures (i.e., random elements in $\mathrm{M}_0(\mathrm{S})$), then the following statements are equivalent :*

1. $\mu_n \xrightarrow{0d} \mu$
2. $\mu_n(f) \xrightarrow{w} \mu(f)$ *for all* $f \in \mathrm{C}^+_{b,0}(\mathrm{S})$
3. $\Psi_{\mu_n}(f) \to \Psi_\mu(f)$ *for all* $f \in \mathrm{C}^+_{b,0}(\mathrm{S})$

Relying on this theorem, we will transpose some results from Resnick (2008) to empirical measures of M_0 in Section 4.

From now on we will specialize to the cases $\mathrm{S} = \mathbb{R}^d$ and $\mathrm{S} = \mathbb{R}^d \times \mathbb{R}^d$ and we will always choose s_o to be the 0 element in the considered space. When no confusion is possible, we will simply write M_0 for $\mathrm{M}_0(\mathbb{R}^d)$ or $\mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$ and $\mathrm{C}^+_{b,0}$ for $\mathrm{C}^+_{b,0}(\mathbb{R}^d)$ or $\mathrm{C}^+_{b,0}(\mathbb{R}^d \times \mathbb{R}^d)$, whenever the underlying space is clear from the context.

1.2 Regularly varying distributions and Extreme Value Theory

Extreme value theory focuses on the study of law with heavy tails which appears in a wide range of subject including among others risk management McNeil et al. (2005), anomaly detection Goix et al. (2017) or machine learning Clémençon et al. (2022). It relies on the concept of regular variation of functions. A positive Borel measurable function b defined on a Borel set D satisfying $[a, +\infty) \subset D$ for some real a is said to be regularly varying with index $\alpha \in \mathbb{R}$ if $\lim_{t\to+\infty} b(t\lambda)/b(t) = \lambda^\alpha$ for each $\lambda > 0$.

In the non-negative one dimensional case, extreme value theory is concerned with distributions whose cumulative distribution function F such that $x \geq 0 \mapsto 1 - F(x)$— i.e., the probability for the random variable to be greater than a threshold x— is regularly varying for some index $\alpha > 0$.

The notion of $\mathcal{M}_0$ convergence then allows us to define regular variation of measures which allows one to deal with the multivariate case.

Definition 1.13. *A probability measure $P \in \mathcal{P}(\mathbb{R}^d)$ is said to be regularly varying with index $\alpha > 0$ and limit measure $\nu \in \mathrm{M}_0(\mathbb{R}^d) \setminus \{0\}$ if there exists a regularly varying function b with index $1/\alpha$ called auxiliary function such that*

$$t\, P(b(t)\cdot) \xrightarrow[t\to\infty]{0} \nu.$$

We summarize in the next theorem some basic facts about the limit measure of regular variation.

Theorem 1.14. *Let $P \in \mathcal{P}(\mathbb{R}^d)$ be regularly varying with index $\alpha > 0$, limit measure $\nu \in \mathrm{M}_0^+(\mathbb{R}^d) \setminus \{0\}$ and auxiliary function b, then*

(a) The limit measure is homogenous in the sense that we have

$$\nu(\lambda^{-1/\alpha}\cdot) = \lambda\nu(\cdot), \qquad \lambda > 0.$$

(b) By redefining b if necessary, we can choose the limit measure ν to be such that

$$\nu(\{x \,:\, |x| > 1\}) = 1.$$

Let $X \sim \nu$ as above, then $|X|$ is regularly varying with the same index and auxiliary function as X and if F denotes its cumulative distribution function, we have that $1 - F$ is regularly varying with index $-\alpha$. Let Q denote the quantile function of $|X|$, then $t \mapsto Q(1 - 1/t)$ is regularly varying with index $1/\alpha$ and we can choose this function for auxiliary function of both X and $|X|$.

We consider the polar transformation

$$\begin{aligned} \mathrm{POLAR} : \mathbb{R}^d \setminus \{0\} &\longrightarrow (\mathbb{R}_+ \times \mathbb{S}^{d-1}) \setminus (\{0\} \times \mathbb{S}^{d-1}) \\ x &\longmapsto (|x|, x/|x|) \end{aligned}$$

where we endow $\mathbb{S}^{d-1}$ with the topology induced by the one of $\mathbb{R}^d$, and $\mathbb{R} \times \mathbb{S}^{d-1}$ with the product topology. It is clear that POLAR is a homeomorphism whose inverse function is simply given by

$$\begin{aligned} \mathrm{POLAR}^{-1} : (\mathbb{R}_+ \times \mathbb{S}^{d-1}) \setminus (\{0\} \times \mathbb{S}^{d-1}) &\longrightarrow \mathbb{R}^d \setminus \{0\} \\ (\rho, \theta) &\longmapsto \rho\theta \end{aligned}$$

We also introduce the family $\nu_\alpha, \alpha > 0$ of Borel measures on $\mathbb{R}_+$ defined by $\nu_\alpha(\{t \in \mathbb{R} : |t| \geq s\}) = s^{-\alpha}$ for every $s > 0$. Setting $\mathrm{C} = \{0\} \times \mathbb{S}^{d-1}$ we define the set $\mathrm{M}_0((0,\infty) \times \mathbb{S}^{d-1})$ and Corollary 4.4 in Lindskog et al. (2014) yields that both $\mu \in \mathrm{M}_0(\mathbb{R}^d) \mapsto \mathrm{POLAR}_\# \, \mu \in \mathrm{M}_0((0,\infty) \times \mathbb{S}^{d-1})$ and $\mu \in \mathrm{M}_0((0,\infty) \times \mathbb{S}^{d-1}) \mapsto \mathrm{POLAR}_\#^{-1} \, \mu \in \mathrm{M}_0(\mathbb{R}^d)$ are continuous. As a consequence we have the following theorem.

Theorem 1.15. *For a random variable X in $\mathbb{R}^d$ with distribution P_X, the two following assertions are equivalent.*

(a) X is regularly varying with index $\alpha > 0$, auxiliary function b.

(b) $\mathrm{POLAR}_\# \, P_X \xrightarrow{0} \nu_\alpha \otimes H$ in $\mathrm{M}_0((0,\infty) \times \mathbb{S}^{d-1})$ for some finite measure H on $\mathbb{S}^{d-1}$.

If (a) or (b) holds, the limit measure $\overline{\nu} \in \mathrm{M}_0^+(\mathbb{R}^d) \backslash \{0\}$ of regular variation of P_X satisfies $\mathrm{POLAR}_\# \, \overline{\nu} = \nu_\alpha \otimes H \in \mathrm{M}_0((0,\infty) \times \mathbb{S}^{d-1})$, or equivalently $\mathrm{POLAR}_\#^{-1} \, (\nu_\alpha \otimes H) = \overline{\nu} \in \mathrm{M}_0(\mathbb{R}^d)$.

Remark 1.16. *Note that this theorem gives the form of a limit measure of regular variation.*

To conclude this subsection, we introduce a decomposition that is usual when dealing with heavy-tailed distribution to build tractable approximations in practice. It will give us some motivation to proceed in a similar way using $\mathbb{C}(1 - 1/t)$ the center-outward quantile regions of order $1 - 1/t, t > 0$ defined in Subsection 1.6 instead of the sets $\{|X| \leq b(t)\}, t > 0$. We will come back to this simple decomposition, and the convergence of each term of it, in each subsequent section to provide the rationale for our work.

For X as above and A a Borel set of $\mathbb{R}^d$, we can write, spliting $\mathbb{R}^d$ in two according to the value taken by $|X|$,

$$
\begin{aligned}
&\mathsf{P}(X \in A) \\
&= \mathsf{P}(X \in A, |X| \leq b(t)) + \mathsf{P}(X \in A, |X| > b(t)) \\
&= \mathsf{P}(X \in A, |X| \leq b(t)) + \mathsf{P}(b(t)^{-1} X \in (b(t)^{-1} A) \cap \{x : |x| > 1\})
\end{aligned}
$$

whence the decomposition

$$
\begin{aligned}
\mathsf{P}(X \in A) = \mathsf{P}(X \in A \cap \mathrm{cl}\, \mathrm{B}_{0,b(t)}) + \\
\mathsf{P}(|X| > b(t)) \, \mathsf{P}(b(t)^{-1} X \in (b(t)^{-1} A) \cap \mathrm{int}\, \mathrm{B}^c_{0,b(t)} \mid |X| > b(t))
\end{aligned}
$$

Notice that $A \cap \operatorname{cl} \mathrm{B}_{0,b(t)}$ and $\{x : |x| > b(t)\}$ are respectively increasing to A and decreasing to $\varnothing$, whence $\mathsf{P}(X \in A \cap \operatorname{cl} \mathrm{B}_{0,b(t)}) \to \mathsf{P}(X \in A)$ and $\mathsf{P}(|X| > b(t)) \to 0$ as $t \to \infty$.

We now focus on the conditional probability in the last term of the sum. Let $\nu_{(t)}, t > 0$ be a sequence of measures in $\mathrm{M}_0(\mathbb{R}^d)$ defined by

$$\begin{aligned}\nu_{(t)} &= \mathsf{P}\left(X/b(t) \in \cdot \cap \{x : |x| > 1\} \mid |X| > b(t)\right) \\ &= \frac{t}{t} \frac{\mathsf{P}\left(X/b(t) \in \cdot \cap \{x : |x| > 1\}\right)}{\mathsf{P}\left(|X|/b(t) > 1\right)}.\end{aligned}$$

Since $|X|$ is regularly varying with index $\alpha > 0$ and auxiliary function b, we have $t\,\mathsf{P}(|X|/b(t) > 1) = t\,\mathsf{P}_X(b(t)\,\{x : |x| > 1\}) \to \bar{\nu}(\{x : |x| > 1\}) = 1$ as $t \to \infty$ for a properly normalized auxiliary function b. As a consequence, to study the M_0-convergence of $\nu_{(t)}$, one just need to deal with $\nu'_{(t)} = t\ \mathsf{P}\left(X/b(t) \in \cdot \cap \{x : |x| > 1\}\right)$. Let A be a Borel set of $\mathbb{R}^d$ such that $0 \notin \operatorname{cl} A$. We claim that $\bar{\nu}(\partial A)) = 0$ implies $\bar{\nu}(\partial(A \cap \{x : |x| > 1\}) = 0$. Indeed, it is clear that $\partial(A \cap \{x : |x| > 1\})$ is included in $\partial A \cup \{x : |x| = 1\}$, hence $\bar{\nu}(\partial(A \cap \{x : |x| > 1\}) \leq \bar{\nu}(\partial A) + \bar{\nu}(\{x : |x| = 1\})$ and the conclusion. Therefore, Theorem 1.8 yields the M_0-convergence of $\nu'_{(t)}$ to $\bar{\nu}(\cdot \cap \{x : |x| > 1\})$. Finally we have the convergence, as $n \to \infty$,

$$\mathsf{P}\left(X/b(t) \in \cdot \mid |X| > b(t)\right) \xrightarrow{0} \bar{\nu}(\cdot \cap \{x : |x| > 1\}).$$

In practice, the decomposition is used together with the latter convergence by assuming that the convergence is achieved for some t^* large enough. Thus we write the approximation

$$\mathsf{P}(X \in A) \approx (1 - p_{t^*})\,\mathsf{P}\left(X \in A \mid |X| \leq b(t)\right) + p_{t^*}\bar{\nu}\left((b(t^*)^{-1}A) \cap \{x : |x| > 1\}\right)$$

where $p_{t^*} = \mathsf{P}(|X| > b(t^*))$.

The rationale behind this approximation is as follows. It seems reasonable to use a simpler model in the area of the space where we have few data. In our context, we will have fewer data available as the norm $|X|$ of the random variable X increases - a distance to the origin. This prompts us to divide the space in half according to the value taken by $|X|$. The empirical counterpart of $p_{t^*} = \mathsf{P}(|X| > b(t^*))$ will be the share the data considered extreme, associated with a threshold $b(t^*)$ for a realization $|x|$ of $|X|$ to be considered extreme. Assuming that the random variable X is regularly varying, the limit in $\mathrm{M}_0(\mathbb{R}^d)$ of $\mathsf{P}\left(X/b(t) \in \cdot \mid |X| > b(t)\right)$ as $t \to \infty$ is quite simpler than a general measure in $\mathrm{M}_0(\mathbb{R}^d)$ since it only depends on the index of regular variation $\alpha > 0$ and on a measure H on the unit sphere $\mathbb{S}^{d-1}$, hence the approximation yields a very simple model for the data above the threshold and we hope to have enough data to estimate the parameters. In one dimensional case, one juste needs to estimate at most three parameters : the index $\alpha > 0$, the probability for an extreme to be respectively negative and positive. Under the assumption that X takes value in $\mathbb{R}_+$ or in $\mathbb{R}_-$, only the index $\alpha > 0$ needs to be estimated. However, as often in statistics, the higher the dimension d, the harder it becomes to estimate the measure H.

To sum up, for t large enough the first term in the approximation carries the non-extremal part of X and the second the extremal one. Extreme value theory is used to estimate the conditional probability for large t and one hopes that the simple model it yields will allow to achieve better estimation than using usual statistical methods when $\operatorname{hzn} A$ is non-empty.

1.3 Multivalued maps and set convergence

In this section, we introduce the concepts needed to deal with subdifferentials of closed convex functions that will arise in optimal transport between finite Borel measures P, Q (see Subsection 1.5) and which are random when dealing with measures that are random too.

To begin with, we introduce multivalued maps. This notion covers the subdifferential of a closed convex function ψ and even seems necessary to deal with, since for some x in the interior of the domain of ψ, the sub-gradient $\partial\psi(x)$ may not be a single element but rather a set. We refer to Rockafellar and Wets (1998) for details, especially Chapter 5.

Definition 1.17 (multivalued maps)**.** *We define a multivalued map T from $\mathcal{X} \subset \mathbb{R}^k$ to $\mathcal{Y} \subset \mathbb{R}^l$ as an application that sends an element x of $\mathcal{X}$ to a subset $T(x)$ of $\mathcal{Y}$ and write*

$$\begin{aligned} T : \mathcal{X} &\rightrightarrows \mathcal{Y} \\ x &\mapsto T(x) \end{aligned}$$

The graph of such a map is $\operatorname{gph}(T) = \{(x, y) \in \mathcal{X} \times \mathcal{Y} \,:\, y \in T(x)\}$*. The domain and the range of T are* $\operatorname{dom} T = \{x \in \mathcal{X} \,:\, T(x) \neq \varnothing\}$ *and* $\operatorname{rge} T = \{y \in \mathcal{Y} \,:\, \exists x \in \mathcal{X}, y \in T(x)\}$*. The inverse of T is simply $T^{-1} : y \in \mathcal{Y} \mapsto \{\, x \in \mathcal{X} \,:\, y \in T(x)\}$ and its graph is defined by* $\operatorname{gph} T^{-1} = \{(y, x) \in \mathcal{Y} \times \mathcal{X} \,:\, (x, y) \in \operatorname{gph} T\}$.

There is a one-to-one correspondence between subsets of $\mathcal{X} \times \mathcal{Y}$ and multivalued maps $\mathcal{X} \rightrightarrows \mathcal{Y}$. A subset of $\mathcal{X} \times \mathcal{Y}$ can be identified with the multivalued map of which it is the graph. Due to this correspondance, when no confusion is possible, we will often write T for $\operatorname{gph} T$ *and $(x, y) \in T$ instead of $(x, y) \in$* $\operatorname{gph} T$.

Since we will be interested in sequences of multivalued maps, it will be useful to define the convergence of such sequences in some sense. Using the identification between multivalued maps and their graphs, one can define the convergence of multivalued maps as the convergence of their graphs. To do so, we first need a notion of set convergence; the one used in Rockafellar and Wets (1998) is the Painlevé–Kuratowski convergence.

Definition 1.18 (Painlevé–Kuratowski convergence)**.** *Let $A_n, n \geq 1$ be a sequence of subsets of $\mathbb{E}$, where $\mathbb{E}$ is itself an open subset of $\mathbb{R}^d$. We define the inner and outer limits of $T_n, n \geq 1$ respectively as*

$$\liminf_{n\to\infty} A_n = \{x \in \mathbb{E} \ : \ \text{there exists } x_n \in A_n, n \geq 1, \text{ such that } x_n \to x \text{ as } n \to \infty\}$$

and

$$\limsup_{n\to\infty} A_n = \{x \in \mathbb{E} \ : \ \text{there exist infinite } N \subset \mathbb{N}, x_n \in A_n, \\ \text{such that } x_n \to x \text{ as } n \to \infty \text{ in } N\}.$$

The sequence $A_n \subset \mathbb{E}, n \geq 1$ converges in the Painlevé–Kuratowski sense to $A \subset \mathbb{E}$ if and only if $A = \liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$. The inner and outer limits are necessarily closed in the trace topology on $\mathbb{E}$.

For a sequence $T_n, n \geq 1$ of maps $\mathcal{X} \rightrightarrows \mathcal{Y}$, we write $\overline{T}$, $\underline{T}$ for the maps respectively associated with the graphs $\limsup_{n\to\infty} \operatorname{gph} T_n$ and $\liminf_{n\to\infty} \operatorname{gph} T_n$.

Definition 1.19 (Graphical convergence)**.** *Let $T, T_n, n \geq 1$ be multivalued maps $\mathbb{R}^k \supset \mathcal{X} \rightrightarrows \mathcal{Y} \subset \mathbb{R}^k$. We say that the sequence of maps T_n converges graphically to T if $\operatorname{gph} T_n$ converges to $\operatorname{gph} T$ in the sense of Painlevé–Kuratowski. We write $T_n \xrightarrow{g} T$ for this convergence.*

Let V an open subset of $\mathcal{X} \subset \mathbb{R}^k$ in the trace topology on $\mathcal{X}$ induced by $\mathbb{R}^k$. We write $T \llcorner V$ for the restriction of the map T to V. It is clear that $\operatorname{gph}(T \llcorner V) = \operatorname{gph} T \cap (V \times \mathcal{Y})$. We say that T_n converges graphically to T relative to the open set V if $T(x) = \overline{T}(x) = \underline{T}(x)$ for every $x \in V$, i.e., $T \llcorner V = \overline{T} \llcorner V = \underline{T} \llcorner V$.

In practice, the multivalued maps we will need to deal with will have closed graphs and will be data dependent, hence random. Using again the identification between multivalued maps and their graphs, this calls for endowing the set $\mathcal{F}(\mathcal{X} \times \mathcal{Y})$ of closed subsets of $\mathcal{X} \times \mathcal{Y}$ with a topology with good properties allowing one to use classical results from probability theory. We do so following Molchanov (2017). The following material is mainly taken from Appendix C in the latter book.

For an open subset $\mathbb{E}$ of $\mathbb{R}^d$, let $\mathcal{F}(\mathbb{E})$ denote the set of all subsets of $\mathbb{E}$ which are closed in the trace topology on $\mathbb{E}$. For A a subset of $\mathbb{E}$, we consider the collection $\mathcal{F}_A = \{F \in \mathcal{F}(\mathbb{E}) : F \cap A \neq \varnothing\}$ of sets that hit A, and the collection of sets $\mathcal{F}^A = \{F \in \mathcal{F}(\mathbb{E}) : F \cap A = \varnothing\}$ that miss A.

Definition 1.20 (Fell topology)**.** *The Fell hit-and-miss topology is defined by the sub-base consisting of $\mathcal{F}_G$ for all open subset G of $\mathbb{E}$ and $\mathcal{F}^K$ for all compact subsets K of $\mathbb{E}$. Let $F, F_n \in \mathcal{F}(\mathbb{E}), n \geq 1$. The sequence $F_n, n \geq 1$ $\mathcal{F}$-converges to F as $n \to \infty$ in the Fell topology if for every open $G \subset \mathbb{E}$ such that $F \in \mathcal{F}_G$ we have $F_n \in \mathcal{F}_G$ for all large n, and for every compact $K \subset \mathbb{E}$ satisfying $F \in \mathcal{F}^K$ we have $F_n \in \mathcal{F}^K$ for all large n. We write $F_n \xrightarrow{\mathcal{F}} F$ for this convergence.*

The following proposition summarizes basic results about the Fell topology which allows one to benefit from general results from probability theory when dealing with random elements in $\mathcal{F}(\mathbb{E})$.

Proposition 1.21. *Let $\mathbb{E}$ be an open subset of $\mathbb{R}^d$ endowed with the trace topology, then the following assertions are true.*

(a) For a sequence $A_n \in \mathcal{F}(\mathbb{E}), n \geq 1$ of closed sets, the convergence in the Fell topology is equivalent to the Painlevé–Kuratowski convergence.

(b) $\mathcal{F}(\mathbb{E})$ is Polish and the Hausdorff–Busemann metric ρ_{HB} metrises the Fell topology. It is defined by

$$\rho_{HB}(F_n, F) = \sup_{x \in \mathbb{R}^d} \exp^{-\rho(x,x_0)} |\rho(x, F) - \rho(x, F_n)|$$

where x_0 is a fixed point of $\mathbb{E}$ and ρ denotes the Euclidean distance on $\mathbb{E}$.

(c) The space $\mathcal{F}(\mathbb{E})$ is compact. As a consequence, any sequence $F_n, n \geq 1$ of random elements in $\mathcal{F}(\mathbb{E})$ is tight.

The following proposition gives a characterization of the graphical convergence relative to an open set in terms of $\mathcal{F}$-convergence when only multivalued maps with closed graphs are considered.

Proposition 1.22 (Characterization of graphical convergence relative to an open set)**.** *Let $T, T_n, n \geq 1$ be multivalued maps $\mathbb{R}^k \rightrightarrows \mathbb{R}^k$, for some $k, l \geq 1$, and let V be an open subset of $\mathbb{R}^k$. Assume $T, T_n, n \geq 1$ are closed, i.e., $\operatorname{gph} T, \operatorname{gph} T_n \in \mathcal{F}(\mathbb{R}^k \times \mathbb{R}^l), n \geq 1$, then the restrictions $T \llcorner V, T_n \llcorner V, n \geq 1$ of the maps to the open set V are closed as sets in the trace topology on $V \times \mathbb{R}^l$, i.e., $\operatorname{gph} T, \operatorname{gph} T_n \in \mathcal{F}(V \times \mathbb{R}^l), n \geq 1$. Moreover, T_n converges graphically to T relative to the open set V if and only if $T_n \llcorner V$ $\mathcal{F}$-converges to $T \llcorner A$ in $\mathcal{F}(V \times \mathbb{R}^l)$.*

The following definition and lemma are taken from Segers (2022) (Definition D.7 and Lemma D.8).

Definition 1.23. *A map from a topological space* $(\mathbb{D}, \mathrm{D})$ *into* $\mathcal{F}(\mathbb{E})$ *is lower semi-continuous if and only if for every open* $G \subset \mathbb{E}$*, the inverse image of* $\mathcal{F}_G$ *is open. A map from a topological space into* $\mathcal{F}(\mathbb{E})$ *is upper semi-continuous if and only if for every compact* $K \subset \mathbb{E}$*, the inverse image of* $\mathcal{F}^K$ *is open.*

Lemma 1.24 (Lower semi-continuity)**.** *Let* $(\mathbb{D}, \rho)$ *be a metric space, let* $\mathbb{E}$ *be a LCHS space, and let* $f : \mathbb{D} \to \mathcal{F}(\mathbb{E})$*. The following statements are equivalent:*

(i) For every open $G \subset \mathbb{E}$*, the set* $f^{-1}(\mathcal{F}_G)$ *is open, i.e.,* f *is lower semi-continuous.*

(ii) If the points $x_n, x \in \mathbb{D}$ *and the open set* $G \subset \mathbb{E}$ *are such that* $f(x) \cap G \neq \varnothing$ *and* $\lim_{n\to\infty} x_n = x$*, then also* $f(x_n) \cap G \neq \varnothing$ *for all large* n*.*

(iii) If $\lim_{n\to\infty} x_n = x$ *in* $\mathbb{D}$*, then* $f(x) \subset \liminf_{n\to\infty} f(x_n)$*.*

(iv) The set $\{(x, F) \in \mathbb{D} \times \mathcal{F} : f(x) \subset F\}$ *is closed in the product topology.*

An application $\mathbb{D} \to \mathcal{F}(\mathbb{E})$ is said to be continuous if it is both upper and lower semi-continuous.

Remark 1.25. *It is important to note that some applications one could hope to be continuous are not. We refer to Propositions E.10 and E.11 in Molchanov (2017) for a short non-exhaustive list of continuous and semi-continuous transformations from* $\mathcal{F}(\mathbb{E})$ *or* $\mathcal{F}(\mathbb{E} \times \mathbb{E})$ *to* $\mathcal{F}(\mathbb{E})$*. For example,* $(F, F') \in \mathcal{F}(\mathbb{E} \times \mathbb{E}) \longmapsto F \cup F' \in \mathcal{F}(\mathbb{E})$ *is continuous while* $(F, F') \in \mathcal{F}(\mathbb{E} \times \mathbb{E}) \longmapsto F \cap F' \in \mathcal{F}(\mathbb{E})$ *is only upper semi-continuous.*

Let $\mathrm{proj}_1 : (x, y) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto x \in \mathbb{R}^d$ and $\mathrm{proj}_2; (x, y) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto y \in \mathbb{R}^d$ denote the projections onto the first and last d components in $\mathbb{R}^d \times \mathbb{R}^d$. In Section 4, we will be interested in the convergence of the sequence $\mathrm{proj}_2(F_n)$, $n \geq 1$ where $F_n \in \mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d)$, $n \geq 1$. It is clear that the maps $\mathrm{proj}_i, i = 1, 2$ are continuous. Let $F \in \mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d)$. The sets $\mathrm{proj}_1(F) = \operatorname{dom} F$ and $\mathrm{proj}_2(F) = \operatorname{rge} F$ are not necessarily closed, hence we would rather consider the functions $\mathrm{cl} \circ \mathrm{proj}_i, i = 1, 2$ which are well-defined from $\mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d)$ to $\mathcal{F}(\mathbb{R}^d)$. One can show—we refer to Lemma D.10 in Segers (2022) for a proof—that $\mathrm{cl} \circ \mathrm{dom}$ and $\mathrm{cl} \circ \mathrm{rge}$ are both lower semi-continuous.

1.4 Convex functions and cyclical monotonicity

In this subsection we introduce convex functions and cyclical monotonicity, which will be shown in Subsection 1.5 to be strongly connected to optimal transport.

Definition 1.26 (Convex functions)**.** *A function $f : \mathbb{R}^d \longrightarrow \mathbb{R} \cup \{+\infty\}$ is said to be convex if*

$$f\left((1-\tau)x + \tau y\right) \leq (1-\tau)f(x) + \tau f(y)$$

holds for every $x, y \in \mathbb{R}^d$ and $\tau \in (0,1)$. It is proper *if $f(x) < +\infty$ for some x in $\mathbb{R}^d$. The domain of f is defined by* $\operatorname{dom} f = \{x \in \mathbb{R}^d : f(x) \in \mathbb{R}\}$*.*

Definition 1.27 (Lower semi-continuity)**.** *The function $f : \mathbb{R}^d \longrightarrow \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous (lsc) at $\bar{x} \in \mathbb{R}^d$ if*

$$\liminf_{x \to \bar{x}} f(x) \geq f(\bar{x}), \ \text{ or equivalently } \liminf_{x \to \bar{x}} f(x) = f(\bar{x}).$$

If f is lower semi-continuous at every point of $\mathbb{R}^d$, we simply say that f is lower semi-continuous or closed.

According to Theorem 1.6 in Rockafellar and Wets (1998), the lower semi-continuity of a function $f : \mathbb{R}^d \longrightarrow \mathbb{R} \cup \{+\infty\}$ is equivalent to the closedness of its epigraph defined by $\operatorname{epi} f = \left\{(x, \alpha) \in \mathbb{R}^d \times \mathbb{R} \ : \ f(x) \leq \alpha\right\}$.

Definition 1.28 (Subgradients and subdifferential)**.** *For any proper, convex function $f : \mathbb{R}^d \longrightarrow \mathbb{R} \cup \{+\infty\}$ and any point $\bar{x}$ lying in* $\operatorname{dom} f$*, we call subgradient of f at $\bar{x}$ any point $v \in \mathbb{R}^d$ satisfying*

$$f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle$$

for every x in $\mathbb{R}^d$. We let $\partial f(\bar{x})$ denote the set containing all such v and call it the subdifferential of f at $\bar{x}$. Defined this way, the subdifferential ∂f of f is a multivalued map $\mathbb{R}^d \rightrightarrows \mathbb{R}^d$.

The following theorem and remark give, for any measure μ vanishing on sets of Hausdorff dimension at most $d-1$, the existence of a set $A \subset \mathbb{R}^d$ such that (a) $\mu(A^c) = 0$ and (b) $\partial f(x)$ is a singleton for any x lying in A, i.e., f is differentiable at every $x \in A$ and $\partial f(x) = \{\nabla f(x)\}$ for such x.

Theorem 1.29 (Theorem 25.5 in Rockafellar (1970))**.** *Let f be a proper convex function on $\mathbb{R}^d$, and let D be the set of points where f is differentiable. Then D is a dense subset of* $\operatorname{int} \operatorname{dom} f$ *and its complement in* $\operatorname{int} \operatorname{dom} f$ *is a set of Lebesgue measure zero. Furthermore, the gradient mapping $\nabla f : x \mapsto \nabla f(x)$ is continuous on D.*

Remark 1.30. *It has been shown in Anderson and Klee (1952) that the set D in the theorem above can be chosen to have a complement of Hausdoff dimension at most $d-1$.*

We now introduce the well known Legendre–Fenchel Transform and recall one of its basic properties.

Definition 1.31 (Legendre–Fenchel Transform)**.** *For any function $f : \mathbb{R}^d \longrightarrow \mathbb{R} \cup \{+\infty\}$ such that $f(x) < \infty$ for some x, we call Legendre–Fenchel transform of f the map $f^* : \mathbb{R}^d \longrightarrow \mathbb{R} \cup \{+\infty\}$ defined by*

$$f^*(v) := \sup_x \{\langle v, x\rangle - f(x)\}$$

and say that f^* is conjugate to f.

Theorem 1.32. *For any function $f : \mathbb{R}^d \longrightarrow \mathbb{R} \cup \{+\infty\}$ such that $f(x) < \infty$ for some x, f^* is proper, lsc and convex.*

The next theorem is taken from Proposition 11.3 in Rockafellar and Wets (1998) and provides a relation between the subdifferentials of f and those of its conjugate f^*.

Theorem 1.33 (Inversion rule for subgradient relations)**.** *For any proper, lsc, convex function f, one has $\partial f^* = (\partial f)^{-1}$ and $\partial f = (\partial f^*)^{-1}$. Indeed,*

$$\bar{v} \in \partial f(\bar{x}) \iff \bar{x} \in \partial f^*(\bar{v}) \iff f(\bar{x}) + f^*(\bar{v}) = \langle \bar{x}, \bar{v}\rangle,$$

whereas $f(x) + f^*(v) \geq \langle x, v\rangle$ for all x, v. Hence $\operatorname{gph} \partial f$ is closed and

$$\partial f(\bar{x}) = \arg\max_v \{\langle \bar{x}, v\rangle - f^*(v)\}, \qquad \partial f^*(\bar{v}) = \arg\max_x \{\langle x, \bar{v}\rangle - f(x)\}.$$

Remark 1.34. *The latter theorem will be used in Subsection 1.6 in the following way. Let μ and ν be two measures on $\mathbb{R}^d$ that vanish on sets of Hausdorff dimension at most $d-1$. Assume that we have a proper, lsc convex function $f : \mathbb{R}^d \longrightarrow \mathbb{R} \cup \{+\infty\}$ and let g denote its conjugate f^*. Applying Theorem 1.32, Theorem 1.29 and Remark 1.30 we get two subsets A, B of $\mathbb{R}^d$ such that $\mu(A^c) = 0$, $\nu(B^c) = 0$ and f, g are differentiable respectively on A, B. We further assume that μ and ν satisfy $\nu = \nabla f_{\#}\mu$ and $\mu = \nabla g_{\#}\nu$. Let $\tilde{A} = A \cap [\nabla f]^{-1}(B)$, $\tilde{B} = B \cap [\nabla g]^{-1}(A)$. It is clear that $\mu(\tilde{A}^c) = \nu(\tilde{B}^c) = 0$. Indeed, we can write successively*

$$\begin{aligned}\mu(\tilde{A}^c) &= \mu\left([\nabla f]^{-1}(B^c)\right) \\ &= \nu(B^c) = 0\end{aligned}$$

and a similar computation can be done for $\nu(\tilde{B}^c)$. As a consequence we have

$$\nabla g \circ \nabla f = \mathrm{Id} \ \ \mu\text{-}a.e \qquad \textit{and} \qquad \nabla f \circ \nabla g = \mathrm{Id} \ \ \nu\text{-}a.e \ .$$

We now introduce the concept of (cyclical) monotonicity which is linked to the convexity of multivalued maps.

Definition 1.35 (Cyclically monotone maps and sets)**.** *A set $T \subset \mathbb{R}^k \times \mathbb{R}^l$ is cyclically monotone if, for every integer $n \geq 1$ and all $(x_1, y_1), \dots, (x_n, y_n) \in T$ with $y_{n+1} := y_1$, we have*

$$\sum_{i=1}^{n} \langle x_i, y_i \rangle \geq \sum_{i=1}^{n} \langle x_i, y_{i+1} \rangle \tag{2}$$

If (2) holds for $n = 2$, T is monotone. The set T is maximal (cyclically) monotone if it is not contained in a larger subset of $\mathbb{R}^d \times \mathbb{R}^d$ that is also (cyclically) monotone. A multivalued map $T : \mathbb{R}^k \rightrightarrows \mathbb{R}^l$ is (cyclically) monotone if its graph $\operatorname{gph} T$ *is (cyclically) monotone.*

The following theorem is given in Rockafellar and Wets (1998) as Proposition 12.6 for maximal monotone extension with a proof relying on Zorn's lemma that holds for maximal cyclically monotone extension.

Theorem 1.36 (Existence of maximal extensions)**.** *For any (cyclically) monotone mapping $T : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ there is a maximal (cyclically) monotone mapping $\bar{T} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$—which is not necessarily unique—such that* $\operatorname{gph} T \subset \operatorname{gph} \bar{T}$.

In general, for a family $\mathcal{F}$ of sets included in $\mathbb{R}^k \times \mathbb{R}^l$, $\mathcal{F}_{(c)m}$ will denote the subfamily consisting of (cyclically) monotone sets. For a family M of measures on the Borel sets of $\mathbb{R}^d$, we write $\mathrm{M}_{(c)m}$ for the family of measures in M whose support is (cyclically) monotone. Likewise we add m(c)m for maximal (cyclically) monotonicity.

The following theorem due to Rockafellar and Wets gives a connection between convex functions and cyclically monotone maps. It is taken from Theorem 24.8 and 24.9 in Rockafellar (1970) and Theorem 12.25 in Rockafellar and Wets (1998).

Theorem 1.37 (Rockafellar and Wets)**.** *(a) A map $T : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ is cyclically monotone if and only if it is contained in the subdifferential of a closed convex function.*

(b) A mapping $T : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ has the form $T = \partial\psi$ for some proper, lsc, convex function $\psi : \mathbb{R}^d \mapsto \mathbb{R} \cup \{+\infty\}$ if and only if T is maximal cyclically monotone. Then ψ is determined by T uniquely up to an additive constant.

We conclude this subsection with a lemma that appeared as Lemma 3.2 in Segers (2022) and will be used primarily in Subsection 2.2.1 to ensure that limit points of sequences of sets inherit some properties. Recall from Subsection 1.3 the notation $\mathcal{F}_A = \{F \in \mathcal{F} : F \cap A = \varnothing\}$ for $A \subset \mathbb{R}^d$.

Lemma 1.38. *The sets $\mathcal{F}_m$, $\mathcal{F}_{cm}$, $\mathcal{F}_{mm} \cup \{\varnothing\}$, $\mathcal{F}_{mcm} \cup \{\varnothing\}$ are all closed in $\mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d)$. The complements of $\mathcal{F}_m$ and $\mathcal{F}_{cm}$ are unions of finite intersections of collections $\mathcal{F}_G$ for open $G \subset \mathbb{R}^d$.*

1.5 Measure transportation and couplings

Throughout this text, optimal transport will always refer to optimal transport with quadratic distance as a cost.

Let P, Q be two probability measures on $\mathbb{R}^d$. We write $\Phi_{\#}\mu$ for the push-forward of the measure μ by the application Φ. The Monge Problem consists in minimizing, for T a Borel measurable map satisfying $T_{\#}P = Q$, the quantity

$$\int_{\mathbb{R}^d} |x - T(x)|^2 \mathrm{d}P(x) \in \mathbb{R}.$$

If the minimum is achieved for some T we call T an optimal transport map. We can easily think of an optimal transport map through the discrete case. For each point x of the support of P, it sends the mass that P puts on x to exactly one point $T(x)$ of $\operatorname{spt}(Q)$ while minimizing the sum of the quadratic displacement of the points. Moreover, $T_{\#}P = Q$ means that there is a conservation of the total mass during the transportation; there is no transport map between measures with different total masses.

Since the Monge Problem may not have solutions without further assumptions, a relaxation of it called the Kantorovich Problem is often considered. We will introduce it now.

Let $\mathrm{proj}_1 : (x, y) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto x \in \mathbb{R}^d$ and $\mathrm{proj}_2; (x, y) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto y \in \mathbb{R}^d$ denote the projections onto the first and last d components in $\mathbb{R}^d \times \mathbb{R}^d$. We denote by $\mathrm{M}^+(\mathrm{S})$ the space of finite, nonnegative Radon measures on (S, d).

Definition 1.39 (Marginals)**.** *Let $\gamma \in \mathrm{M}^+(\mathbb{R}^d \times \mathbb{R}^d)$. We call* $\mathrm{proj}_{1\#}\,\gamma$ *and* $\mathrm{proj}_{2\#}\,\gamma$ *the left and right marginals of μ, and introduce the map $\phi_i : \gamma \in \mathrm{M}^+(\mathbb{R}^d \times \mathbb{R}^d) \mapsto$* $\mathrm{proj}_{i\#}\,\gamma \in \mathrm{M}^+(\mathbb{R}^d)$ *for $i = 1, 2$.*

Note that the space $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ is a subset of $\mathrm{M}^+(\mathbb{R}^d \times \mathbb{R}^d)$. Endowing the spaces $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with the usual weak topologies, we derive the following lemma.

Lemma 1.40 (Marginals of probability measures)**.** *The restriction of ϕ_1 and ϕ_2 to $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$, the space of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$, is a continuous function $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \to \mathcal{P}(\mathbb{R}^d)$.*

Proof. One just needs to apply the continuous mapping theorem; we refer to Theorem 5.27 in Kallenberg (2021). $\square$

Definition 1.41 (Coupling)**.** *Let $\mu, \nu \in \mathrm{M}^+(\mathbb{R}^d)$, the set $\Pi(\mu, \nu)$ of couplings between μ and ν is made up of all Borel measures π on $\mathbb{R}^d \times \mathbb{R}^d$ such that for each $A \in \mathcal{B}(\mathbb{R}^d)$,*

$$\pi(A \times \mathbb{R}^d) = \mu(A) \quad \textit{and} \quad \pi(\mathbb{R}^d \times A) = \nu(A).$$

In other words, $\pi \in \mathrm{M}^+(\mathbb{R}^d \times \mathbb{R}^d)$ lies in $\Pi(\mu, \nu)$ if and only if

$$\phi_1(\pi) = \mu \quad \textit{and} \quad \phi_2(\pi) = \nu.$$

We will write $\Pi_{cm}(P, Q)$ for the set of couplings between P and Q whose support is cyclically monotone according to the definition given later on.

Remark 1.42.

1. *Whenever $P, Q \in \mathrm{M}^+(\mathbb{R}^d)$ have the same finite non-zero mass m, the set $\Pi(\mu, \nu)$ is not empty. Indeed, one just needs to consider $\pi = \mu \otimes \nu / m$.*

2. *In particular, for $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d)$ with finite equal mass, $\Pi(\tilde{\mu}, \tilde{\nu})$ is not empty, and, if $\tilde{\pi} \in \Pi(\tilde{\mu}, \tilde{\nu})$, then necessarily $\tilde{\pi}$ puts no mass on $\{0\} \times \mathbb{R}^d$ and $\mathbb{R}^d \times \{0\}$. We use here a notation with a tilde over π since $\tilde{\pi}(\{0\}) = 0$ and the restriction $\mathrm{res}(\tilde{\pi})$ of $\tilde{\pi}$ to $(\mathbb{R}^d \times \mathbb{R}^d) \setminus \{0\}$ lies in $\mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$ using arguments similar to Lemma 2.5.*

3. *If both μ and ν have infinite mass, the simple example of coupling used in previous points no longer works. However, if we set $\bar{\pi} = \tilde{\mu} \otimes \delta_0 + \delta_0 \otimes \tilde{\nu} + \alpha \delta_0$ for some $\alpha \geq 0$ and measures $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d)$ not necessarily with equal mass, then $\Pi(\mu', \nu')$ is not empty for some $\mu' \in \bar{\mu}$ and $\nu' \in \bar{\nu}$. Indeed, if $\bar{\pi} \in \Pi(\mu', \nu')$, we necessarily have*

$$\mu'(\{0\}) = \bar{\pi}(\{0\} \times \mathbb{R}^d) = \bar{\pi}(\{0\}) + \bar{\pi}(\{0\} \times (\mathbb{R}^d \setminus \{0\})) = \alpha + \nu(\mathbb{R}^d \setminus \{0\}),$$
$$\nu'(\{0\}) = \bar{\pi}(\mathbb{R}^d \times \{0\}) = \bar{\pi}(\{0\}) + \bar{\pi}((\mathbb{R}^d \setminus \{0\}) \times \{0\}) = \alpha + \mu(\mathbb{R}^d \setminus \{0\})$$

and $\bar{\pi}(\{0\}) = \alpha$. Conversely, if we freely choose $\alpha \geq 0$, and fix $\mu'(\{0\})$, $\nu'(\{0\})$ as above, a simple computation proves that $\bar{\pi} \in \Pi(\mu', \nu')$.

In particular, for $A \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$, we have $\bar{\pi}(A \times \mathbb{R}^d) = \mu(A)$ and $\bar{\pi}(\mathbb{R}^d \times A) = \nu(A)$ which doesn't depend on the choice of the elements in the equivalence classes.

4. *For $\gamma \in \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$, one should notice that for $i = 1, 2$, the measure $\phi_i(\tilde{\gamma})$ cannot in general be written $\tilde{\mu}$ for some $\mu \in \mathrm{M}_0(\mathbb{R}^d)$ since γ may put mass on $\{0\} \times \mathbb{R}^d$, but we can always write $\phi_i(\tilde{\gamma}) \in \bar{\mu}$ for some $\mu \in \mathrm{M}_0(\mathbb{R}^d)$.*

Let P, Q be probability measures on $\mathbb{R}^d$. The Kantorovich Problem consists in minimizing, for π in $\Pi(P, Q)$, the quantity

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \, \mathrm{d}\pi(x, y) \in \mathbb{R}.$$

The Kantorovich Problem can also be defined for measures μ, ν in $\mathrm{M}^+(\mathbb{R}^d)$ with equal finite mass $m \in \mathbb{R}_+ \setminus \{0\}$. One just needs to consider $P = \mu/m$ and $Q = \nu/m$ and multiply every $\pi \in \Pi(P, Q)$ by m to get the collection $\Pi(\mu, \nu)$. This simple scaling argument will allow us to use results proved for probability measures for measures with equal finite nonzero mass.

A sufficient condition for this problem to admit solutions is that μ and ν have finite second order moments. In this special case, an optimal transport plan π turns out to be a joint distribution which maximizes $\int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle \, \mathrm{d}\pi(x, y)$, i.e., the sum of the product moments $\mathsf{E}[X_j Y_j]$, for $j \in \{1, \ldots, d\}$, if $(X, Y) \sim \pi$. In dimension $d = 1$, this means that the joint distribution π maximizes the covariance between X and Y. When we have $X = \mathrm{T}(Y)$ where T is the optimal transport map, it is the relation which in some sense maximizes the dependance between the variables. This gives some motivation to use it in statistics.

We now introduce the relation between optimal transport and cyclical monotonicity—the latter was defined in Subsection 1.4.

Remark 1.43. *Let $\mathcal{S}_n$ denote the set of permutations of $\{1, 2, \ldots, n\}$. Condition* (2) *in Definition 1.35 is equivalent to*

$$\sum_{i=1}^{n} |x_i - y_i|^2 \leq \min_{\sigma \in \mathcal{S}_n} \sum_{i=1}^{n} |x_i - y_{\sigma(i)}|^2 \tag{3}$$

Therefore, if T has finite cardinal and only contains pairs, T seen as a function $\mathbb{R}^d \to \mathbb{R}^d$ is a solution to the optimal transport problem between $\lambda \sum_{x \in \operatorname{dom} T} \delta_x$ and $\lambda \sum_{y \in \operatorname{rge} T} \delta_y$ for every $\lambda > 0$.

Let $\mathrm{M}_{0,cm}$ denote the family of measures μ in M_0 whose support $\operatorname{spt} \mu = \operatorname{spt} \tilde{\mu}$ is cyclically monotone. The following theorem due to McCann (1995) gives mild assumptions on the probability measure P under which there always exists a closed convex function ψ such that $T = \nabla \psi$ pushes P forward to Q.

Theorem 1.44 (Main Theorem in McCann(1995)). *Let $P, Q \in \mathcal{P}(\mathbb{R}^d)$. There exists a coupling measure $\pi \in \Pi(P, Q)$ with cyclically monotone support contained in the graph of the subdifferential of a closed convex function ψ on $\mathbb{R}^d$. If P vanishes on all sets of Hausdorff dimension at most $d-1$, then ψ is differentiable P-almost everywhere, $\nabla\psi_{\#}P = Q$, and the map $\nabla\psi$ is uniquely determined P-almost everywhere.*

Thanks to Knott and Smith (1984), such a T is the solution to the Monge Problem if both distributions have finite second order moments.

Theorem 1.45 (Knott and Smith (1984)). *Let $P, Q \in \mathcal{P}(\mathbb{R}^d)$. Assume that $\iint |x - y|^2 P \otimes Q(dx, dy)$ is finite, then a sufficient condition for $T : \mathbb{R}^d \mapsto \mathbb{R}^d$ to be an optimal transport map is that (a) there exists some convex function ψ satisfying $T = \nabla\psi$ and (b) $T_{\#}P = Q$.*

Remark 1.46. *A direct consequence of the last two theorems is that when μ vanishes on all sets of Hausdorff dimension at most $d-1$, we just need to find one $\nabla\psi$ - with ψ convex- pushing P to Q to get the unique possible optimal transport map.*

Remark 1.47. *This result can be straightforwardly extended to measures $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d)$ with equal finite mass $m > 0$ by an elementary scaling argument. One just needs to consider $\mu/m, \nu/m \in \mathcal{P}(\mathbb{R}^d)$ and finally multiply the restriction γ of $\Pi_{cm}(\mu/m, \nu/m)$ by m since for every $A \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$ that is bounded away from zero, $A \times \mathbb{R}^d$ is bounded away from zero too and we can write $m\gamma(A \times \mathbb{R}^d) = m\mu(A \times \mathbb{R}^d)/m = \mu(A \times \mathbb{R}^d)$ and likewise $m\gamma(\mathbb{R}^d \times A) = m\nu(A)/m = \nu(A)$.*

This means that for measures $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d)$ with finite equal mass, under mild assumptions on μ, there always exists a map T which is the gradient of a convex function such that

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \mathrm{d}\pi(x, y) = \int_{\mathbb{R}^d} |x - T(x)|^2 \,\mathrm{d}\mu(x) \in \mathbb{R} \cup \{\infty\}.$$

Remark 1.48. *Whenever a solution to the initial Monge–Kantorovich problem exists, it is the one given by McCann's result relying on geometric ideas. In this sense, when P vanishes on all sets of Hausdorff dimension at most $d-1$, coupling measures with cyclically monotone support contained in the graph of the subdifferential of a closed convex function can be seen in some sense as a generalization of optimal transport.*

Since the results by McCann require very few assumptions, and provide some kind of generalization of optimal transport, it has naturally attracted statisticians who where eager to get results useful when dealing with real data without making strong assumptions like existence of second-order moments.

The following counter-example shows that the Main Theorem in McCann (1995) cannot be extended straightforwardly to measures in $\mathrm{M}_0(\mathbb{R}^d)$ with infinite mass. It is taken from an unpublished note by Johan Segers.

Example 1.49. *Let $\mu, \nu \in \mathrm{M}_0(\mathbb{R})$ and suppose that μ has no atoms and that $\tilde{\mu}(\mathbb{R}_+) = \tilde{\nu}(\mathbb{R}_-) = \infty$ while $\tilde{\mu}(\mathbb{R}_-) = \tilde{\nu}(\mathbb{R}_+) = 0$. We claim that the conclusions of Theorem 1.44 do not hold. Indeed, if we suppose that there exists some T monotone (non-decreasing) pushing $\tilde{\mu}$ to $\tilde{\nu}$, then there is a contradiction. Let $x > 0$. If $y = T(x) < 0$, then, since $\tilde{\mu}(\mathbb{R}_+) = \infty$, we would have,*

$$\tilde{\nu}((-\infty, y]) = \tilde{\mu}(\{z \in \mathbb{R} : T(z) \leq y\}) \geq \tilde{\mu}((-\infty, x]) = \infty$$

in contradiction to the assumption that $\nu \in \mathrm{M}_0(\mathbb{R})$. Hence $T(x) \geq 0$ for all $x > 0$. But then we have a contradiction since

$$\infty = \tilde{\nu}(\mathbb{R}_-) = \tilde{\nu}(\{z \in \mathbb{R} : T(z) < 0\}) \leq \tilde{\mu}(\mathbb{R}_-) = 0.$$

Even Theorem 6 in McCann (1995) giving the existence of a cyclically monotone coupling cannot be extended to measures in $\mathrm{M}_0(\mathbb{R}^d)$ if we impose the choice of the elements in $\bar{\mu}, \bar{\nu}$ to be $\tilde{\mu}, \tilde{\nu}$ for example.

Example 1.50. *In Example 1.49, there is no cyclically monotone coupling between $\tilde{\mu}$ and $\tilde{\nu}$. Indeed, for $\tilde{\pi} \in \Pi(\tilde{\mu}, \tilde{\nu})$, we let $T = \mathrm{spt}(\tilde{\pi}) \in \mathcal{F}(\mathbb{R} \times \mathbb{R})$ seen as a multivalued map $\mathbb{R} \rightrightarrows \mathbb{R}$. We know (see Remark 1.42) that $\tilde{\pi}$ puts no mass on $\{0\} \times \mathbb{R}$ and $\mathbb{R} \times \{0\}$. We must have $T(\mathbb{R}_+ \setminus \{0\}) \subset \mathbb{R}_- \setminus \{0\}$ otherwise $\tilde{\nu}(\mathbb{R}_+) = 0$ would be violated. Let $x > 0$ and suppose we have $y < 0$ satisfying $y \in T(x)$. We have*

$$\tilde{\nu}((-\infty, y]) = \tilde{\pi}(\mathbb{R} \times (-\infty, y]) \geq \tilde{\pi}(T^{-1}((-\infty, y]) \times (-\infty, y]).$$

Since T is monotone, it is clear that $(-\infty, x) \subset T^{-1}((-\infty, y])$ and that $T((-\infty, x)) \subset (-\infty, y]$. Whence

$$\tilde{\nu}((-\infty, y]) \geq \tilde{\pi}((-\infty, x) \times \mathbb{R}) = \tilde{\mu}((-\infty, x)) = \tilde{\mu}((-\infty, x]) = \infty$$

where we used that $\tilde{\mu}$ has no atoms, in contradiction to the assumption that $\nu \in \mathrm{M}_0(\mathbb{R})$. We just proved that there is no coupling between $\tilde{\mu}$ and $\tilde{\nu}$ with cyclically monotone support.

However, there may exist some $\mu' \in \bar{\mu}$ and $\nu' \in \bar{\nu}$ such that $\Pi_{cm}(\mu', \nu')$ is not empty.

Example 1.51 (Continuation of Example 1.50)**.** *If we consider*

$$\psi(x) = \begin{cases} +\infty & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

then $\partial\psi = (\{0\} \times \mathbb{R}_-) \cup (\mathbb{R}_+ \times \{0\})$ *is cyclically monotone and we can easily find a coupling* $\bar{\pi}$ *between some* $\mu' \in \bar{\mu}$ *and* $\nu' \in \bar{\nu}$ *satisfying* $\operatorname{spt} \bar{\pi} \subset \partial\psi$ *using Remark 1.42(3).*

One should keep in mind that even if there are no cyclically monotone couplings, there may still be couplings. Imposing cyclic monotonicity on couplings is a real constraint.

Example 1.52 (Continuation of Example 1.50)**.** *In Example 1.50, if we further assume that* μ *and* ν *are symmetric in the sense that for every* $A \in \mathcal{B}(\mathbb{R}_+ \setminus \{0\})$ *we have* $\mu(A) = \nu(-A)$*, then there exists a simple coupling between* μ *and* ν*. One can consider the map* $T : x \in \mathbb{R} \mapsto -x \in \mathbb{R}$ *and remark that* $\pi = [\mathrm{Id} \otimes T]_{\#}\mu$ *belongs to* $\Pi(\tilde{\mu}, \tilde{\nu})$ *but obviously not to* $\Pi_{cm}(\tilde{\mu}, \tilde{\nu})$ *since* T *is non-increasing.*

To study heavy-tailed laws, one needs to deal with measures in M_0 with infinite mass for which the Kantorovich Problem would have solutions only under very strong assumptions (second order moment existence). The heuristic is then to only look for conditions sufficient for some kind of cyclically monotone coupling to exist. However, it is clear from Example 1.50 that there may be no cyclically monotone coupling between $\tilde{\mu}, \tilde{\nu} \in \mathrm{M}_0(\mathbb{R}^d)$, i.e., that the assumption that the coupling puts no mass on the axis is too strong. One should therefore look for cyclically monotone couplings between μ' and ν' for some $(\mu', \nu') \in (\bar{\mu} \times \bar{\nu})$. Unfortunately, finding the values $\bar{\mu}(\{0\})$, $\bar{\nu}(\{0\})$ and $\bar{\gamma}(\{0\})$ such that some cyclically monotone coupling exists and if such values does exist is tricky. To overcome this difficulty, Johan Segers introduced zero-couplings, getting inspiration from Guillen et al. (2019) and Catalano et al. (2021), instead of coupling in a talk he gave at the Joint Statistical Meeting 2023 about his on-going work in collaboration with Cees de Valk. Section 2 will be devoted to the study of some of their properties.

1.6 Center-outward quantile

The definition of ranks, quantiles and distribution functions in $\mathbb{R}^d$ with good properties has been a long-standing open problem in statistics. What are good properties is itself an open question but for sure the new object should at least match the usual one in the special cases where it is properly defined.

Building on Monge–Kantorovitch depth presented in Chernozhukov et al. (2017), Hallin et al. (2021) introduced center-outward quantiles and distribution functions relying on optimal transportation of a distribution of interest Q toward a reference distribution P they choose to be the spherical uniform distribution U on the unit ball $\mathrm{B}_{0,1}$ defined by $U = \mathrm{Unif}(0,1) \times \mathrm{Unif}(\mathbb{S}^{d-1})$. They reserved the symbols $\mathbf{Q}_\pm$ and $\mathbf{F}_\pm$ for center-outward quantiles and distribution functions respectively and also proposed center-outwards ranks $|\mathbf{F}_\pm|$ and signs $\mathbf{F}_\pm/|\mathbf{F}_\pm|$ relying on their newly introduced center-outward distribution function. Under suitable assumptions, there exist a U-almost everywhere unique optimal transport plan $\mathbf{Q}_\pm$ called center-outward quantile function pushing P to Q and a Q-almost everywhere unique optimal transport plan $\mathbf{F}_\pm$ called center-outward distribution function pushing Q to P. The Monge–Kantorovitch depth region of probability τ for Q was introduced in Chernozhukov et al. (2017) as $K_Q^{MK}(\tau) = \mathbf{Q}_\pm(K_P^{Tukey}(\tau))$ where $K_P^{Tukey}(\tau)$ is the half-space depth region of probability τ for P. When the reference distribution is $P = U$, Hallin et al. (2021) call the last quantity center-outward quantile of order τ and write $\mathbb{C}(\tau)$. Moreover, they proved that the new object enjoys properties that make them good candidates for multivariate quantiles.

First, the new objects are linked with the usual notions in well understood simple cases. In dimension 1, according to Proposition 2.1 and 2.2 in Hallin et al. (2021), we have $\mathbf{F}_\pm = 2F - 1$ where F is the traditional distribution function, $\mathbf{Q}_\pm$ coincides with $\inf\{x|F(x) \geq (1+u)/2\}$ for $u \in (-1,1)$ and $\mathbb{C}(\tau) = K_Q^{MK}(\tau)$, $q \in (0,1)$ coincides with

$$[\inf\{x|F(x) \geq (1-q)/2\}, \inf\{x|F(x) \geq (1+q)/2\}] \cap \mathrm{cl}\,\mathrm{spt}(P)$$

which matches the half-space depth region of probability content τ for Q.

Something similar can be stated for measures in the family $\mathcal{P}_{ell}^d$ of elliptical distributions on $\mathbb{R}^d$. Let $\mathrm{P}_{0,I,g}$ denote the distribution whose polar decomposition consists of a radial measure with density g over $\mathbb{R}_+$, and the uniform distribution on the unit sphere. We call the cumulative distribution function G of g the radial distribution function. Note that $X \sim \mathrm{P}_{0,I,g}$ holds iff $\mathbf{F}_\pm(X) \sim \mathrm{U}$ where $\mathbf{F}_\pm : x \mapsto x\, G(|x|)/|x|$ is the center-outward distribution function. Indeed, as G is associated with the density g with respect to Lebesgue, G is continuous and we have $G \circ G^{\leftarrow} = \mathrm{Id}$ (see for example Property 1 on page 44 in Guyader (2023)), hence $G(|X|)$ has uniform distribution over $(0,1)$. We also define $\mathbf{Q}_\pm(X) : x \mapsto x\, G^{\leftarrow}(|x|)/|x|$ the center-outward quantile function which satisfies $\mathbf{Q}_\pm(Y) \sim \mathrm{P}_{0,I,g}$ for $Y \sim U$. It is clear that $\mathbf{F}_\pm \circ \mathbf{Q}_\pm = \mathrm{Id}$, while one can show that $\mathbf{Q}_\pm \circ \mathbf{F}_\pm = \mathrm{Id}$ holds $\mathrm{P}_{0,I,g}$-almost everywhere - it will be proven later in a more general case. Let $\tilde{T}_F(x) = \int_0^x F(s)ds$ and $T_F(x) = \tilde{T}_F(|x|)$ for every non-decreasing map $F : x \in \mathbb{R}^d \to \mathbb{R}^d$. Such a map has derivative $\nabla T_F(x) = xF(|x|)/|x|$. Moreover, as $|\cdot|$ is convex and $\tilde{T}_F$ is non-decreasing and convex—since it has non-decreasing derivatives—, $T_{G^{\leftarrow}}, T_G$ are convex functions and $\mathbf{Q}_\pm = \nabla T_{G^{\leftarrow}}$, $\mathbf{F}_\pm = \nabla T_G$. Applying Theorem 1.44, $\mathbf{Q}_\pm$ and $\mathbf{F}_\pm$ are the optimal transport plans pushing respectively U to $\mathrm{P}_{0,I,g}$ and $\mathrm{P}_{0,I,g}$ to U.

In the spherical setting, a common choice to define quantile regions is the half-space depth for which the depth region $K_Q^{Tukey}(\tau)$ with probability τ is simply the closed ball with radius $G^{\leftarrow}(\tau)$ centered at the origin. Note that $K_P^{Tukey}(\tau) = \mathbf{Q}_\pm(B(0,\tau))$. We say that X has distribution $\mathrm{P}_{\mu,\Sigma,g}$, where μ is a location parameter and Σ is a symmetric positive definite real matrix, iff $Y = \Sigma^{-1/2}(X-\mu)$ has spherical distribution $\mathrm{P}_{0,I,g}$. Let $\mathcal{P}^d_{ell} = \{\ \mathrm{P}_{\mu,\Sigma,g}\}$ denotes the family of all full-rank elliptical distributions over $\mathbb{R}^d$ $(d > 1)$ with radial densities over elliptical support sets. The depth region of probability τ is defined with $Y \sim \mathrm{P}_{0,I,g}$ and sent to the original distribution through the transformation $y \mapsto \Sigma^{1/2}y + \mu$.
Recall that the Monge–Kantorovitch depth region of probability τ was introduced in Chernozhukov et al. (2017) as $K_Q^{MK}(\tau) = \mathbf{Q}_\pm(B(0,\tau))$. The discussion above yields $K_Q^{Tukey}(\tau) = K_Q^{MK}(\tau)$ so the two concepts coincide.

Moreover, under the assumption that $P = U$ and Q has a density with respect to the Lebesgue measure, it has been proved in Proposition 2.5 in Hallin et al. (2021) that center-outward ranks $|\mathbf{F}_\pm|$ and associated order statistics benefit from distribution-freeness and maximal ancillarity. Therefore, center-outward quantiles, ranks and order statistics gives extension of the usual distributional properties of univariate order statistics and ranks that make them successful in semi-parametric estimation in one dimension and pave the way for multivariate counterparts. This good properties of center-outward quantiles will be out of the scope of this text which.

Here we briefly introduce $\mathbf{Q}_\pm$ and $\mathbf{F}_\pm$ for our purpose and refer to Hallin et al. (2021) for details. Recall that the spherical uniform distribution can be written in polar decomposition as $U = \mathrm{Unif}(0,1) \otimes \mathrm{Unif}(\mathbb{S}^{d-1})$. Since U vanishes on sets of Hausdorff dimension at most $d-1$, Theorem 1.44 yields the existence of a Borel set A of $\mathbb{R}^d$ and a closed convex function ψ such that $\nabla\psi$ is defined on A, $U(A^c) = 0$ and $\nabla\psi_\# U = Q$. If Q has finite second order moments—it is clear that U does—then we can apply Theorem 1.45 to see that $\nabla\psi$ is the U-almost everywhere unique optimal transport map pushing U forward to Q.

Remark 1.53. *In the rest of this section, one can replace U by any other probability measure P whose polar decomposition involves the uniform distribution on the unit sphere and which vanishes on sets of Hausdorff dimension at most $d-1$, and $\mathbb{C}_{U:Q}(q)$, $\mathcal{C}_{U:Q}(q)$ by $\mathbb{C}_{P:Q}(q) = \nabla\psi(F^{\leftarrow}(q)\mathrm{B}_{0,1})$, $\mathcal{C}_{P:Q}(q) = \nabla\psi(F^{\leftarrow}(q)\mathbb{S}^{d-1})$ respectively where F is the cumulative distribution function of $|\cdot|_\# P$, without altering the conclusions, except for Theorem 1.61. In Section 4 we will specialize to the case where Q is* *regularly varying; this remark will allow us to define center-outward quantile regions and contours using a reference distribution which is regularly varying too, and then derive asymptotic results for some rescaling of the center-outward quantile fonction.*

Definition 1.54. *Call center-outward quantile function $\boldsymbol{Q}_\pm$ of $Q \in \mathcal{P}(\mathbb{R}^d)$ the P-a.e. unique element $\nabla\psi$ with ψ a closed convex function given in Theorem 1.44 pushing U forward to Q.*

If we assume that Q vanishes on sets of Hausdorff dimension at most $d-1$, then we can define the center-outward distribution function $\mathbf{F}_\pm$ in a similar way. Let $\phi := \psi^*$ denote the Legendre transform of ψ. Since ϕ is lsc and convex, Proposition 11.3 in Rockafellar and Wets (1998) gives $\partial\psi = (\partial\psi^*)^{-1}$ where both sides of the equality are seen as multivalued maps. Since Q vanishes on sets of Hausdorff dimension at most $d-1$, Theorem 25.5 in Rockafellar (1970) yields the existence of a Borel set B of $\mathbb{R}^d$ such that $\nabla\phi$ is defined on B and $Q(B^c) = 0$. for every x in $A \cap [\nabla\psi]^{-1}(B)$, $\nabla\phi \circ \nabla\psi(x) = x$ and since Q is the push-forward of P by $\nabla\psi$, $U([A \cap [\nabla\psi]^{-1}(B)]^c) = 0$, i.e., $\nabla\phi \circ \nabla\psi = \mathrm{Id}$ P-almost everywhere. As a consequence, for every Borel set E of $\mathbb{R}^d$, $U(E) = U([\nabla\phi \circ \nabla\psi]^{-1}(E))$, whence $U(E) = Q([\nabla\phi]^{-1}(E))$ and finally $U = \nabla\phi_{\#}Q$. Note that we can also show that $\nabla\psi \circ \nabla\phi = \mathrm{Id}$ Q-almost everywhere using similar arguments as it is done in the discussion preceding Definition 2.2 in Hallin et al. (2021). Applying Theorem 1.44, $\nabla\phi$ is the Q-almost everywhere unique candidate optimal transport plan pushing Q forward to U.

Definition 1.55. *Call center-outward distribution function $\boldsymbol{F}_\pm$ of $Q \in \mathcal{P}(\mathbb{R}^d)$ the Q-a.e. unique element $\nabla\phi$ with ϕ a closed convex function given in Theorem 1.44 pushing Q forward to P.*

We now introduce the notion of the greatest interest for our purpose.

Definition 1.56. *For every q lying in $(0,1)$, we define the center-outward quantile region and contour of order q of Q respectively as*

$$\mathbb{C}_{U:Q}(q) = \nabla\psi(q\mathrm{B}_{0,1}) = \{\nabla\psi(x) \,:\, |x| \leq q\}$$

and

$$\mathcal{C}_{U:Q}(q) = \nabla\psi(q\mathbb{S}^{d-1}) = \{\nabla\psi(x) \,:\, |x| = q\}.$$

When no confusion is possible, we will omit to mention the distributions U, Q and simply write $\mathcal{C}(q)$, $\mathbb{C}(q)$.

For these quantities to be relevant, one would naturally expect from quantile regions and contours of order $q \in (0,1)$ to satisfy :

(a) $Q(\mathbb{C}_{U:Q}(q)) = q$, i.e., the quantile region of order q has probability content q

(b) $\mathcal{C}_{U:Q}(q) = \partial\mathbb{C}_{U:Q}(q)$

The following example shows that for these quantities to be relevant, we need to assume that Q vanishes on sets of Hausdorff dimension at most $d-1$, as highlighted by the following example.

Example 1.57. *We define the application ψ on $\mathbb{R}^2$ as*

$$\psi(x,y) = \begin{cases} \langle (x,y), (-1,0)\rangle & \text{if } (x,y) \in \mathbb{R}_- \times \mathbb{R}, \\ \langle (x,y), (1,0)\rangle^2/2 & \text{if } (x,y) \in \mathbb{R}_+ \times \mathbb{R}. \end{cases}$$

The function ψ, whose values only depend on its first argument, is continuous and convex. A simple computation shows that $\nabla\psi$ is defined as soon as $x \neq 0$ and satisfies

$$\nabla\psi(x,y) = \begin{cases} (-1,0) & \text{if } x < 0, \\ (x,0) & \text{if } x > 0, \end{cases}$$

whence $\nabla\psi$ is the gradient of a convex function and we can choose $Q = \nabla\psi_{\#}P$ where P is the spherical uniform distribution. Since Q has finite second order moment, $\nabla\psi$ is the P-almost everywhere unique transport plan pushing P toward Q. Defined this way, Q puts the mass $1/2$ on the singleton $\{(-1,0)\}$ while it puts the mass $1/2$ on the segment $[(0,0),(1,0)]$ following a distribution whose description is useless for our purpose.

Let $\alpha > 0$. Since $[\nabla\psi]^{-1}\mathbb{C}(\alpha) = (\mathrm{B}_{0,1} \cap [\mathbb{R}_- \times \mathbb{R}]) \cup (\mathrm{B}_{0,\alpha} \cap [(0,\alpha)\times\mathbb{R}])$ we can write $Q(\mathbb{C}(\alpha)) = Q(\cup_{0<\beta<\alpha}\mathbb{C}(\beta))$ and $Q(\mathbb{C}(\alpha)) = Q(\cap_{\alpha<\beta}\mathbb{C}(\beta))$ using the sequential limit property of measures. As P vanishes on small sets, Q gives the same mass to both sets. Whence $\xi : \alpha \in (0,1) \mapsto Q(\mathbb{C}(\alpha)) \in [0,1]$ is continuous and increasing - the last information is always true. We further claim that there exists some $\zeta : (0,1) \to [0,1]$ such that $\xi(\alpha) = 1/2 + \zeta(\alpha) + \circ(1)$ as α goes to 1. The first term in the expression of $\xi(\alpha)$ comes from the simple fact $\mathrm{B}_{0,1} \cap \mathbb{R}_- \times \mathbb{R} \subset [\nabla\psi]^{-1}\nabla\psi(\mathrm{B}_{0,\alpha})$ for every positive α. From the figure it is clear that we have $\xi(\alpha) \geq 1/2 + \zeta(\alpha)$ for $\zeta(\alpha) = (1-\alpha)\arcsin(\alpha)/\pi$ and we can check that $\xi(\alpha) - 1/2 - \zeta(\alpha)$ goes to zero as α goes to 1. Note that $\circ(1) > 0$ and $\zeta(\alpha) > 0$ for every α in $(0,1)$.

Let $q \in (0,1)$. We have $\psi(q) > q$ and we can find, thanks to the previous computations, some $\alpha < q$ satisfying

$$q \leq Q(\mathbb{C}(\alpha)) < Q(\mathbb{C}(q))$$

for every $q \in (0,1)$ we can find $\alpha \in (0,q)$ such that $q \leq Q(\mathbb{C}(\alpha)) < Q(\mathbb{C}(q))$ while one would wish the quantile region of order q to be the smallest one of the form $\mathbb{C}(\cdot)$ with Q-probability content at least q.

We give another example in dimension 2 of the same phenomenon for some Q which is atomless such that $\alpha \mapsto Q(\mathbb{C}(\alpha))$ is continuous.

Example 1.58. *We proceed as above by defining the application ψ on $\mathbb{R} \times \mathbb{R}$ as*

$$\psi(x,y) = \begin{cases} (x^2+y^2)/2 & \text{if } x \geq 0 \\ -x + y^2/2 & \text{if } x < 0 \end{cases}.$$

The function ψ is continuous and convex as the sum of two one-dimensional convex functions (one with argument x and the other with argument y in the formula above). A simple computation shows that $\nabla\psi$ is defined as soon as $x \neq 0$ and satisfies

$$\nabla\psi(x,y) = \begin{cases} (x,y) & \text{if } x > 0 \\ (-1,y) & \text{if } x < 0 \end{cases}$$

whence $\nabla\psi$ is the gradient of a convex function and we can choose $Q = \nabla\psi_{\#}P$. Defined this way, Q matches the spherical uniform distribution on $\mathbb{R}_+ \times \mathbb{R}$ while it puts the mass $1/2$ on the segment $[(-1,-1),(-1,1)]$ following a distribution whose description is useless for our purpose.

Let $\alpha > 0$. Since $[\nabla\psi]^{-1}\mathcal{C}(\alpha) = (\mathrm{B}_{0,\alpha} \cap \mathbb{R}_+ \times \mathbb{R}) \cup (\mathrm{B}_{0,\alpha} \cap \mathbb{R} \times (-\alpha,\alpha))$, we can write $Q(\mathcal{C}(\alpha)) = Q(\cup_{0<\beta<\alpha}\mathcal{C}(\beta))$ and $Q(\mathcal{C}(\alpha)) = Q(\cap_{\alpha<\beta}\mathcal{C}(\beta))$, and using that P vanishes on small sets, Q gives the same mass to both sets. Whence $\xi : \alpha \in (0,1) \mapsto Q(\mathcal{C}(\alpha)) \in [0,1]$ is continuous and increasing - the last information is always true. We further claim that there exists some $\zeta : (0,1) \to [0,1]$ such that $\xi(\alpha) = \alpha + \zeta(\alpha) + \circ(1)$ as α goes to 1. The first term comes from the simple fact $\mathrm{B}_{0,\alpha} \subset [\nabla\psi]^{-1}\nabla\psi(\mathrm{B}_{0,\alpha})$ which is always true. From the figure it is clear that we have $\xi(\alpha) \geq \alpha + \zeta(\alpha)$ for $\zeta(\alpha) = (1-\alpha)(1/2 - \arccos(\alpha)/\pi)$ and we can check that $\xi(\alpha) - \alpha - \zeta(\alpha)$ goes to zero as α goes to 1. Note that $\circ(1) > 0$ and $\zeta(\alpha) > 0$ for every α living in $(0,1)$.

Let $q \in (0,1)$. We have $\psi(q) > q$ and we can find, thanks to the previous computations, some $\alpha < q$ satisfying

$$q \leq Q(\mathcal{C}(\alpha)) < Q(\mathcal{C}(q)).$$

As a consequence of the last two examples, when no assumptions are made on the distribution Q the center-outward quantile regions and contours introduced before can be deceptive regarding the expected properties (a) and (b) although contours and regions can still be defined. The potential practitioner should keep this in mind when dealing with the possible center-outward quantile since it is a major limitation.

Theorem 1.59. *Let $Q \in \mathcal{P}(\mathbb{R}^d)$ and $q \in (0,1)$. The center-outward quantile region $\mathbb{C}(q)$ has Q-probability content at least q. If we further assume that Q vanishes on sets of Hausdorff dimension at most $d-1$, then there is equality in the last statement.*

Proof. Since $\mathrm{B}_{0,q} \subset [\nabla\psi]^{-1}(\mathbb{C}_Q(q)$, it is clear that the relation

$$Q(\mathbb{C}_Q(q)) = P([\nabla\phi]^{-1}(\mathbb{C}_Q(q))) \geq q$$

holds. When Q vanishes on sets of Hausdorff dimension at most $d-1$, we have seen that $\nabla\phi \circ \nabla\psi = \mathrm{Id}$ P-almost everywhere so

$$P([\nabla\psi]^{-1}\nabla\psi(\mathrm{B}_{0,q})) = P(\nabla\phi \circ \nabla\psi(\mathrm{B}_{0,q})) = P(\mathrm{B}_{0,q}) = q.$$

As a consequence, equality holds instead of the inequality. □

Remark 1.60. *When the assumptions made on Q does not hold, a similar definition of the center-outward quantile region can be given as follows. Let q live in $(0,1)$. We define $\mathbb{C}_{\text{general}}(q)$, the general center-outward quantile of order q, as*

$$\mathbb{C}_{\text{general}}(q) = \bigcap_{\alpha \in \mathrm{E}} \mathbb{C}(\alpha)$$

where $\mathrm{E} = \{\alpha > 0 \,:\, Q(\mathbb{C}(\alpha)) \geq q\}$.

Since the family $\mathbb{C}(\alpha)$, for $\alpha \in (0,1)$, is nested, we could alternatively set $\mathbb{C}_{\text{general}}(q) = \mathbb{C}(\alpha^*)$ for some α^* lying in $\arg\min\{\alpha \,:\, Q(\mathbb{C}(\alpha)) \geq q\}$. It may be simpler to compute in practice but at the cost of nestedness. When Q vanishes on sets of Hausdorff dimension at most $d-1$, it matches the definition given previously.

Under additional assumptions on the gradient $\nabla\psi$ pushing U to Q, Hallin et al. (2021) showed that center-outward quantile regions and contours enjoy some properties that one would expect from them.

Theorem 1.61 (Corollary 2.1 in Hallin et al. (2021))**.** *Let $P \in \mathcal{P}(\mathbb{R}^d)$ be given by $P = \nabla\Upsilon_{\#}Q$ for some convex function Υ such that (i) $\nabla\Upsilon$ is a homeomorphism from $\mathbb{S}^{d-1} \setminus \{0\}$ to $\nabla\Upsilon(\mathbb{S}^{d-1} \setminus \{0\})$ and (ii) $\nabla\Upsilon(\{0\})$ is a compact convex set of Lebesgue measure zero. For any $q \in [0,1)$, the quantile region $\mathbb{C}(q)$ is closed, connected and nested, with boundary $\mathcal{C}(q)$ of Lebesgue measure 0.*

Remark 1.62. *The nestedness of quantile regions is an immediate consequence of their definition and requires no further assumption.*

The main aim of this text is to extend center-outward quantiles to the limit measures of regularly varying probability measures. In Section 4, we introduce center-outward tail quantiles and discuss their basic properties in a discussion similar to the one above, and we also provide ways to estimate them in practice.

2 Zero-couplings and tails of transport plans

We begin by explaining the rationale for our study. Let $P, Q \in \mathcal{P}(\mathbb{R}^d)$ be regularly varying with index $\alpha > 0$, auxiliary function b and limit measure $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d) \setminus \{0\}$. Let X, Y be random vectors with laws P and Q. In Subsection 1.2 we have seen that the decomposition

$$\mathsf{P}(X \in A) = \mathsf{P}(X \in A \cap \mathrm{cl}\,\mathrm{B}_{0,b(t)}) + \mathsf{P}[|X| > b(t)]\ \mathsf{P}[b(t)^{-1}X \in b(t)^{-1}A \mid |X| > b(t)]$$

is useful to derive simple asymptotic approximations in Extreme Value Theory. Recall from the same subsection that we can choose $b(t) = F^{\leftarrow}(1 - 1/t)$ where $F^{\leftarrow}$ is the generalized inverse of $F(t) = |\cdot|_{\#}Q([t, \infty))$. Assume P vanishes on sets of Hausdorff dimension at most $d-1$. Then relying on Subsection 1.6, we can define $\mathbb{C}(q) = \mathbb{C}_{P:Q}(q)$ the center-outward quantile region of order $q \in (0, 1)$; note that unlike in Hallin et al. (2021) we do not choose P to be the spherical uniform distribution, since the latter is not regularly varying. Letting $q = 1 - 1/t$ for $t > 1$ yields the alternative decomposition

$$\mathsf{P}\left(Y \in A\right) = \mathsf{P}\left(Y \in A \cap \mathbb{C}(1 - 1/t)\right) + \mathsf{P}\left(Y \in \mathbb{C}(1 - 1/t)^c)\right) \mathsf{P}\left(Y \in A \mid Y \in \mathbb{C}(1 - 1/t)^c\right).$$

Assume Q vanishes on sets of Hausdorff dimension at most $d - 1$ too, then

$$t\ \mathsf{P}\left(Y \in \mathbb{C}(1 - 1/t)^c)\right) = t\, P(b(t)\, (\mathbb{R}^d \setminus \mathrm{B}_{0,1})).$$

Therefore $t\ \mathsf{P}\left(Y \in \mathbb{C}(1 - 1/t)^c)\right)$ goes to $\mu(\mathbb{R}^d \setminus \mathrm{B}_{0,1}) = 1$ as $n \to \infty$ if we further assume that μ vanishes on sets of Hausdorff dimension at most $d - 1$ [AM:I am not sure this hypothesis is required (don't think so)]. As a consequence, wanting to build an approximation like in Subsection 1.2 and using arguments from there, we are interested in the M_0-convergence of the quantity

$$t\ \mathsf{P}\left(Y \in \cdot \cap \mathbb{C}(1 - 1/t)^c\right) = t\, Q\left(\cdot \cap \mathbb{C}(1 - 1/t)^c\right) = t\, P\left([\nabla\psi]^{-1}(\cdot) \cap \mathrm{B}^c_{0,b(t)}\right),$$

which can also be written as

$$t\, \pi\left(b(t)\left[\mathbb{R}^d \times (\cdot/b(t) \cap \mathbb{C}(1 - 1/t)^c/b(t))\right]\right)$$

where $\pi \in \Pi(P, Q)$ is the unique cyclically monotone coupling between P and Q. In order to proceed like in Subsection 1.2, we would like that $t\, \pi(b(t)\cdot)$ admits some limit in $\mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$.

Let $P, Q \in \mathcal{P}(\mathbb{R}^d)$ be regularly varying with common indice $\alpha > 0$, auxiliary function b and limit measures $\mu, \nu \in \mathrm{M}^+_0(\mathbb{R}^d) \setminus \{0\}$ respectively. We suppose that

P vanishes on sets of Hausdorff dimension at most $d-1$ and let π denote the unique cyclically monotone coupling between P and Q. The main result of this section—Theorem 2.38—asserts that the family $t \operatorname{res} \pi\,(b(t)\cdot)$ is relatively compact in $\mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$. Moreover, assuming further that (a) the limit measure μ puts no mass on sets of Hausdorff dimension at most $d-1$ and (b) $\operatorname{spt} \mu$ is "large enough" relatively to $\operatorname{spt} \nu$, we get that $t \operatorname{res} \pi\,(b(t)\cdot)$ M_0-converges to some $\gamma \in \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$ characterized by $\gamma = [\mathrm{Id} \otimes \nabla \bar{\psi}]_{\#}\mu$ for some closed convex function $\bar{\psi}$ whose gradient is uniquely determined μ-almost everywhere and satisfies the homogeneity relation $\nabla\bar{\psi}(\lambda x) = \lambda \nabla \bar{\psi}(x)$ for every $\lambda > 0$ for μ-almost every x in $\mathbb{R}^d$. Under the assumptions (a) and (b), Theorem 2.38 also gives the $\mathcal{F}$-convergence of the rescaling $(b(t))^{-1}\partial\psi(b(t)\,\cdot\,)$ of the subdifferential of a closed convex function ψ satisfying $\operatorname{spt}\pi \subset \partial\psi$ toward $\partial\bar{\psi}$ the subdifferential of $\bar{\psi}$.

Note that in the assertion above, we had to consider $t \operatorname{res} \pi\,(b(t)\cdot)$ instead of $t\,\pi\,(b(t)\cdot)$ since the latter is not an element of $\mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$. As a limit of restrictions of cyclically monotone coupling, one could expect γ to be the restriction of a cyclically monotone coupling too. Recall from the discussion at the end of Subsection 1.5 that a coupling between $\tilde{\mu}$ and $\tilde{\nu}$ may not exist. However we know that there exists at least one coupling $\bar{\pi}$ between μ' and ν' where $\mu' \in \bar{\mu}$ and $\nu' \in \bar{\nu}$. As a consequence, we only hope that $\gamma = \operatorname{res}\gamma'$ for some $\gamma' \in \bar{\gamma}$ satisfying $\gamma' \in \pi_{cm}(\mu', \nu')$. Unfortunately, finding the values $\bar{\mu}(\{0\})$, $\bar{\nu}(\{0\})$ and $\bar{\gamma}(\{0\})$ such that some cyclically monotone coupling exists and if such values do exist is tricky. To overcome this difficulty, Johan Segers introduced zero-couplings, getting inspiration from Guillen et al. (2019) and Catalano et al. (2021), instead of coupling, in a talk—Segers (2023)—he gave at the Joint Statistical Meeting 2023 about his on-going work in collaboration with Cees de Valk.

To begin with, in Subsection 2.1 we introduce the family $\Gamma_{0,(cm)}(\mu, \nu)$ of (cyclically monotone) zero-couplings and prove some basic properties while emphasizing the differences and similarities with the usual couplings introduced in Subsection 1.5. We also define the sub-family $\Gamma_h(\mu, \nu)$ of h-couplings that will be proved later to contain all the zero-couplings we are interested in and exhibit additional properties.

In Subsection 2.2, we extend some results for couplings taken from Segers (2022) to the zero-coupling framework. These results are applied to study the convergence of a sequence $\gamma_n, n \geq 1$ of zero-couplings between measures $\mu_n, \nu_n, n \geq 1$ in $\mathrm{M}_0(\mathbb{R}^d)$ that are M_0-converging to μ, ν respectively. Then the later latter result is extended to the case where $\mu_n, \nu_n, n \geq 1$ are random elements in $\mathrm{M}_0(\mathbb{R}^d)$ that are weakly M_0-converging to μ, ν respectively—μ, ν are still deterministic. In both cases, the convergence of their possibly random support $\operatorname{spt}\gamma_n$ is also discussed in the Fell topology introduced in Subsection 1.3.

In Subsection 2.3, we turn to cyclically monotone zero-couplings and finally prove the main theorem of the section. First, taking ideas from de Valk and Segers (2018) and building on McCann (1995) as well as on results on sequences of zero-couplings from Subsection 2.2, we prove that there is at least one cyclically monotone zero-coupling γ between μ and ν. Moreover, assuming γ is also a h-coupling, we prove that $\gamma = [\mathrm{Id} \otimes \nabla\bar{\psi}]_{\#}\mu$ for every closed convex function $\bar{\psi}$ satisfying $\operatorname{spt}\gamma \subset \partial\bar{\psi}$. Moreover, the gradient of $\bar{\psi}$ is uniquely determined μ-almost everywhere and satisfies $\nu = \operatorname{res}(\nabla\bar{\psi}_{\#}\mu)$. Then we prove that there is at most one cyclically monotone h-coupling and provide a simple criterion on $\operatorname{spt}\mu$ for every cyclically monotone zero-coupling to match the first requirement of being a h-coupling. Finally, we prove that the family $t \operatorname{res}\pi\left(b(t)\cdot\right), t > 0$ is relatively compact and that any limit point is a cyclically monotone zero-coupling matching the second requirement of being a h-coupling. As a consequence, assuming that the criterion on $\operatorname{spt}\mu$ is satisfied, all limit points are the same, hence the convergence of $t \operatorname{res}\pi\left(b(t)\cdot\right), t > 0$ along the whole sequence toward the single h-coupling is proved. As in subsection 2.2.1 we also investigate the convergence of the subdifferential $\partial\bar{\psi}$ for $\bar{\psi}$ defined above.

2.1 General definitions and basic properties

Recall from Subsection 1.5 the definition of the maps ϕ_1, ϕ_2 which send a finite, nonnegative Radon measure $\mu \in \mathrm{M}^+(\mathbb{R}^d \times \mathbb{R}^d)$ to its left and right marginals $\operatorname{proj}_{1\#}\mu$, $\operatorname{proj}_{2\#}\mu$ in $\mathrm{M}^+(\mathbb{R}^d)$. Since we are interested here in measures in $\mathrm{M}_0(\mathbb{R}^d)$, we would rather consider $\operatorname{res}\operatorname{proj}_{1\#}\mu$ and $\operatorname{res}\operatorname{proj}_{2\#}\mu$, with res defined in Eq. (1).

Definition 2.1 (Zero-marginals)**.** *Let $\gamma \in \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$. We call $\operatorname{res}\operatorname{proj}_{1\#}\tilde{\gamma}$ and $\operatorname{res}\operatorname{proj}_{2\#}\tilde{\gamma}$ the left and right zero-marginals of μ, and introduce the map $\varphi_i : \gamma \in \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d) \mapsto (\operatorname{proj}_{i\#}\tilde{\gamma})|_{\mathbb{R}^d\setminus\{0\}} \in \mathrm{M}_0(\mathbb{R}^d)$ for $i = 1, 2$.*

The following lemma is similar to Lemma 1.40, with zero-marginals instead of marginals and another space of measures.

Lemma 2.2. *The maps $\varphi_i, i \in \{1, 2\}$ are continuous.*

Remark 2.3. *From the definitions above, it is clear that φ_i is related to ϕ_i by the simple expression $\varphi(\cdot) = \operatorname{res}(\phi_i(\tilde{\cdot}))$*

Definition 2.4 (Zero-couplings)**.** *Let $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d)$. The set $\Gamma_0(\mu, \nu)$ of zero-couplings between μ and ν consists of all nonnegative Borel measures γ on $(\mathbb{R}^d \times \mathbb{R}^d) \setminus \{0\}$ such that for each $A \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$, we have*

$$\gamma(A \times \mathbb{R}^d) = \mu(A) \quad \textit{and} \quad \gamma(\mathbb{R}^d \times A) = \nu(A).$$

In other words, $\gamma \in \mathrm{M}^+((\mathbb{R}^d \times \mathbb{R}^d) \setminus \{0\})$ *lies in* $\Gamma_0(\mu,\nu)$ *if and only if*

$$\varphi_1(\gamma) = \mu \quad \text{and} \quad \varphi_2(\gamma) = \nu.$$

We will let $\Gamma_{0,cm}(\mu,\nu)$ *denote the set of zero-couplings between* μ *and* ν *whose support is cyclically monotone.*

In the rest of the text, we choose to use γ for a zero-coupling and π for a usual coupling to emphasize the difference in their very nature. Note that in the definition above, we used $\varphi_i(\gamma), i \in \{1,2\}$—whose definition involves $\tilde{\gamma}$—for a zero-coupling γ lying in the space $\mathrm{M}^+((\mathbb{R}^d\times\mathbb{R}^d)\setminus\{0\})$ of non-negative Borel measures on $(\mathbb{R}^d\times\mathbb{R}^d)\setminus\{0\}$. The following lemma ensures that γ lies in $\mathrm{M}_0(\mathbb{R}^d\times\mathbb{R}^d)$, hence $\tilde{\gamma}$, $\bar{\gamma}$ are defined without ambiguity.

Lemma 2.5. *For every* μ,ν *in* $\mathrm{M}_0(\mathbb{R}^d)$, $\Gamma_0(\mu,\nu)$ *is a subset of* $\mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$.

Remark 2.6. *We bring together here a few observations on the newly introduced object.*

- *Notice that if* $\gamma \in \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$, *then* $\gamma \in \Gamma_0(\varphi_1(\gamma),\varphi_2(\gamma))$. *The map* φ_i *plays for zero-couplings a role similar to the one played by* ϕ_i *for couplings.*

- *A zero-coupling between two measures* μ *and* ν *in* $\mathrm{M}_0(\mathbb{R}^d)$ *always exists. One just needs to consider* $\gamma = \mu \otimes \delta_0 + \delta_0 \otimes \nu$. *Since* γ *puts mass on the axes* $\{0\} \times \mathbb{R}^d$ *and* $\mathbb{R}^d \times \{0\}$, *the measure* $\tilde{\gamma}$ *is no coupling between* $\tilde{\mu}$ *and* $\tilde{\nu}$ *(see Remark 1.42).*

- *It is easy to see that* $\mathrm{res}(\Pi(\tilde{\mu},\tilde{\nu}))$ *is a subset of* $\Gamma_0(\mu,\nu)$ *(we refer to the proof of Lemma 2.7 below), but the example given in the previous point shows that the converse is not true.*

- *However, we can always find some* $(\mu',\nu') \in \bar{\mu}\times\bar{\nu}$ *such that* $\tilde{\gamma}$*(or* $\bar{\gamma}$*)* $\in \Pi(\mu',\nu')$. *We call this a representation of* γ. *Using the same idea, for* $\gamma \in \Gamma_0(\mu,\nu)$, *we have* $\phi_1\overline{\gamma} = \overline{\mu}$ *and* $\phi_2\overline{\gamma} = \overline{\nu}$ *when seen as families of elements.*

The following lemma emphasizes that zero-couplings are a natural way to proceed when one wants to avoid the problem of finding, if they exist, values $\bar{\mu}(\{0\})$ and $\bar{\nu}(\{0\})$ such that $\Pi_{cm}(\bar{\mu},\bar{\nu})$ is not empty.

Lemma 2.7 (Representation)**.** *For every* $\mu,\nu \in \mathrm{M}_0(\mathbb{R}^d)$ *we have*

$$\{\overline{\gamma} : \gamma \in \Gamma_0(\mu,\nu)\} = \{\pi \in \Pi(\alpha,\beta) : \alpha \in \overline{\mu}, \beta \in \overline{\nu}\}.$$

For $\alpha \in \overline{\mu}$ *and* $\beta \in \overline{\nu}$ *satisfying* $\alpha(\{0\}) = \gamma(\{0\} \times (\mathbb{R}^d \setminus \{0\}))$ *and* $\beta(\{0\}) = \gamma((\mathbb{R}^d \setminus \{0\}) \times \{0\})$, *we have* $\tilde{\gamma} \in \Pi(\alpha,\beta)$.

Remark 2.8. *When dealing with cyclically monotone zero-couplings which are not necessarily optimal transport plans, the interpretation in terms of mass transportation doesn't stand any more. Consider two measures* μ, ν *in* $\mathrm{M}_0(\mathbb{R})$ *respectively supported in* $\mathbb{R}_-$ *and* $\mathbb{R}_+$. *Assume* $\mu(\mathbb{R}_- \setminus \{0\}) = \infty$ *while* $\nu(\mathbb{R}_+ \setminus \{0\}) = 1$, *then* γ *defined by* $\gamma = \mu \otimes \delta_0 + \delta_0 \otimes \nu$ *is a cyclically monotone zero-coupling between* μ *and* ν *while* $\mu(\mathbb{R} \setminus \{0\})$ *and* $\nu(\mathbb{R} \setminus \{0\})$ *differ.*

We give two simple examples which illustrate the importance of letting the zero-coupling send mass to the origin but also take mass from the origin.

In this first example, there is too little mass put by μ on the right-hand side of the origin, so the missing mass is taken from the origin which is not part of the punctured space.

Example 2.9 (Mass taken from zero). *Let* $\mu, \nu \in \mathrm{M}_0(\mathbb{R})$ *and suppose that* μ *has no atoms and that* $\tilde{\mu}(\mathbb{R}_-) = \tilde{\nu}(\mathbb{R}_-) = \infty$ *while* $\tilde{\mu}(\mathbb{R}_+) = 0$ *but* $\tilde{\nu}(\mathbb{R}_+) = \infty$. *We claim that there is no zero-coupling* γ *between* μ *and* ν *with cyclically monotone support such that* $\gamma(\{0\} \times (\mathbb{R}_+ \setminus \{0\})) = 0$.

Assume that there exists such a zero-coupling γ, *and let* $T = \operatorname{spt} \gamma$, *seen as a multivalued map* $\mathbb{R} \rightrightarrows \mathbb{R}$. *We will show that it implies a contradiction. Let* $(x, y), (a, b) \in T$. *Since* T *is monotone,* $a \geq x \implies T(a) \ni b \geq y \in T(x)$. *Therefore, if* $T(a) \cap \mathbb{R}_+ \neq \varnothing$ *for some* $a \in \mathbb{R}$, *then for every* $x > a$ *we have* $T(x) \subset \mathbb{R}_+$. *As a consequence, there exists some* $a \in \mathbb{R}$ *such that* $T((-\infty, a)) \subset \mathbb{R}_- \setminus \{0\}$, *otherwise* $\tilde{\nu}(\mathbb{R}_-) = \infty$ *would not hold, and* $x = \sup\{a \in \mathbb{R} \,:\, T((-\infty, a)) \subset \mathbb{R}_- \setminus \{0\}\}$ *exists. As* $\tilde{\nu}(\mathbb{R}_+) > 0$, *the inequality* $x < \infty$ *is necessarily satisfied. Since* $T = \operatorname{spt} \gamma$ *we have*

$$\operatorname{spt} \gamma \cap [\mathbb{R} \times (\mathbb{R}_- \setminus \{0\})] = T^{-1}(\mathbb{R}_- \setminus \{0\}) \times (\mathbb{R}_- \setminus \{0\}).$$

By construction, it is clear that $T^{-1}(\mathbb{R}_- \setminus \{0\}) \subset (-\infty, x]$. *Thus we can write*

$$\begin{aligned} \infty &= \nu(\mathbb{R}_- \setminus \{0\}) \\ &= \gamma(\mathbb{R} \times (\mathbb{R}_- \setminus \{0\})) \\ &= \gamma(T^{-1}(\mathbb{R}_- \setminus \{0\}) \times (\mathbb{R}_- \setminus \{0\})) \\ &\leq \gamma((-\infty, x] \times (\mathbb{R}_- \setminus \{0\})), \end{aligned}$$

whence x *is necessarily nonnegative, otherwise* γ *would not belong to* $\mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$. *Likewise, using* $\tilde{\nu}(\mathbb{R}_+) = \infty$ *one can show* $\gamma(([x, +\infty) \times (\mathbb{R}_+ \setminus \{0\})) = \infty$, *whence* $x = 0$. *To conclude, note that* $\tilde{\mu}(\mathbb{R}_+) = 0$ *implies* $\operatorname{spt} \gamma \cap (\mathbb{R}_+ \setminus \{0\} \times \mathbb{R}) = \varnothing$. *Thus we have the equality* $\infty = \gamma(([0, +\infty) \times (\mathbb{R}_+ \setminus \{0\})) = \gamma(\{0\} \times (\mathbb{R}_+ \setminus \{0\}))$ *which is in contradiction with our assumption.*

In the example above, one should notice that since $\gamma(\{0\} \times (\mathbb{R} \setminus \{0\})) \neq 0$, it is impossible to find closed convex function ψ such that $\gamma = [\mathrm{Id} \otimes \nabla\psi]_{\#}\mu$ because this γ would send no mass on $\{0\} \times (\mathbb{R} \setminus \{0\})$.

In this second example, there is too much mass put by μ (relative to the one put by ν) on the right side of the origin, so the "excess" of mass is sent to the origin out of the punctured space.

Example 2.10 (Mass sent to zero)**.** *Let $\mu, \nu \in \mathrm{M}_0(\mathbb{R})$ and suppose that μ has no atoms and that $\tilde{\mu}(\mathbb{R}_-) = \tilde{\nu}(\mathbb{R}_-) = \infty$ while $\tilde{\mu}(\mathbb{R}_+) = \infty$ but $\tilde{\nu}(\mathbb{R}_+) = 0$. We claim that there is no zero-coupling γ between μ and ν such that $\gamma\left((\mathbb{R} \setminus \{0\} \times \{0\}\right) = 0$. Again, let $T = \operatorname{spt}\gamma$ seen as a multivalued map $\mathbb{R} \rightrightarrows \mathbb{R}$. Using arguments similar to those in the previous example, we get a contradiction otherwise.*

We give the details here. Assume that there exists such a zero-coupling γ, and let $T = \operatorname{spt}\gamma$, seen as a multivalued map $\mathbb{R} \rightrightarrows \mathbb{R}$. Let $(x, y), (a, b) \in T$. Since T is monotone, $a \geq x \implies T(a) \ni b \geq y \in T(x)$. Therefore, if $T(a) \cap \mathbb{R}_+ \neq \varnothing$ for some $a \in \mathbb{R}$, then for every $x > a$ we have $T(x) \subset \mathbb{R}_+$. As a consequence, there exists some $a \in \mathbb{R}$ such that $T\left((-\infty, a)\right) \subset \mathbb{R}_- \setminus \{0\}$, otherwise $\tilde{\nu}(\mathbb{R}_-) = \infty$ would not hold, and $x = \sup\{a \in \mathbb{R} \ : \ T\left((-\infty, a)\right) \subset \mathbb{R}_- \setminus \{0\}\}$ exists. Contrary to Example 2.9, it is not clear at this point that $x < \infty$ holds. Since $T = \operatorname{spt}\gamma$ we have

$$\operatorname{spt}\gamma \cap [\mathbb{R} \times (\mathbb{R}_- \setminus \{0\})] = T^{-1}(\mathbb{R}_- \setminus \{0\}) \times (\mathbb{R}_- \setminus \{0\}).$$

By construction, it is clear that $T^{-1}(\mathbb{R}_- \setminus \{0\}) \subset (-\infty, x]$. Thus we can write

$$\begin{aligned}
\infty &= \nu(\mathbb{R}_- \setminus \{0\}) \\
&= \gamma(\mathbb{R} \times (\mathbb{R}_- \setminus \{0\})) \\
&= \gamma(T^{-1}(\mathbb{R}_- \setminus \{0\}) \times (\mathbb{R}_- \setminus \{0\})) \\
&\leq \gamma((-\infty, x] \times (\mathbb{R}_- \setminus \{0\})),
\end{aligned}$$

whence x is necessarily nonnegative, otherwise γ would not belong to $\mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$. On the other hand, if $x > 0$, we have $a > 0$ such that $T\left((-\infty, a]\right) \subset (-\infty, -\epsilon]$ for some $\epsilon > 0$. This yields $\tilde{\nu}\left((-\infty, -\epsilon]\right) = \infty$ which is absurd, hence $x = 0$. Moreover, $T(\mathbb{R}_+) = \{0\}$ otherwise $\tilde{\nu}(\mathbb{R}_+) = 0$ would not hold. As a consequence, we have the inequality $\gamma\left((\mathbb{R} \setminus \{0\} \times \{0\}\right) \geq \tilde{\mu}(\mathbb{R}_+) = \infty$ which is in contradiction with our assumption.

As opposed to the previous example, it is not impossible that we find a closed convex function ψ such that $\gamma = [\mathrm{Id} \otimes \nabla\psi]_{\#}\mu$. In fact, a cyclically monotone zero-coupling γ admits such an expression under some hypothesis (see Remark 2.28).

Example 2.11 (Continuation of Example 2.10). *In the last example, if we also suppose that* $\tilde{\nu}|_{\mathbb{R}_-} = \tilde{\mu}|_{\mathbb{R}_-}$, *then we can write* $\operatorname{spt}\gamma = \{(x,x) \in (\mathbb{R})^2 : x \in \mathbb{R} \setminus \{0\}\} \cup \{(x,0) \in (\mathbb{R})^2 : x \in \mathbb{R}_+\}$. *We can consider the closed convex function* $\psi : \mathbb{R} \mapsto \mathbb{R}$ *defined by*

$$\psi(x) = \begin{cases} x^2/2 & \text{if } x \leq 0, \\ 0 & \text{if } x > 0, \end{cases}$$

for which we can easily check that

$$\nabla\psi(x) = \begin{cases} 0 & \text{if } x \geq 0, \\ x & \text{if } x < 0. \end{cases}$$

Therefore we have $\operatorname{spt}\gamma = \partial\psi$ *and we get* $\gamma = [\mathrm{Id} \otimes \nabla\psi]_{\#}\mu$ *through a simple computation. In particular, we have* $\nu = \varphi_2[\mathrm{Id} \otimes \nabla\psi]_{\#}\mu = (\nabla\psi_{\#}\mu)|_{\mathbb{R}\setminus\{0\}}$.

The example above illustrate that under an adequate hypothesis, one can express ν as the restriction of a push-forward of μ by an appropriate function. Since the expression of ν as some kind of push-forward of μ plays an important role in the theory of center-outward quantiles, a part of this text will be devoted to find sufficient conditions for such expression to hold. We refer to Subsections 2.3.1 and 2.3.2.

Next we introduce a family of zero-couplings that will be of great interest since we will show that its elements have nice properties. Under good assumptions on μ, this family contains all the cyclically monotone zero-couplings we will be interested in when both μ and ν are limit measures in the definition of regular variation.

Definition 2.12 (H-coupling). *For* $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d)$, *let* $\Gamma_h(\mu,\nu)$ *denote the family of zero-couplings* γ *between* μ *and* ν *satisfying* $\gamma(\{0\} \times (\mathbb{R}^d \setminus \{0\})) = 0$ *and such that* $\operatorname{spt}\gamma$ *is homogenous, i.e.*

$$(x,y) \in \operatorname{spt}\gamma \implies (\lambda x, \lambda y) \in \operatorname{spt}\gamma \qquad \forall\lambda > 0.$$

We will write $\Gamma_{h,cm}(\mu,\nu)$ *for* $\Gamma_h(\mu,\nu) \cap \Gamma_{0,cm}(\mu,\nu)$.

Remark 2.13. *By definition we have* $\Gamma_h(\mu,\nu) \subset \Gamma_0(\mu,\nu)$ *but the converse inclusion is false. We refer to Remark 2.9 for a counter-example.*

2.2 Sequences of zero-coupling and their support

This section relies heavily on the work by Segers on the convergence of sequences of couplings $\pi_n \in \Pi(P_n, Q_n), n \geq 1$ where $P_n, Q_n, n \geq 1$ are sequences of probability measures on $\mathbb{R}^d$, and the convergence of cyclically monotone sets containing their supports, i.e., $T_n \in \mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d), n \geq 1$ satisfying $\operatorname{spt} \pi_n \subset T_n$. Although being formulated for probability measures, most of the results in Segers (2022) are in fact given with a demonstration that extends straightforwardly to M_0. Indeed, thanks to Theorem 2.7 (Prohorov) and Theorem 2.4 (Portmanteau) for M_0 (which are quite different from the ones in the set of probability measures, see Section 1.1) in Hult and Lindskog (2006), most of Segers' arguments can be reused almost directly. In what follows, we adapt the results given in Segers (2022) to our context and give complete proofs while highlighting the changes made and the reasons for them.

The main consequence of the following Lemma for our purpose is that if a sequence $(\gamma_n, T_n) \in \Gamma_0(\mu_n, \nu_n) \times \mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d)$ satisfying $\operatorname{spt} \gamma_n \subset T_n$ converges in the product topology to some (γ, T), then the limit satisfies $\operatorname{spt} \gamma \subset T$. It is adapted from Lemma D.11 (Lower semi-continuity of the support map) in Segers (2022).

Lemma 2.14 (Lower semi-continuity of the support map)**.** *Let $\mathbb{E}$ be an open subset of $\mathbb{R}^k$ for some $k \geq 1$. If $\mu_n \xrightarrow{0} \mu$ in $\mathrm{M}_0(\mathbb{E})$—seen as a subset of $\mathrm{M}_0(\mathbb{R}^d)$—as $n \to \infty$, then $\operatorname{spt} \mu \subset \liminf_{n\to\infty} \operatorname{spt} \mu_n$, that is, the map*

$$\operatorname{spt} : \mathrm{M}_0(\mathbb{E}) \to \mathcal{F}(\mathbb{E}) : \mu \mapsto \operatorname{spt} \mu$$

is lower semi-continuous. In particular, $\{(\mu, F) \in \mathrm{M}_0(\mathbb{E}) \times \mathcal{F}(\mathbb{E}) : \operatorname{spt} \mu \subset F\}$ is closed in the product topology.

The following Lemma is adapted from Lemma 4.1 in Segers (2022) and will be usefull when dealing with converging sequences.

Lemma 2.15 (Closure and compactness properties of sets of zero-coupling measures)**.** *The following assertions are true.*

(a) The sets $\mathrm{M}_{0,cm}(\mathbb{R}^d \times \mathbb{R}^d)$ and $\Gamma_{0,cm}(\mu, \nu)$ for $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d)$, are closed in $\mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$;

(b) The sets

$$\{(\mu, \nu, \gamma) : \mu, \nu \in \mathrm{M}_0, \gamma \in \Gamma_0(\mu, \nu)\} \qquad \textit{and}$$
$$\{(\mu, \nu, \gamma) : \mu, \nu \in \mathrm{M}_0, \gamma \in \Gamma_{0,cm}(\mu, \nu)\}$$

are closed in $\mathrm{M}_0(\mathbb{R}^d) \times \mathrm{M}_0(\mathbb{R}^d) \times \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$;

(c) If $K, L \subset \mathrm{M}_0(\mathbb{R}^d)$ are compact (in $\mathrm{M}_0(\mathbb{R}^d)$), then

$$\{(\mu, \nu, \gamma) : \mu \in K, \nu \in L, \gamma \in \Gamma_0(\mu, \nu)\} \tag{4}$$

is compact too.

The following Lemma is a particular case of Lemma A.5 in de Valk and Segers (2018) and is similar to Lemma C.1 in Segers (2022) which concerns probability measures. We recall from Sections 1.5 and 2.1 that for every $\gamma \in \mathrm{M}_0(\mathbb{R}^d)$ the operators ϕ_i, φ_i, $i = 1, 2$ are defined by $\phi_i(\gamma) = \operatorname{proj}_{i\#} \gamma \in \mathrm{M}^+(\mathbb{R}^d)$ and $\varphi_i(\gamma) = (\operatorname{proj}_{i\#} \gamma)|_{\mathbb{R}^d \setminus \{0\}} \in \mathrm{M}_0(\mathbb{R}^d)$.

Lemma 2.16 (Support of a zero-margin)**.** *Let $\gamma \in \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$ and let $\mu = \varphi_1(\gamma) \in \mathrm{M}_0(\mathbb{R}^d)$ denote its left zero-marginal. Then $\operatorname{spt} \mu \subset \operatorname{spt} \phi_1(\gamma) = \operatorname{cl} \operatorname{proj}_1(\operatorname{spt} \gamma)$. As a consequence, if $\operatorname{spt} \gamma \subset T$ for some $T \subset \mathbb{R}^d \times \mathbb{R}^d$, then $\operatorname{spt} \mu \subset \operatorname{cl}(\operatorname{dom} T)$. In particular, if T is maximal monotone, then $\operatorname{int}(\operatorname{spt} \mu) \subset \operatorname{dom} T$.*

Remark 2.17. *Since $\varphi_1(\gamma) = \operatorname{res} \phi_1(\gamma)$ we have the inclusions $\operatorname{spt} \phi_1(\gamma) \subset \operatorname{spt} \varphi_1(\gamma) \cup \{0\}$ and $\operatorname{spt} \phi_1(\gamma) \supset \operatorname{spt} \varphi_1(\gamma)$.*

When dealing with measures in $\mathrm{M}_0(\mathbb{R}^d)$ with infinite mass like the ones that appear as limit measures in the definition of regular variation, one can further refine the result of the previous theorem.

Remark 2.18. *In the particular case where $\mu \in \mathrm{M}_0(\mathbb{R}^d)$ has infinite mass and $\gamma \in \Gamma_0(\mu, \nu)$ for some $\nu \in \mathrm{M}_0(\mathbb{R}^d)$, we can use an equality symbol in the last statement to write $\operatorname{spt} \mu = \operatorname{spt} \phi_1(\gamma) = \operatorname{cl} \operatorname{proj}_1(\operatorname{spt} \gamma)$. The identity $\operatorname{spt} \mu = \operatorname{spt} \phi_1(\gamma)$ is clear since μ has infinite mass. Indeed, as μ puts finite mass on sets bounded away from the origin, we necessarily have $0 \in \operatorname{spt} \mu \cap \operatorname{spt} \phi_1(\gamma)$, hence the two sets coincide.*

Moreover, under the same hypothesis, we can replace $\operatorname{spt} \mu$ by $\operatorname{proj}_1 \operatorname{spt} \gamma$ in our computations as they have the same μ-mass. It will be useful since $\operatorname{spt} \gamma$—seen as a multivalued map—is defined on $\operatorname{proj}_1 \operatorname{spt} \gamma$ but not necessarily on $\operatorname{spt} \mu$.

Remark 2.19. *Let $\gamma \in \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$ be a zero coupling between some $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d)$ with $\mu(\mathbb{R}^d \setminus \{0\}) = +\infty$, then $\operatorname{proj}_1 \operatorname{spt} \gamma$ is a μ-continuity set, i.e., $\mu\left(\operatorname{bnd} \operatorname{proj}_1 \operatorname{spt} \gamma\right) = \mu\left(\operatorname{spt} \mu \setminus \operatorname{proj}_1 \operatorname{spt} \gamma\right) = 0$. Indeed, as $\gamma \in \Gamma_0(\mu, \nu)$, γ has infinite mass too, hence $0 \in \operatorname{spt} \gamma$ and $0 \in \operatorname{proj}_1 \operatorname{spt} \gamma$. Thus $(\operatorname{proj}_1 \operatorname{spt} \gamma)^c \subset \mathbb{R}^d \setminus \{0\}$ and we can write successively*

$$\begin{aligned}
\mu\left((\operatorname{proj}_1 \operatorname{spt} \gamma)^c\right) &= \gamma\left(\operatorname{spt} \gamma \cap \left[(\operatorname{proj}_1 \operatorname{spt} \gamma)^c \times \mathbb{R}^d\right]\right) \\
&\leq \gamma\left(\left[\operatorname{proj}_1 \operatorname{spt} \gamma \times \operatorname{proj}_2 \operatorname{spt} \gamma\right] \cap \left[(\operatorname{proj}_1 \operatorname{spt} \gamma)^c \times \mathbb{R}^d\right]\right) \\
&= 0.
\end{aligned}$$

We recall from Subsection 1.3 that we write $T \llcorner V$ for the restriction of the multivalued map $T : \mathbb{R}^k \rightrightarrows \mathbb{R}^l$ to a subset V of $\mathbb{R}^k$. The following lemma will be useful to prove the graphical convergence, relative to the interior V of the support of μ, of a sequence of maximal cyclically monotone sets $T_n \in \mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d), n \geq 1$ containing the support of γ_n by allowing one to prove that the limit points of $T_n \llcorner V$ are all the same. It is adapted from Lemma 4.2 in Segers (2022).

Lemma 2.20 (Uniqueness of maximal monotone extension on the interior of the support)**.** *Let* $\gamma \in \Gamma_0(\mu, \nu)$ *for* $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d)$ *and let* $S, T \in \mathcal{F}_{\mathrm{mm}}(\mathbb{R}^d \times \mathbb{R}^d)$ *satisfy* $\operatorname{spt} \gamma \subset S \cap T$*. Put* $V = \operatorname{int}(\operatorname{spt} \mu)$ *and* $W = \operatorname{int}(\operatorname{spt} \nu)$*. Then* $V \subset \operatorname{dom} S \cap \operatorname{dom} T$ *and* $S \llcorner V = T \llcorner V$*, as well as* $W \subset \operatorname{rge} S \cap \operatorname{rge} T$ *and* $S^{-1} \llcorner W = T^{-1} \llcorner W$*.*

2.2.1 Deterministic sequences

For two sequences $\mu_n, \nu_n \in \mathrm{M}_0(\mathbb{R}^d), n \geq 1$ that are M_0-converging to $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d)$ respectively, we give here results about the stability of the possible convergence of any sequence γ_n of zero-couplings between μ_n and ν_n for each $n \geq 1$, and of the sequence of the supports $\operatorname{spt} \gamma_n$.

These results provide us powerful tools, because it allows one to use approximations; one can easily construct sequences such that both μ_n and ν_n have the same finite mass and use results about couplings between $\tilde{\mu}_n$ and $\tilde{\nu}_n$ possibly with cyclically monotone support to construct a sequence of zero-couplings with the desired properties.

Theorem 2.21 (Stability)**.** *Let* $\mu_n, \nu_n \in \mathrm{M}_0(\mathbb{R}^d), n \geq 1$ *and* $\gamma_n \in \Gamma_0(\mu, \nu), n \geq 1$ *be such that* μ_n*,* ν_n M_0*-converge to* $\mu, \nu \in \mathrm{M}_0$ *respectively, and suppose that we have a sequence of closed sets* $T_n \in \mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d), n \geq 1$ *such that* $\operatorname{spt} \gamma_n \subset T_n$ *for each* $n \geq 1$*, then there exists an infinite subset* N *of* $\mathbb{N}$ *such that*
(a) $\gamma_n \xrightarrow{0} \gamma$ *as* $n \to \infty$ *in* N *for some* $\gamma \in \Gamma_0(\mu, \nu)$*,*
(b) $T_n \xrightarrow{\mathcal{F}} T$ *as* $n \to \infty$ *in* N *for some* $T \in \mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d)$*,*
(c) $\operatorname{spt} \gamma \subset T$*.*
If we further assume that for every $n \geq 1$*,* $\gamma_n \in \Gamma_{0,cm}(\mu_n, \nu_n)$ *and* $T_n \in \mathcal{F}_{\mathrm{mcm}}(\mathbb{R}^d \times \mathbb{R}^d)$*, then all limits points of* $\gamma_n, n \geq 1$ *belong to* $\Gamma_{0,cm}(\mu, \nu)$ *while all limit points of* $T_n, n \geq 1$ *lie in* $\mathcal{F}_{\mathrm{mcm}}(\mathbb{R}^d \times \mathbb{R}^d)$*.*

Remark 2.22. *In Theorem 2.21, one cannot replace* $\Gamma_{0,cm}$ *by* $\Gamma_{h,cm}$ *since the latter is not closed and* $\Gamma_{h,cm}(\mu, \nu)$ *may potentially be empty.*

When $\Gamma_{0,cm}(\mu, \nu)$ is a singleton, the following corollary states that any sequence of cyclically monotone zero-couplings M_0-converges to the unique element in $\Gamma_{0,cm}(\mu, \nu)$.

Corollary 2.23. *Under the assumptions of Theorem 2.21, assume that for every* $n \geq 1$*,* $\gamma_n \in \Gamma_{0,cm}(\mu_n, \nu_n)$*,* $T_n \in \mathcal{F}_{\mathrm{mcm}}(\mathbb{R}^d \times \mathbb{R}^d)$ *and that there exist some* $T \in$

$\mathcal{F}_{\mathrm{cm}}(\mathbb{R}^d \times \mathbb{R}^d)$*,* $\gamma \in \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$ *such that* $\Gamma_{0,cm}(\mu,\nu) = \{\gamma\}$*,* $\operatorname{spt}\gamma \subset T$*. Then we have, as* $n \to \infty$*,*

$$\gamma_n \xrightarrow{0} \gamma \qquad \text{and} \qquad T_n \llcorner V \xrightarrow{\mathcal{F}} T \llcorner V \quad \text{in } \mathcal{F}(V \times \mathbb{R}^d).$$

where $V := \operatorname{int}(\operatorname{spt}(\mu))$*.*

2.2.2 Random sequences

In this section, we extend Corollary 2.23 to sequences of random measures $\mu_n, \nu_n \in \mathrm{M}_0(\mathbb{R}^d), n \geq 1$ weakly converging to $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d)$ respectively—the limits are deterministic elements—and an associated sequence of random zero-couplings $\gamma_n \in \Gamma_0(\mu_n, \nu_n)$.

This result is adapted from Theorem 1.2(a) in Segers (2022) and will be extensively used in Section 4 to deal with sequences of zero-couplings between empirical measures.

Theorem 2.24. *Let* $\mu_n, \nu_n, n \geq 1$ *be sequences of random elements of* $\mathrm{M}_0(\mathbb{R}^d)$ *satisfying* $\mu_n \xrightarrow{w} \mu$ *and* $\nu_n \xrightarrow{w} \nu$ *in* M_0 *with* μ, ν *deterministic elements of* $\mathrm{M}_0(\mathbb{R}^d)$*. We consider a sequence of random zero-couplings* $\gamma_n \in \Gamma_{0,cm}(\mu_n, \nu_n), n \geq 1$ *and a non-random* $\gamma \in \Gamma_{0,cm}(\mu, \nu)$*. Assume that there exist maximal cyclically monotone* $T_n, T \in \mathcal{F}_{\mathrm{mcm}}(\mathbb{R}^d \times \mathbb{R}^d)$ *containing almost surely* $\operatorname{spt}\gamma_n$ *and* γ *respectively. If* $V := \operatorname{int}(\operatorname{spt}(\mu))$ *is non-empty and if* $\Gamma_{0,cm}(\mu,\nu) = \{\gamma\}$*, then we have* $V \subset dom(T)$ *and, as* $n \to \infty$*,*

$$\gamma_n \xrightarrow{w} \gamma \quad \text{in } \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d) \qquad \text{and} \qquad T_n \llcorner V \xrightarrow{w} T \llcorner V \quad \text{in } \mathcal{F}(V \times \mathbb{R}^d).$$

Remark 2.25. *Using similar arguments, one can also write an extension to the random setting of Theorem 2.21 about weakly* M_0*-convergent subsequences when we do not assume that there exists a unique cyclically monotone zero-coupling between* μ *and* ν*. We did not do it since such a result would be useless for our purpose. Indeed, we will see in Theorem 2.38 that for* μ *and* ν *in* $\mathrm{M}_0(\mathbb{R}^d)$ *arising as limit measures in the definition of regular variation,* $\Gamma_{0,cm}(\mu,\nu)$ *will be a singleton under mild assumptions on* μ *and* $\operatorname{spt}\mu$*.*

2.3 Cyclically monotone zero-coupling

2.3.1 Existence of cyclically monotone zero-coupling

We first prove that for every $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d)$, the set $\Gamma_{0,cm}(\mu,\nu)$ is non-empty. This result is a direct consequence of Theorem 2.21 (Stability) and of Theorem 6 in McCann

(1995). For sequences $\mu_n, \nu_n \in \mathrm{M}_0(\mathbb{R}^d), n \geq 1$ approximating μ, ν such that for every $n \geq 1$, the measures μ_n and ν_n have equal finite mass, a rescaling argument and McCann's result give the existence of a cyclically monotone coupling $\tilde{\pi}_n$ between $\tilde{\mu}_n$ and $\tilde{\nu}_n$. Setting $\gamma_n = \operatorname{res}(\tilde{\pi}_n) \in \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$ we get a sequence of cyclically monotone zero-couplings between μ_n and ν_n. Thanks to Theorem 2.21 (Stability) it must converge along some subsequence to an element of $\Gamma_{0,cm}(\mu, \nu)$.

Theorem 2.26 (Existence)**.** *Let $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d)$ have equal nonzero total mass, then $\Gamma_{0,cm}(\mu, \nu) \neq \varnothing$.*

The next result is quite similar to Proposition 10 in McCann (1995). It links the expression of a cyclically monotone zero-coupling γ between μ and ν such that μ satisfies a mild assumption to the gradient of any closed convex function ψ such that the support of γ is contained in $\partial\psi$. Contrary to McCann's result for finite measures, there is an extra term that can be interpreted as a compensator or error term which puts mass on the axis $\{0\} \times (\mathbb{R}^d \setminus \{0\})$ when we need to "take" mass from zero like in Example 2.9.

Theorem 2.27 (Representation)**.** *Let $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d)$ have equal nonzero total mass, let $\gamma \in \Gamma_{0,cm}(\mu, \nu)$, and assume that μ vanishes on sets of Hausdorff dimension at most $d-1$. Then, for any closed convex function ψ satisfying $\operatorname{spt} \gamma \subset \partial\psi$ we can write*

$$\gamma = [\mathrm{Id} \otimes \nabla\psi]_{\#}\mu + \delta_0 \otimes [\nu - \nabla\psi_{\#}\mu],$$

in the sense that for every Borel set $B \subset (\mathbb{R}^d \times \mathbb{R}^d) \setminus \{0\}$, we have

$$[\mathrm{Id} \otimes \nabla\psi]_{\#}\mu(B) = \mu\left(\{x \in \operatorname{dom} \nabla\psi \setminus \{0\} : (x, \nabla\psi(x)) \in B\}\right),$$

and for every Borel set $B \subset (\mathbb{R}^d \times \mathbb{R}^d) \setminus \{0\}$ bounded way from the origin in $\mathbb{R}^d \times \mathbb{R}^d$, we have

$$\left(\delta_0 \otimes [\nu - \nabla\psi_{\#}\mu]\right)(B) = [\nu - \nabla\psi_{\#}\mu](\{x \in \mathbb{R}^d \setminus \{0\} : (0, x) \in B\})$$

where $\{x \in \mathbb{R}^d \setminus \{0\} : (0, x) \in B\}$ is bounded away from the origin in $\mathbb{R}^d$, preventing the case $\infty - \infty$ in the latter equality from arising.

Remark 2.28. *In the proof of the theorem above, we have shown that $\gamma = [\mathrm{Id} \otimes \nabla\psi]_{\#}\mu$, for every closed convex function ψ satisfying $\operatorname{spt} \gamma \subset \partial\psi$, whenever $\gamma(\{0\} \times (\mathbb{R}^d \setminus \{0\})) = 0$. In particular, this equality holds when γ lies in $\Gamma_{h,cm}(\mu, \nu)$.*

In the rest of this text, we will be interested in cases where $\gamma = [\mathrm{Id} \otimes \nabla\psi]_{\#}\mu$, for which ν can be expressed as the restriction of the push-forward of μ by $\nabla\psi$, i.e., $\nu = \operatorname{res}((\nabla\psi)_{\#}\mu)$.

2.3.2 Uniqueness

Looking at Corollary 2.23, we would like to have sufficient conditions for the family $\Gamma_{0,cm}(\mu,\nu)$ of cyclically monotone zero-couplings between μ and ν to be a singleton in order to get the convergence of the whole sequence $\gamma_n \in \Gamma_{0,cm}(\mu_n,\nu_n), n \geq 1$ to the unique cyclically monotone zero-coupling $\gamma \in \Gamma_{0,cm}(\mu,\nu)$ where μ_n, ν_n M_0-converge to μ, ν respectively.

General results are given in an unpublished manuscript by de Valk, C. and Segers, J. However, since the proof is long and tricky, we choose another approach in this text. We will only prove the uniqueness in the cases we are interested in: when the cyclically monotone couplings put no mass on $\{0\} \times \mathbb{R}^d$ and have homogenous support, i.e., for cyclically monotone h-couplings. We write $H(x), H_+(x)$ and $H_-(x)$ for the hyperplane, positive and negative half-spaces induced by x and defined as follows;

$$\begin{cases} H(x) = \{y : \langle x, y\rangle = 0\}, \\ H_+(x) = \{y : \langle x, y\rangle > 0\}, \\ H_-(x) = \{y : \langle x, y\rangle < 0\}. \end{cases}$$

Theorem 2.29. *For $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d)$ with equal nonzero mass such that μ vanishes on sets of Hausdorff dimension at most $d-1$, the set $\Gamma_{h,cm}(\mu,\nu)$ contains at most one element. Assume that $\Gamma_{h,cm}(\mu,\nu)$ is nonempty, then there exists a maximal cyclically monotone map $T : \mathbb{R}^d \to \mathbb{R}^d$ which is unique μ-almost everywhere such that*

$$\begin{cases} \gamma = [\mathrm{Id} \otimes T]_{\#}\mu \\ \nu = \mathrm{res}(\nabla\psi_{\#}\mu) \end{cases}$$

Moreover, T satisfies $T = \nabla\psi$ μ-almost everywhere for every closed convex function ψ for which we have $\operatorname{spt}\gamma \subset \partial\psi$.

In particular, if we also suppose in Theorem 2.27 that $\Gamma_{0,cm}(\mu,\nu) \subset \Gamma_{h,cm}(\mu,\nu)$, i.e.

$$\gamma(\{0\} \times (\mathbb{R}^d \setminus \{0\})) = 0\,, \quad \forall\gamma \in \Gamma_{0,cm}(\mu,\nu)$$

$$(x,y) \in \operatorname{spt}\gamma \implies (\lambda x, \lambda y) \in \operatorname{spt}\gamma\,, \quad \forall\gamma \in \Gamma_{0,cm}(\mu,\nu),\ \forall\lambda > 0$$

then $\Gamma_{0,cm}(\mu,\nu)$ is a singleton.

The rest of this subsection will be devoted to conditions under which we have $\Gamma_{0,cm}(\mu,\nu) \subset \Gamma_{h,cm}(\mu,\nu)$. When the support of cyclically monotone zero-couplings are homogenous, we have very simple criteria to show $\Gamma_{0,cm}(\mu,\nu) \subset \Gamma_{h,cm}(\mu,\nu)$. From Section 4 onwards, all the measures with which we are going to work will satisfy this property.

Lemma 2.30. *For every $A \subset \mathbb{R}^d \times \mathbb{R}^d$ satisfying*

(a) monotonicity : $(x, y), (a, b) \in A \implies \langle x - a, y - b\rangle \geq 0$

(b) homogeneity : $\forall \lambda > 0$, $(x, y) \in A \implies (\lambda x, \lambda y) \in A$

the set $A \cap (\{0\} \times H_+(x))$ is empty for every x lying in $\operatorname{proj}_1 A$.

Remark 2.31. *Let $\gamma \in \Gamma_0(\mu, \nu)$ have cyclically monotone and homogenous support. For $A = \operatorname{spt}\gamma$, we demonstrated in the proof of the above lemma that for every $(x, y) \in \operatorname{spt}\gamma$, $\langle x, y\rangle \geq 0$, i.e., $y \in H(x) \cup H_+(x)$.*

This simple lemma gives us two useful results.

Proposition 2.32. *Let $E \subset \mathbb{R}^d$ be such that for every $x, y \in E$ we have $\langle x, y\rangle \geq 0$. Let $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d)$ with non-zero equal mass and with μ vanishing on sets of Hausdorff dimension at most $d-1$. If $\operatorname{spt}(\mu) \cup \operatorname{spt}(\nu) \subset E$, then for every $\gamma \in \Gamma_{0,cm}(\mu, \nu)$ whose support is homogenous we have $\gamma(\{0\} \times (\mathbb{R}^d \setminus \{0\})) = 0$, so that in fact γ lies in $\Gamma_{h,cm}(\mu, \nu)$.*

Remark 2.33. *On should notice that orthants are particular cases of such sets E. In fact, any set satisfying the assumptions is included in some orthant.*

Theorem 2.34. *Let $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d)$ have homogenous support and non-zero equal mass. If there exist $x_1, \ldots, x_k \in \operatorname{spt}\mu$ for some $k \in \mathbb{N}$ such that $\bigcup_{i=1}^k H_+(x_i) = \mathbb{R}^d \setminus \{0\}$, then for every $\gamma \in \Gamma_{0,cm}(\mu, \nu)$ whose support is homogenous we have $\gamma(\{0\} \times (\mathbb{R}^d \setminus \{0\})) = 0$, so that in fact γ lies in $\Gamma_{h,cm}(\mu, \nu)$.*

A direct application of Theorem 2.34 above together with Theorem 2.29 yields the following corollary.

Corollary 2.35. *Let μ, ν satisfy the assumptions of Theorem 2.34, then there is a unique cyclically monotone zero-coupling with homogenous support γ between μ and ν. Moreover, γ lies in $\Gamma_{h,cm}(\mu, \nu)$.*

Remark 2.36. *Thanks to the last two theorems, to show that $\Gamma_{0,cm}(\mu, \nu)$ is contained in $\Gamma_{h,cm}(\mu, \nu)$, it suffices to show that every $\gamma \in \Gamma_{0,cm}(\mu, \nu)$ has homogenous support and satisfies the hypotheses of one of the theorems.*

To prove the convergence of any sequence of cyclically monotone zero-couplings $\gamma_n \in \Gamma_{0,cm}(\mu_n, \nu_n), n \geq 1$ where μ_n, ν_n M_0-converge to $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d)$ respectively, it suffices to show that any limit point of $\gamma_n, n \geq 1$ has homogenous support and satisfies the hypothesis of one of the theorems above.

Remark 2.37. *In Section 4, we will consider a special μ with homogenous support satisfying $\operatorname{spt}\mu = \mathbb{R}^d$. For $e_1, \ldots, e_n$ an orthogonal basis of $\mathbb{R}^d$, it suffices to choose $e_1, \ldots, e_n, -e_1, \ldots, -e_n$ to fulfill the requirements of Theorem 2.34.*

2.3.3 Tails of transport plans between regularly varying distributions

In this last subsubsection, we apply the theory developed so far to study the unique cyclically monotone coupling, given by Main Theorem in McCann (1995), between probability measures $P, Q \in \mathcal{P}(\mathbb{R}^d)$ regularly varying with common auxiliary function b, common indice $\alpha > 0$, and limit measures μ and ν in $\mathrm{M}_0(\mathbb{R}^d)$ respectively. Under a mild assumption on μ and its support $\operatorname{spt}\mu$, we show that Main Theorem in McCann (1995) admits an extension to cyclically monotone zero-coupling between the limit measures μ and ν in the definition of regular variation. Moreover, we will show the M_0-convergence of the restriction of a rescaled version of the unique cyclically monotone coupling π between P and Q, and the $\mathcal{F}$-convergence of a restriction of the rescaled version of the subgradient of a closed convex function ψ satisfying $\operatorname{spt}\pi \subset \partial\psi$.

Theorem 2.38. *Let P and Q in $\mathcal{P}(\mathbb{R}^d)$ be regularly varying with common auxiliary function b, indice $\alpha > 0$, and limit measures μ and ν in $\mathrm{M}_0(\mathbb{R}^d)$. Let ψ be a closed convex function such that the graph of $\partial\psi$ contains the cyclically monotone support of some $\pi \in \Pi(P, Q)$.*

(a) Every sequence of positive numbers $t_n \to \infty$ contains a subsequence such that, as $n \to \infty$ along the subsequence, we have

$$\left.\begin{aligned} t_n \operatorname{res}\pi(b(t_n)\,\cdot\,) &\xrightarrow{0} \gamma \qquad \text{in } \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d) \\ b(t_n)^{-1}\partial\psi(b(t_n)\,\cdot\,) &\xrightarrow{\mathcal{F}} \partial\bar{\psi} \end{aligned}\right\} \tag{5}$$

for some $\gamma \in \Gamma_{0,cm}(\mu, \nu)$ and some closed convex function $\bar{\psi}$ satisfying $\operatorname{spt}\gamma \subset \partial\bar{\psi}$.

Moreover, the support of γ is homogenous,

$$\lambda^{1/\alpha} \operatorname{spt}\gamma = \operatorname{spt}\gamma, \qquad \lambda > 0,$$

and γ satisfies

$$\gamma(\lambda^{-1/\alpha}\,\cdot\,) = \lambda\gamma, \qquad \lambda > 0. \tag{6}$$

(b) If the limiting measure μ vanishes on sets of Hausdorff dimension not larger than $d-1$ and if we assume that no γ in $\Gamma_{0,cm}(\mu, \nu)$ puts mass on $\{0\} \times (\mathbb{R}^d \setminus \{0\})$, then

$$\gamma = [\mathrm{Id} \otimes \nabla\bar{\psi}]_{\#}\mu$$

is the unique zero-coupling between μ and ν having cyclically monotone support. The gradient $\nabla\bar{\psi}$ is determined uniquely μ-almost everywhere and satisfies $\nabla\bar{\psi}_{\#}\mu|_{\mathbb{R}^d\setminus\{0\}} = \nu$ and

$$\nabla\bar{\psi}(\lambda x) = \lambda\nabla\bar{\psi}(x), \qquad \lambda > 0$$

μ-almost everywhere.

(c) *If we further assume that* $\operatorname{spt}\mu = \mathbb{R}^d$*, then* $\Gamma_{0,cm}(\mu,\nu)$ *is a singleton whose single element belongs to* $\Gamma_{h,cm}(\mu,\nu)$*, and the convergences in (a) take place along the whole sequence and the limits are characterized by (b).*

Under assumption (a), (b) and (c), the theorem says that the restriction $\operatorname{res}\pi$ of the unique cyclically-monotone coupling between P and Q is regularly varying with the same indice and auxiliary function that P and Q—but in $\mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$ instead of $\mathrm{M}_0(\mathbb{R}^d)$—and that the limit measure in the definition of regular variation is the unique cyclically monotone zero-coupling between the limit measures μ and ν.

Remark 2.39. *The additional assumption in (c), i.e.,* $\operatorname{spt}\mu = \mathbb{R}^d$*,can obviously be relaxed since it is only used to apply Theorem 2.34 through Remark 2.37. If there exist* $x_1, \ldots, x_k \in \operatorname{spt}\mu$ *for some* $k \in \mathbb{N}$ *such that* $\bigcup_{i=1}^k H_+(x_i) = \mathbb{R}^d \setminus \{0\}$*, then the conclusion of (c) still holds. We state it this way since it is the form that will be useful in the following sections.*

The latter theorem will be of constant use in the rest of this paper and will allow us to define center center-outward tail quantiles in Section 4 and will give some rationale to the construction of some approximations in practice in Subsubsection 4.4.2.

3 New material

In Subsection 4.1 below I defined the center outward tail quantile region and contour as

$$\mathbb{C}^{\tau}(q) = \nabla\bar{\psi}(\mathrm{B}_{0,(1-q)^{-1/\alpha}}) = \{\nabla\bar{\psi}(x) \,:\, x \in D_{\bar{\psi}},\, |x| \le (1-q)^{-1/\alpha}\}$$

and

$$\mathcal{C}^{\tau}(q) = \nabla\bar{\psi}((1-q)^{-1/\alpha}\mathbb{S}^{d-1}) = \{\nabla\bar{\psi}(x) \,:\, x \in D_{\bar{\psi}},\, |x| = (1-q)^{-1/\alpha}\},$$

where $D_{\bar{\psi}}$ denotes the set of points where $\bar{\psi}$ is differentiable. I thought it was in some sense natural to do so since I was interested in the convergence of the cloud of point $\hat{\mathcal{C}}_n^{\tau}(0)$—built from the support of $\hat{\gamma}_{n,t_n}$—toward $\mathcal{C}^{\tau}(0)$ and the complement of $D_{\bar{\psi}}$ being of null mass for μ one could expect to have no points in this set.
However, as Anne suggested to me last year, it could be wiser to look at the convergence of a well chosen interpolation of transport plan toward the contour. Bearing this in mind, we should rather define the tail contour and quantiles as

$$\mathbb{C}^{t}(q) = \partial\bar{\psi}(\mathrm{cl}\,\mathrm{B}_{0,(1-q)^{-1/\alpha}})$$

and

$$\mathcal{C}^{t}(q) = \partial\bar{\psi}((1-q)^{-1/\alpha}\mathbb{S}^{d-1}).$$

Doing so, the two above sets are closed—in fact they are compact by virtue of Lemma 3.4 in Segers (2022)—since $\partial\bar{\psi}$ is maximal monotone.

For a set $C \subset \mathbb{R}^d$, the horizon cone is the closed cone $C^{\infty} \subset \mathbb{R}^d$ defined by

$$C^{\infty} = \begin{cases} \{x \mid \exists x_n \in C, \lambda_n \to 0, \lambda_n > 0, n \in \mathbb{N}, \text{ with } \lambda_n x_n \to x\} & when\, C \neq \varnothing \\ \{0\} & when\, C = \varnothing \end{cases}$$

Relying on Theorem 2.38 one can show the identity

$$\partial\bar{\psi} = (\partial\psi)^{\infty}$$

We make the following assumptions:

- The measure P_X (P_Y) is regularly varying with limit measure μ (ν) tail index α, and auxiliary function b.
- The measures P_X and P_Y vanish on sets of Hausdorff dimension at most $d-1$.
- μ and ν vanish on sets of Hausdorff dimension at most $d-1$.

- There exist $x_1, \ldots, x_k \in \operatorname{spt} \mu$ for some $k \in \mathbb{N}$ such that $\bigcup_{i=1}^k H_+(x_i) = \mathbb{R}^d \setminus \{0\}$ and the same holds for $\operatorname{spt} \nu$.

Taking ideas from Remark 1.34 and relying on Theorem 2.38 one can prove that

$$\partial \overline{\psi^*} = [\partial \bar{\psi}]^{-1} = \partial (\bar{\psi}^*)$$

i.e.

$$(\partial \psi^*)^\infty = [(\partial \psi)^\infty]^{-1}$$

3.1 Simple extremal approximations and quantile regions

Writing $p_t = \mathsf{P}(|Y| > b(t))$ we have

$$\mathsf{P}(Y \in A) = (1 - p_t)\, \mathsf{P}\left(Y \in A \mid |Y| \leq b(t)\right) + p_t\, \mathsf{P}\left(Y \in A \mid |Y| > b(t)\right).$$

Assume $A = (A^\infty \cap B^c) \cup (A \cap B)$ where $B = \bigcup_{i \in \mathrm{N} \subset \mathbb{N}} (B_i \cap \mathrm{B}_{0,r_i})$ for some family $(B_i)_{i \in \mathrm{N}}$, $\mathrm{N} \subset \mathbb{N}$, of cones whose interior is nonempty and $0 < r_i < \infty$. This is equivalent to the assumption

$$\exists M > 0, \quad A \cap \mathrm{B}_{0,M}^c = A^\infty \cap \mathrm{B}_{0,M}^c.$$

Thus, for t large enough, since $\mathbb{S}^{d-1}$ is compact, we have $b(t)^{-1} A = A^\infty$. As a consequence, as $t \to \infty$,

$$\mathsf{P}\left(Y \in A \mid |Y| > b(t)\right) \to \nu\left(A^\infty \cap (\operatorname{cl} \mathrm{B}_{0,1})^c\right).$$

Assuming that the convergence is achieved in the above decomposition—in order to use the simple model from Extreme Value Theory for the areas of the space with few points—, we write, for some large t^*

$$\mathsf{P}(Y \in A) \approx (1 - p_{t^*})\, \mathsf{P}\left(Y \in A \mid |Y| \leq b(t^*)\right) + p_{t^*} \nu\left(A^\infty \cap (\operatorname{cl} \mathrm{B}_{0,1})^c\right).$$

We now deal with the optimal transport based approach developed throughout this text. Let $p_t = \mathsf{P}\left(Y \notin \mathbb{C}(1/t^*)\right)$, we have the simple decomposition

$$\mathsf{P}\left(Y \in A\right) = (1 - p_t)\, \mathsf{P}\left(Y \in A \mid Y \in \mathbb{C}(1/t)\right) + p_t\, \mathsf{P}\left(Y \in A \mid Y \notin \mathbb{C}(1/t)\right).$$

Under suitable assumptions [AM:exactly the same as above] we will prove that, as $t \to \infty$,

$$\mathsf{P}\left(Y \in A \mid Y \notin \mathbb{C}(1/t)\right) \to \mu\left(\left([\nabla \psi]^{-1}(A)\right)^\infty \cap (\operatorname{cl} \mathrm{B}_{0,1})^c\right).$$

Assuming like before that the convergence is achieved in the above decomposition, we write, for some large t^*

$$\mathsf{P}\left(Y \in A\right) \approx (1 - p_{t^*})\, \mathsf{P}\left(Y \in A \mid Y \in \mathbb{C}(1/t^*)\right) + p_{t^*}\, \mu\left(\left([\nabla\psi]^{-1}(A)\right)^{\infty} \cap (\operatorname{cl} \mathrm{B}_{0,1})^c\right)$$

i.e.

$$\mathsf{P}\left(Y \in A\right) \approx (1 - p_{t^*})\, \mathsf{P}\left(Y \in A \mid Y \in \mathbb{C}(1/t^*)\right) + p_{t^*}\, \nu\left(\nabla\bar{\psi}\left[\left([\nabla\psi]^{-1}(A)\right)^{\infty}\right] \cap \mathbb{C}^{\tau}(0)^c\right)$$

Remark 3.1. *[AM: This simple approximation provides a easy way to generate new points. The Mzp-convergence of the approximation is trivial but I wonder how to relevantly measure the quality of the approximation.*
In practice, in addition to the transport plan estimated to build quantile regions, it would require to estimate the tail index α to generate points in the reference distribution. Using $\hat{\alpha}$ we define the measure $\hat{\mu}$ as a substitute for μ whose polar decomposition only depends on α, then $\nabla\hat{\bar{\psi}}$ allow us to define $\hat{\nu} = \nabla\hat{\bar{\psi}}_{\#}\hat{\mu}$. Finally we use $\hat{\nu}$ to draw extremal points. For the non-extremal part, since the polar decomposition of the reference distribution involves the uniform on the unit sphere, it reduces to the one-dimensional case and one can simply use the generalized inverse of the empirical distribution function of the radial part.
One should apply the estimated function $\hat{\bar{\psi}}$ only to points on the unit sphere, using homogeneity to avoid using it on regions of the space where it may be a poor approximation of $\bar{\psi}$.]

We now propose a way to build extreme quantile regions.

In the simple case of a reference distribution given by a polar decomposition with uniform distribution on the unit sphere for angular distribution, a very natural way to build quantile regions is to use balls centered at the origin.

Let X be such a random variable with regularly varying (μ, α, b) distribution P_X vanishing on small sets. We have

$$\mu(\mathrm{B}^c_{0,(1-\beta)^{-1/\alpha}}) = 1 - \beta$$

and the convergence, as $t \to \infty$

$$t\, \mathsf{P}(b(t)^{-1}X \in \mathrm{B}^c_{0,(1-\beta)^{-1/\alpha}}) \to \mu(\mathrm{B}^c_{0,(1-\beta)^{-1/\alpha}}) = 1 - \beta$$

For $\beta \in [0, 1]$, the simple approximation from previous page can be written, for large t,

$$\mathsf{P}(X \notin (b(t)\mathrm{B}_{0,(1-\beta)^{-1/\alpha}})) \approx p_{t^*}\nu\left(\mathrm{B}^c_{0,(1-\beta)^{-1/\alpha}}\right) = p_{t^*}(1 - \beta)$$

As a consequence we suggest to use for quantile region of order q the set $b(t^*)\mathrm{B}_{0,(1-\beta)^{-1/\alpha}}$ where $\beta = p_{t^*}^{-1}(q-1+p_{t^*})$ Notice that the set

$$b(t^*)\mathrm{B}_{0,(1-\beta)^{-1/\alpha}}$$

can be written $b(t^*)\mathrm{B}_{0,(1-\beta)^{-1/\alpha}} \cup \mathrm{B}_{0,b(t)}$ since β lies in $[0,1]$.

Likewise, dealing with center-outward quantile regions proposed in Hallin et al. (2021), we propose to take for quantile region of order q the set

$$\mathbb{C}(1-1/t) \cup b(t)\mathbb{C}^\tau(\beta)$$

where β is chosen as before.

This choice is motivated by the above approximation and the following convergence.

Conjecture 3.2. *For $\beta \in (0,1)$ we have the convergence*

$$t\,\mathsf{P}\left(Y \in b(t)\left(\mathbb{C}^\tau(\beta)^c \cap b(t)^{-1}\mathbb{C}(1-1/t)^c\right)\right) \longrightarrow \nu\left(\mathbb{C}^\tau(\beta)^c\right) = (1-p)(1-\beta)$$

as $t \to \infty$, where $p = \mu\left([\nabla\psi]^{-1}(\{0\}) \cap \mathrm{B}_{0,1}^c\right)$ as in Theorem 4.11. Let $\chi = \{x \in \mathbb{R}^d : \operatorname{spt}\nu \cap \operatorname{cl} H_+(x) = \{0\}\}$ like in Theorem 4.10. If we further assume that $\mu(\chi) = 0$, then $p = 0$ in the expression above.

[AM:I have a proof when $p=0$. Hopefully it easily extends to the general case.]

3.2 Estimation of $\mathcal{C}^\tau(0)$ and $\mathbb{C}^\tau(0)$

Relying on Hallin et al. (2021), where a construction of a continuous maximal cyclically monotone interpolation of the discrete optimal transport plan is proposed in Section 3, we can define an empirical version $\partial\hat{\psi}_n$—which is single valued—of $\partial\psi$ (where ψ is a closed convex function whose gradient pushes P forward to Q). This allows one to define, using Theorem 2.38, an empirical rescaled interpolation

$$\partial\hat{\bar{\psi}}_{n,t_n} = \hat{b}_n(t_n)^{-1}\partial\hat{\psi}_n(\hat{b}_n(t_n)\cdot)$$

for which a simple computation yields the expected relation

$$\begin{aligned}\partial\hat{\bar{\psi}}_{n,t_n}\Big(\frac{|X_i|U_i}{\hat{b}_n(t)}\Big) &= \hat{b}_n(t_n)^{-1}\partial\hat{\psi}_n\Big(\hat{b}_n(t_n)\frac{|X_i|U_i}{\hat{b}_n(t)}\Big)\\ &= \hat{b}_n(t_n)^{-1}\partial\hat{\psi}_n\Big(\hat{b}_n(t_n)\frac{|X_i|U_i}{\hat{b}_n(t)}\Big) = \frac{X_{\hat{\sigma}_n(i)}}{\hat{b}_n(t)},\end{aligned}$$

whence $\operatorname{spt}\hat{\gamma}_{n,t} \subset \partial\hat{\bar{\psi}}_{n,t_n}$. Since μ vanishes on sets of Hausdorff dimension at most $d-1$ and $\operatorname{spt}\mu = \mathbb{R}^d$, by Theorem 2.38 $\Gamma_{0,cm}(\mu,\nu)$ is a singleton and Theorem 2.24 yields

$$\partial\hat{\psi}_{n,t_n} \xrightarrow{w} \partial\bar{\psi} \ \text{ in } \mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d) \text{ i.e., } \partial\hat{\psi}_{n,t_n} \xrightarrow{g} \partial\bar{\psi} \ \text{ in probability.}$$

Remark 3.3. *The empirical subdifferential $\partial\hat{\bar{\psi}}_{n,t_n}$ does not satisfy the homogeneity of $\partial\bar{\psi}$ from Theorem 5.1 in de Valk and Segers (2018).*

We then define an alternative estimator of the center-outward tail contour of order 0 by

$$\hat{\mathcal{C}}_n^\tau(0) = \partial\hat{\psi}_{n,t_n}(\mathbb{S}^{d-1}).$$

Recall that thanks to the homogeneity of $\partial\bar{\psi}$ from Theorem 5.1 in de Valk and Segers (2018), estimating $\mathcal{C}^\tau(0)$ also gives an estimation of the set $\mathbb{C}^\tau(q)$ for every $q \in (0,1)$.

Remark 3.4. *Defined this way, $\hat{\mathcal{C}}_n^t(0)$ does not rely on the homogeneity of the limit $\partial\bar{\psi}$ and does not necessarily use the most extreme points since it takes them into account solely through the construction of the interpolation which is some kind of smoothing of a piecewise constant map. It is simply a rescaling of a specific center-outward contour defined like in Hallin et al. (2021) (with the spherical uniform reference distribution replaced by the reference distribution defined by the polar decomposition $|\cdot|_\# Q \otimes \mathrm{Unif}(\mathbb{S}^{d-1})$).*

Since $\partial\bar{\psi}$ may not be single valued on $\mathbb{S}^{d-1}$, the nice result from Theorem 1.1(b) in Segers (2022)—which concludes some uniform convergence from the Fell convergence—cannot be applied here directly. Using the theory developed in Segers (2022), we will now prove the uniform convergence of an estimator toward a reasonable restriction of the center-outward tail quantile contour $\mathcal{C}^\tau(0)$.

We write $\bar{\psi}^*$ for the Legendre-Fenchel transform of the closed convex function $\bar{\psi}$. The map $\bar{\psi}^*$ satisfies $\partial\bar{\psi}^* = (\partial\bar{\psi})^{-1}$. Let $D_{\bar{\psi}}$ and $D_{\bar{\psi}^*}$ denote the sets of points where $\partial\bar{\psi}$ and $\partial\bar{\psi}^*$ are single valued respectively. Under the assumption—which is already used in Section 4—that both μ and ν vanish on sets of Hausdorff dimension at most $d-1$, we have $\mu(D_{\bar{\psi}} \cap \{0\}^c) = \nu(D_{\bar{\psi}^*} \cap \{0\}^c) = 0$. We define the set A by

$$A = (D_{\bar{\psi}} \cup \{0\}) \cap [\partial\bar{\psi}^*](D_{\bar{\psi}^*} \cup \{0\}) \cap \{0\}^c$$

and claim that $\mu(A^c) = 0$ [AM:the complement is taken in the set $\mathbb{R}^d$ as usual when dealing with measures in $\mathrm{M}_0(\mathbb{R}^d)$]. It suffices to show that $\mu(([\partial\bar{\psi}^*](D_{\bar{\psi}^*} \cup \{0\}) \cap \{0\}^c)^c) = 0$. This is clear since

$$\begin{aligned}\mu(A^c) &= \mu(([\partial\bar{\psi}^*](D_{\bar{\psi}^*} \cup \{0\}) \cap \{0\}^c)^c) \\ &= \mu(([\nabla\bar{\psi}]^{-1}(D_{\bar{\psi}^*} \cup \{0\}) \cap \{0\}^c)^c) \\ &= \nu((D_{\bar{\psi}^*} \cup \{0\})^c) = 0\end{aligned}$$

It seems reasonable to approximate the set $\mathcal{C}^\tau(0) = \partial\bar{\psi}(\mathbb{S}^{d-1})$ by $\partial\bar{\psi}(\mathbb{S}^{d-1} \cap D_{\bar{\psi}})$ since $\nu((\partial\bar{\psi}(\mathbb{B}_{0,1}) \setminus \partial\bar{\psi}(\mathbb{B}_{0,1} \cap D_{\bar{\psi}})) \cap \{0\}^c) = 0$ thanks to the above equalities. [AM:I wrote the latter identity using tail quantile regions instead of contours since $\nu(\mathcal{C}^\tau(0))$ may be equal to zero while $\nu(\mathbb{C}^\tau(0))$ cannot; approximating a set with zero ν-mass by a smaller one with zero ν-mass made less sens to me.]

To begin with, it happens that there exist some set W whose complement in $\mathbb{R}^d$ is of Hausdoff dimension at most $d-1$ (hence $\mu(W^c) = 0$) such that $\mathrm{d}_H(\partial\hat{\psi}_{n,t_n}(W \cap \mathbb{S}^{d-1}), \partial\bar{\psi}(W \cap \mathbb{S}^{d-1}))$ goes to zero in probability as $n \to \infty$. We will choose W to be the set $D_{\bar{\psi}}$ of points at which $\partial\bar{\psi}$ is single valued, i.e., where $\nabla\bar{\psi}$ is defined.

Remark 3.5. *Under the assumption—which is already done in Section 4—that both μ and ν vanish on sets of Hausdorff dimension at most $d-1$, it is immediate (we refer to Remark 1.34 for details) that we have*

$$\begin{aligned}\nu(\partial\bar{\psi}(D_{\bar{\psi}} \cap \mathbb{S}^{d-1})^c) &= \mu([\nabla\bar{\psi}]^{-1}\nabla\bar{\psi}(D_{\bar{\psi}} \cap \mathbb{S}^{d-1})^c \cap \{0\}^c) \\ &= \tilde{\mu}((D_{\bar{\psi}} \cap \mathbb{S}^{d-1})^c) = 0\end{aligned}$$

As a consequence, it seems reasonable to look for the convergence of some estimator toward the set $\partial\bar{\psi}(D_{\bar{\psi}} \cap \mathbb{S}^{d-1})$ instead of toward the whole set $\partial\bar{\psi}(\mathbb{S}^{d-1})$.

We can achieve this using a slightly modified version of the following proposition which is taken from Segers (2022).

Proposition 3.6 (Local uniform convergence, Segers (2022))**.** *Let $T \in \mathcal{F}_{\mathrm{mm}}$, let $V \subset \operatorname{dom} T$ be open and let $K \subset V$ be compact and non-empty. Assume that T is single-valued on K. For every $\epsilon > 0$ there exists an open neighbourhood $\mathcal{G}$ of $T \llcorner V$ in $\mathcal{F}(V \times \mathbb{R}^d)$ such that for all $T' \in \mathcal{F}_{\mathrm{mm}}$ with $T' \llcorner V \in \mathcal{G}$, we have $K \subset \operatorname{dom} T'$ and*

$$\sup_{x \in K} \sup_{y \in T'(x)} |y - T(x)| \leq \epsilon.$$

As a consequence, for any sequence $(T_n)_n$ in $\mathcal{F}_{\mathrm{mm}}$, if $T_n \xrightarrow{v} T$ as $n \to \infty$, then $K \subset \operatorname{dom} T_n$ for all but finitely many n and $\sup_{x \in K} \sup_{y \in T_n(x)} |y - T(x)| \to 0$ as $n \to \infty$.

A proof of it is given in the same text as follows.

Proof of Proposition 3.6. For every $x \in K$, Lemma 3.5 in Segers (2022) applied to the compact set $\{x\}$ guarantees there exists $\delta_x \in (0, \epsilon)$ and a neighbourhood $\mathcal{G}_x$ of $T \llcorner V$ in $\mathcal{F}(V \times \mathbb{R}^d)$ such that for any $T' \in \mathcal{F}_{\mathrm{mm}}$ with $T' \llcorner V \in \mathcal{G}_x$ we have $x + \delta_x \mathbb{B} \subset \operatorname{dom} T'$ and

$$d_H\big(T'(x + \delta_x \mathbb{B}), T(x)\big) \leq \epsilon/2.$$

By reducing δ_x if necessary, we may, by Lemma 2.1 in Segers (2022), ensure that

$$T(x + \delta_x \mathbb{B}) \subset T(x) + (\epsilon/2)\mathbb{B}.$$

Also for this reduced δ_x, we have

$$T'(x + \delta_x \mathbb{B}) \subset T(x) + (\epsilon/2)\mathbb{B}.$$

The compact set K is covered by the union of the open balls $x + \delta_x \mathbb{B}^\circ$ for $x \in K$. Let $Z \subset K$ be a finite set such that K is already covered by the union of the balls $z + \delta_z \mathbb{B}$ for $z \in Z$. Put $\mathcal{G} = \bigcap_{z \in Z} \mathcal{G}_z$.

Let $T' \in \mathcal{F}_{\mathrm{mm}}$ be such that $T' \llcorner V \subset \mathcal{G}$. Then $K \subset \bigcup_{z \in Z} (z + \delta_z \mathbb{B}) \subset \operatorname{dom} T'$. For any $x \in K$, we can find $z \in Z$ such that $x \in z + \delta_z \mathbb{B}$ and thus

$$\begin{aligned} T(x) &\in T(z + \delta_z \mathbb{B}) \subset T(z) + (\epsilon/2)\mathbb{B}, \\ T'(x) &\subset T'(z + \delta_z \mathbb{B}) \subset T(z) + (\epsilon/2)\mathbb{B}. \end{aligned}$$

Since $T(z)$ is a singleton, both $T(x)$ and $T'(x)$ are contained in the same ball with radius $\epsilon/2$. It follows that the maximal distance between $T(x)$ and a point in $T'(x)$ is not larger than ϵ. $\square$

One can notice that the proof given in Segers (2022) does not require K to be compact, thus we can reuse it, without doing any modification, in a different context; we only require K to be precompact and T to be singlevalued on K.

Proposition 3.7 (adapted from Proposition 3.6 in Segers (2022))**.** *Let $T \in \mathcal{F}_{\mathrm{mm}}$, let $V \subset \operatorname{dom} T$ be open and let $K \subset V$ be precompact and non-empty. Assume that T is single-valued on K. For every $\epsilon > 0$ there exists an open neighbourhood $\mathcal{G}$ of $T \llcorner V$ in $\mathcal{F}(V \times \mathbb{R}^d)$ such that for all $T' \in \mathcal{F}_{\mathrm{mm}}$ with $T' \llcorner V \in \mathcal{G}$, we have $K \subset \operatorname{dom} T'$ and*

$$\sup_{x \in K} \sup_{y \in T'(x)} |y - T(x)| \leq \epsilon.$$

As a consequence, for any sequence $(T_n)_n$ in $\mathcal{F}_{\mathrm{mm}}$, if $T_n \xrightarrow{v} T$ as $n \to \infty$, then $K \subset \operatorname{dom} T_n$ for all but finitely many n and $\sup_{x \in K} \sup_{y \in T_n(x)} |y - T(x)| \to 0$ as $n \to \infty$.

Since $\mathbb{S}^{d-1}$ is compact, it is clear that the set $\mathbb{S}^{d-1} \cap D_{\bar{\psi}}$ is precompact. Moreover, by definition $\partial \bar{\psi}$ is singlevalued on $\mathbb{S}^{d-1} \cap D_{\bar{\psi}}$. Since $\operatorname{spt} \gamma \subset \partial \bar{\psi}$, we can take $V = \operatorname{int} \operatorname{spt} \mu = \mathbb{R}^d$ which satisfies $V \subset \operatorname{dom} \partial \bar{\psi}$ (see Lemma 2.16). Thus, a direct application of the last theorem yields the existence, for every $\epsilon > 0$ of a neighborhood $\mathcal{G}$ of $\partial\bar{\psi}$ in $\mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d)$ such that for all $T' \in \mathcal{F}_{\mathrm{mm}}$ with $T' \in \mathcal{G}$ we have

$$\mathrm{d}_H(\partial\bar{\psi}(\mathbb{S}^{d-1} \cap D_{\bar{\psi}}), T'(\mathbb{S}^{d-1} \cap D_{\bar{\psi}})) \leq \epsilon.$$

Relying on this, we now prove the announced convergence. Recall that we already know that $\partial\hat{\psi}_{n,t_n}$ weakly converges to $\partial\bar{\psi}$ in $\mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d)$. Using the usual Portmanteau theorem for weak convergence in the metric space $\mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d)$, we can write $\liminf_{n\to\infty} \mathsf{P}(\partial\hat{\psi}_{n,t_n} \in \mathcal{G}) \geq \mathsf{P}(\partial\bar{\psi} \in \mathcal{G}) = 1$. As a consequence, we have

$$\liminf_{n\to\infty} \mathsf{P}(\mathrm{d}_H(\partial\bar{\psi}(\mathbb{S}^{d-1} \cap D_{\bar{\psi}}), T'(\mathbb{S}^{d-1} \cap D_{\bar{\psi}})) \leq \epsilon) = 1$$

and as ϵ was chosen arbitrarily, this finally gives the convergence in probability of $\mathrm{d}_H(\partial\bar{\psi}(\mathbb{S}^{d-1} \cap D_{\bar{\psi}}), T'(\mathbb{S}^{d-1} \cap D_{\bar{\psi}})$ toward 0.

Theorem 3.8.

$$\mathrm{d}_H(\partial\hat{\psi}_{n,t_n}(\mathbb{S}^{d-1} \cap D_{\bar{\psi}}), \partial\bar{\psi}(\mathbb{S}^{d-1} \cap D_{\bar{\psi}})) \xrightarrow{\mathbb{P}} 0$$

4 Old: Center-outward tail quantiles and their estimation

[AM:Almost everything in this section is to be reworked apart from the results about empirical measures convergence and the ones giving conditions for the complement to the center-outward tail quantile regions to have the desired mass.]

To begin with, we explain the rationale for our study. Let $P, Q \in \mathcal{P}(\mathbb{R}^d)$ be regularly varying with common index $\alpha > 0$, common auxiliary function $b(t) = |\cdot|_\# Q([1-1/t, \infty))$ and limit measures $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d) \setminus \{0\}$. Assume that P, Q, μ all vanish on sets of Hausdorff dimension at most $d-1$, and $\operatorname{spt} \mu = \mathbb{R}^d$. By Theorem 1.44, $\Pi_{cm}(P, Q)$ is a singleton whose single element π can be written $\pi = [\mathrm{Id} \times \nabla\psi]_\# P$ where ψ is a closed convex function satisfying $\operatorname{spt} \pi \subset \partial\psi$ and $\nabla\psi$ is uniquely determined P-almost everywhere. By Theorem 2.38(c), we have $\Gamma_{0,cm}(\mu, \nu) = \{\gamma\}$ for some $\gamma \in \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$ which can be written $\gamma = [\mathrm{Id} \times \nabla\bar{\psi}]_\# \mu$ for every closed convex function $\bar{\psi}$ such that $\operatorname{spt} \gamma \subset \partial\bar{\psi}$. The gradient $\nabla\bar{\psi}$ is uniquely determined μ-almost everywhere. Moreover, by Theorem 2.38 we have the convergences

$$\begin{cases} t \operatorname{res} \pi(b(t)\,\cdot\,) \xrightarrow{0} \gamma & \text{in } \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d), \\ b(t)^{-1}\partial\psi(b(t)\,\cdot\,) \xrightarrow{\mathcal{F}} \partial\bar{\psi} & \text{in } \mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d). \end{cases}$$

Let Y be a random vector in $\mathbb{R}^d$ with distribution Q. Recall the usual decomposition introduced in Subsection 1.2. Let $p_t = \mathsf{P}(|Y| > b(t))$. Then for every Borel set A of $\mathbb{R}^d$ and $t > 0$ we have

$$\mathsf{P}(Y \in A) = (1 - p_t)\,\mathsf{P}(Y \in A \mid |Y| \le b(t)) + p_t\,\mathsf{P}(Y \in A \mid |Y| > b(t)).$$

According to the same subsection we can choose $b(t) = F^{\leftarrow}(1 - 1/t)$ where $F^{\leftarrow}$ is the generalized inverse of $F(t) = |\cdot|_\# Q([t, \infty))$. From Remark 1.53 we can define $\mathbb{C}(q) = \mathbb{C}_{P:Q}(q)$ the center-outward quantile region of order $q \in (0, 1)$ with target measure Q and reference measure P instead of the spherical uniform distribution used Hallin et al. (2021) by $\mathbb{C}(q) = \nabla\psi(\mathrm{B}_{0,F^{\leftarrow}(q)})$. This change is motivated by the simple fact that since the latter distribution is not regularly varying, we cannot use it in Theorem 2.38 to obtain the regular variation of the restriction to $(\mathbb{R}^d \times \mathbb{R}^d) \setminus \{0\}$ of the unique cyclically monotone coupling. To benefit from the center-outward quantile regions proposed in Chernozhukov et al. (2017), Hallin et al. (2021) and introduced in Subsection 1.6, for some $t > 0$ we will substitute $\mathbb{C}(1 - 1/t)$ the center-outward quantile region of order $1 - 1/t$ for $\{x : |x| \le b(t)\}$. For every Borel set A of $\mathbb{R}^d$ and $t > 0$ we have the decomposition

$$\mathsf{P}(Y \in A) = \mathsf{P}(Y \in A \cap \mathbb{C}(1 - 1/t)) + \mathsf{P}(Y \in \mathbb{C}(1 - 1/t)^c))\,\mathsf{P}(Y \in A \mid Y \in \mathbb{C}(1 - 1/t)^c).$$

Since both P and Q vanish on sets of Hausdorff dimension at most $d-1$, we can write the identity

$$t\ \mathsf{P}\left(Y \in \mathbb{C}(1-1/t)^c\right)) = t\, P(b(t)\,(\mathbb{R}^d \setminus \mathrm{B}_{0,1})),$$

and as μ vanish on sets of Hausdorff dimension at most $d-1$ too, $t\,\mathsf{P}\left(Y \in \mathbb{C}(1-1/t)^c\right))$ goes to $\mu(\mathbb{R}^d \setminus \mathrm{B}_{0,1}) = 1$ as $n \to \infty$. As a consequence, for the purpose of building an approximation like in Subsection 1.2 and using arguments from there, we are interested in the M_0-convergence of

$$t\ \mathsf{P}\left(Y \in \cdot \cap \mathbb{C}(1-1/t)^c\right) = t\,\pi\left(b(t)\left[\mathbb{R}^d \times (\cdot/b(t) \cap \mathbb{C}(1-1/t)^c/b(t))\right]\right)$$

where $\pi \in \Pi(P,Q)$ is the unique cyclically monotone coupling between P and Q. In order to proceed like in Subsection 1.2, we assume that the M_0-convergence of $t \operatorname{res} \pi(b(t)\cdot)$ toward γ is achieved for some $t^* > 0$ and write

$$t^*\,\pi\left(b(t^*)\left[\mathbb{R}^d \times (\cdot/b(t^*) \cap \mathbb{C}(1/t^*)^c/b(t^*))\right]\right) \approx \gamma\left(\mathbb{R}^d \times [\cdot/b(t^*) \cap \mathbb{C}(1/t^*)^c/b(t^*)]\right).$$

Moreover, since $b(t)^{-1}\partial\psi(b(t)\,\cdot\,)$ $\mathcal{F}$-converges to $\partial\bar{\psi}$, it seems reasonable to approximate $\mathbb{C}(1-1/t^*)/b(t^*)$ by $\nabla\bar{\psi}(\mathrm{B}_{0,1})$.

Therefore, for any Borel set A of $\mathbb{R}^d \setminus \{0\}$ which is bounded away from zero and satisfies $\nu(\partial A) = 0$ we propose the approximation

$$t^*\,\pi\left(b(t^*)\left[\mathbb{R}^d \times (A/b(t^*) \cap \mathbb{C}(1/t^*)^c/b(t^*))\right]\right) \approx \gamma\left(\mathbb{R}^d \times \left[A/b(t^*) \cap \nabla\bar{\psi}(\mathrm{B}_{0,1})^c\right]\right).$$

Writing $p_{t^*} = \mathsf{P}\left(Y \notin \mathbb{C}(1/t^*)\right)$, this finally leads to the approximation

$$\mathsf{P}\left(Y \in A\right) \approx (1-p_{t^*})\,\mathsf{P}\left(Y \in A \mid Y \in \mathbb{C}(1/t^*)\right) + p_{t^*}\,\nabla\bar{\psi}_{\#}\mu\left(A/b(t^*) \cap \nabla\bar{\psi}(\mathrm{B}_{0,1})^c\right).$$

As a consequence, we will be interested in the quantities $\nabla\bar{\psi}\big(\mathrm{B}_{0,(1-q)^{-1/\alpha}}\big)$, $q \in (0,1)$ which appear as limits of center-outward quantile regions through some rescaling. They will prompt us, in Subsection 4.1, to define the concept of center-outward tail quantile functions, contours $\mathcal{C}^\tau(q)$ and regions $\mathbb{C}^\tau(q)$ which extend the idea of their center-outward counterparts developed in Chernozhukov et al. (2017) and Hallin et al. (2021) and recalled in Subsection 1.6 to heavy-tailed distributions in $\mathrm{M}_0(\mathbb{R}^d)$. Still in Subsection 4.1 we derive some of their basic properties including homogeneity and relations between tail quantiles contours and regions, and provide necessary and sufficient conditions for the complement to the tail quantile region of order $q \in (0,1)$ in $\mathbb{R}^d$ to have the expected ν-mass of q. When the support of the spectral measure associated to ν is strictly smaller than the unit sphere, we derive under mild assumptions on ν and μ an exact formula giving the ν-mass of the complement to the tail quantile region in $\mathbb{R}^d$.

In Subsection 4.2 we provide empirical counterparts $\hat{\mu}_{n,t}$, $\hat{\nu}_{n,t}$, $\hat{\gamma}_{n,t}$ to the measures μ, ν, γ where γ lies in $\Gamma_0(\mu, \nu)$ and prove their weak M_0-convergence which was conjectured by Segers (2023) in a talk that he gave at Joint Statistical Meeting in Toronto, Ontario in 2023. The proof is delayed to Appendix B and largely relies on arguments taken from Resnick (2007), Resnick (2008) and adapted to the space M_0 and on the theory developed in Subsection 2.2. The three convergences of the empirical measures will allow us to study the convergence of $\operatorname{spt} \hat{\gamma}_{n,t}$ using results from Subsubsection 2.2.2.

In Subsection 4.3, we focus on the estimation of the center-outward tail quantile contour and regions. Thanks to the homogeneity of both tail quantile contours and regions, and thanks to the relation between tail quantile contours and region, it reduces to the estimation of the center-outward tail quantile contour of order 0 of which we propose an estimator $\tilde{\mathcal{C}}^\tau(0)$. This section is largely devoted to giving some rationale to the proposed estimator $\hat{\mathcal{C}}^\tau(0)$. We introduce an intermediate set $\tilde{\mathcal{C}}^\tau(0)$ we hope to be a good approximation of $\mathcal{C}^\tau(0)$ and let as a conjecture that the identity $\nu(\mathcal{C}^\tau(0) \Delta \tilde{\mathcal{C}}^\tau(0)) = 0$ holds for a target measure ν smooth enough. We prove the $\mathcal{F}$-convergence of some restriction of $\operatorname{spt} \hat{\gamma}_{n,t_n}$ toward some restriction of $\partial\bar{\psi}$—where $\bar{\psi}$ is as introduced at the beginning of the current section—and show that $\hat{\mathcal{C}}^\tau(0)$ can be build using the restriction of $\operatorname{spt} \hat{\gamma}_{n,t_n}$ that we just considered when the conjecture is assumed to be true. Finally, we prove some kind of convergence of the proposed estimator $\hat{\mathcal{C}}^\tau(0)$ toward $\tilde{\mathcal{C}}^\tau(0)$.

Finally, in Subsection 4.4, we apply the theory developed so far to estimate center-outward quantile regions of order $q \in (0, 1)$ close to 1. In Subsubsection 4.4.1, we present our method which relies on the simple decomposition introduced at the beginning of the current section. To build our estimate, we use at the same time the center-outward quantile region $\mathbb{C}(1 - 1/t^*)$, where t^* is as at the beginning of the current section, and a center-outward tail quantile region $\hat{\mathbb{C}}^t(\beta)$, for some $\beta \in (0, 1)$ which will be determined, which allows to extrapolate in direction where there is extreme values. We'll leave the problem of interpolating the contour from the points to a further work. In Subsubsection 4.4.2 we illustrate our method on simple examples built from elliptical distributions, and reproduce some experiments that Hallin et al. (2021) conducted for center-outward quantile contours by substituting mixtures of elliptical distribution with heavy-tailed radial distributions for mixtures of multivariate gaussian distributions.

4.1 General definitions and basic properties

In this first subsection, we introduce the concept of center-outward tail quantiles and discuss some of its basic properties while emphasizing similarities with center-outward quantiles defined in Subsection 1.6.

To extend the idea of center-outward quantiles to heavy-tailed distributions in $\mathrm{M}_0(\mathbb{R}^d)$, one needs to consider a reference distribution different from the spherical uniform distribution used in Chernozhukov et al. (2017) or Hallin et al. (2021) since it has bounded support and hence cannot be regularly varying as defined in Subsection 1.2. Let Q be a regularly varying probability measure on $\mathbb{R}^d$ with index $\alpha > 0$, limit measure ν lying in $\mathrm{M}_0(\mathbb{R}^d) \setminus \{0\}$, and auxiliary function b. Recall from Subsection 1.2 that the limit measures in the definition of regular variation have polar decomposition of the form $\nu_\alpha \otimes H$ where H is a finite measure on the unit sphere and ν_α is the Borel measure on $\mathbb{R}_+$ satisfying $\nu_\alpha([x, \infty)) = x^{-\alpha}$.

Since the spherical uniform distribution used in Chernozhukov et al. (2017) or Hallin et al. (2021) has polar decomposition of the form $\mathrm{Unif}(0,1) \otimes \mathrm{Unif}(\mathbb{S}^{d-1})$, it seems natural to look for a reference distribution with polar decomposition of the form $P = R_\beta \otimes \mathrm{Unif}(\mathbb{S}^{d-1})$ for some $\beta > 0$ and radial distribution R_β on $\mathbb{R}_+$ regularly varying with limit measure in the definition of regular variation ν_β. Here we just replaced the radial uniform distribution on $(0,1)$ in Chernozhukov et al. (2017) or Hallin et al. (2021) by R_β. Using usual polar decomposition arguments, we can show that such P is regularly varying with index $\beta > 0$ and limit measure μ which can be expressed as $\mu = \nu_\beta \times \mathrm{Unif}(\mathbb{S}^{d-1})$ and is therefore symmetric too. In view of Theorem 2.38 a natural choice is $\beta = \alpha$, which is achieved by simply setting $R_\alpha = |\cdot|_\# Q$.

It is clear that $\operatorname{spt} \mu = \mathbb{R}^d$, and μ vanishes on sets of Hausdorff dimension at most $d-1$ since the measure μ is absolutely continuous with respect to d-dimensional Lebesgue measure, and sets of Hausdorff dimension less than d are Lebesgue null sets. Thus a direct application of Theorem 2.38(c) gives that for every $\nu \in \mathrm{M}_0(\mathbb{R}^d)$, we have $\Gamma_{0,cm}(\mu, \nu) = \{\gamma\}$ where

$$\gamma = [\mathrm{Id} \otimes \nabla\bar{\psi}]_\# \mu$$

for some closed convex function $\bar{\psi}$ satisfying $\operatorname{spt} \gamma \subset \partial\bar{\psi}$. Moreover, the gradient $\nabla\bar{\psi}$, which is determined uniquely μ-almost everywhere, satisfies $\operatorname{res} \nabla\bar{\psi}_\# \mu = \nu$ and is homogenous in the sense that

$$\nabla\bar{\psi}(\lambda x) = \lambda \nabla\bar{\psi}(x), \qquad \forall \lambda > 0$$

μ-almost everywhere. More precisely, the complement of the set of $x \in \mathbb{R}^d$ on which the identity is true for all $\lambda > 0$ is a μ-null set.

Remark 4.1. *There is a connection between our reference distribution P and the spherical uniform distribution U used in Chernozhukov et al. (2017), Hallin et al. (2021) and introduced in Subsection 1.6. Let F_R denote the cumulative distribution function of R_α. Assume Q vanishes on sets of Hausdorff dimension at most $d-1$, then F_R is continuous. Following the discussion at the beginning of Subsection 1.6 and using notation from there, let $\tilde{T}_F(x) = \int_0^x F(s)ds$ and $T_F(x) = \tilde{T}_F(|x|)$ for every non-decreasing map $F : x \in \mathbb{R}^d \to \mathbb{R}^d$. Such a map has derivative $\nabla T_F(x) = xF(|x|)/|x|$. Moreover, as $|\cdot|$ is convex and $\tilde{T}_F$ is non-decreasing and convex—since it has non-decreasing derivatives—, $T_{F_R^{\leftarrow}}$, T_{F_R} are convex functions and $\boldsymbol{Q}_\pm = \nabla T_{F_R^{\leftarrow}}$, $\boldsymbol{F}_\pm = \nabla T_{F_R}$ are respectively the center-outward quantile and distribution functions. Applying Theorem 1.44, $[\mathrm{Id} \times \boldsymbol{Q}_\pm]_\# U$ and $[\mathrm{Id} \times \boldsymbol{F}_\pm]_\# P$ are the unique cyclically monotone zero-coupling respectively between U and P, P and U. We mention this connection here even if we didn't investigate further the link between the center-outward tail quantiles that we will introduce in the current subsection and the center-outward quantiles as defined by Hallin et al. (2021) using the spherical uniform distribution.*

Since limit measures in the definition of regular variation have infinite mass on neighborhoods of the origin, we need to adopt a definition slightly different from Definition 1.56 in order to define $\mathbb{C}_\nu^\tau(q)$ the center-outward tail quantile region of order q of ν. We consider the mass left in the tails and expect $\nu(\mathbb{R}^d \setminus \mathbb{C}_\nu^\tau(q))$ to be equal to $1-q$ while $\nu(\mathbb{C}_\nu^\tau(q))$ should have infinite value. If ν were a probability measure, it would be equivalent, under appropriate assumptions, to $\nu(\mathbb{C}_\nu^\tau(q)) = q$.

To define center-outward tail quantile contours and regions we consider the family $\nabla\bar{\psi}(\mathrm{B}_{0,r}) \setminus \{0\}, r > 0$ instead of $\nabla\psi(\mathrm{B}_{0,r}), r > 0$ for center-outward quantiles. Note that by definition, for $s > 0$, $\mu(\mathbb{R}^d \setminus \mathrm{B}_{0,r}) = s$ if and only if $r = s^{-1/\alpha}$.

Definition 4.2 (Center-outward tail quantile function, region and contour)**.** *Call center-outward tail quantile function of $Q \in \mathcal{P}(\mathbb{R}^d)$ the μ-a.e. unique element $\boldsymbol{Q}_\pm = \nabla\bar{\psi}$ given in Theorem 2.38.*

Let $D_{\bar{\psi}}$ denote the sets of points x where $\bar{\psi}$ is differentiable, i.e., $\partial\bar{\psi}(x)$ is a singleton. For every $q \in (0,1)$, define the center-outward tail quantile region of order q of ν as

$$\mathbb{C}_\nu^\tau(q) = \nabla\bar{\psi}(\mathrm{B}_{0,(1-q)^{-1/\alpha}}) = \{\nabla\bar{\psi}(x) \,:\, x \in D_{\bar{\psi}},\, |x| \leq (1-q)^{-1/\alpha}\}$$

and the center-outward tail quantile contour of order q of ν as

$$\mathcal{C}_\nu^\tau(q) = \nabla\bar{\psi}((1-q)^{-1/\alpha}\mathbb{S}^{d-1}) = \{\nabla\bar{\psi}(x) \,:\, x \in D_{\bar{\psi}},\, |x| = (1-q)^{-1/\alpha}\}.$$

Remark 4.3. *In Definition 4.2, one may have been tempted to define the center-outward tail quantile regions and contour as*

$$\mathbb{C}_\nu^\tau(q) = \mathrm{rge}\left(\partial\bar{\psi} \llcorner \mathrm{B}_{0,(1-q)^{-1/\alpha}}\right).$$

and

$$\mathcal{C}^\tau_\nu(q) = \operatorname{rge}\left(\partial\bar{\psi}\llcorner\left\{x \in \mathbb{R}^d \ : \ |x| = (1-q)^{-1/\alpha}\right\}\right)$$

but it would be less convenient in practice. Indeed, as the support $\operatorname{spt}\gamma$ *of the cyclically monotone zero-coupling* γ *is a subset of* $\partial\bar{\psi}$ *but is not necessarily equal to it, we may lack information about the map* $\partial\bar{\psi}$ *on the whole space when working with empirical data and an empirical version of the zero-coupling. To circumvent this problem we choose to rather consider smaller sets that we hope will have the same* ν*-mass since the restrictions of both* $\operatorname{spt}\gamma$ *and* $\partial\bar{\psi}$*—seen as a multivalued map—to* $D_{\bar{\psi}}$ *coincide and* $\mu(D^{\mathrm{c}}_{\bar{\psi}}) = 0$ *by Theorem 1.29.*

Remark 4.4. *Since* $\nabla\bar{\psi}$ *is homogenous, we have* $\nabla\bar{\psi}(\mathrm{B}_{0,s}) = s\nabla\bar{\psi}(\mathrm{B}_{0,1})$ *for any* $s > 0$*. As a consequence,* $\mathcal{C}^\tau_\nu$ *and* $\mathbb{C}^\tau_\nu$ *inherit some kind of homogeneity, in the sense that we have*

$$\mathcal{C}^\tau_\nu(q) = (1-q)^{-1/\alpha}\,\mathcal{C}^\tau_\nu(0), \qquad \mathbb{C}^\tau_\nu(q) = (1-q)^{-1/\alpha}\,\mathbb{C}^\tau_\nu(0)$$

and for any $\lambda > 0$*,*

$$\mathcal{C}^\tau_\nu(\lambda q) = \left(\frac{1-\lambda q}{1-q}\right)^{-1/\alpha}\mathcal{C}^\tau_\nu(q), \qquad \mathbb{C}^\tau_\nu(\lambda q) = \left(\frac{1-\lambda q}{1-q}\right)^{-1/\alpha}\mathbb{C}^\tau_\nu(q).$$

This property will be useful for extrapolation and allows one only to look for a good estimation of $\mathcal{C}^\tau_\nu(0)$ *and* $\mathbb{C}^\tau_\nu(0)$*. An intuitive relation between the last two quantities will be shown later on.*

For these quantities to be relevant, one would naturally expect from center-outward tail quantile regions and contours of order $q \in (0,1)$ to satisfy :

(a) $\nu(\mathbb{R}^d \setminus \mathbb{C}^\tau(q)) = q$, i.e., the tail quantile region of order q has probability content q

(b) $\mathcal{C}^\tau(q) = \partial\mathbb{C}^\tau(q)$

These two conditions are similar to those proposed in Subsection 1.6 for center-outward quantiles. The next result, which gives a simple relation between $\mathcal{C}^\tau_\nu(q)$ and $\mathbb{C}^\tau_\nu(q)$, is immediate since $\nabla\bar{\psi}$ is homogenous.

Lemma 4.5. *For every* q *in* $(0,1)$ *we have*

$$\begin{aligned}\mathbb{C}^\tau_\nu(q) &= \bigcup_{y \in \mathcal{C}^\tau_\nu(q)} [0, y] \\ &= \{[0, \nabla\bar{\psi}(x)] \ : \ |x| = (1-q)^{-1/\alpha}\} \\ &= (1-q)^{-1/\alpha}\{[0, \nabla\bar{\psi}(x)] \ : \ |x| = 1\}.\end{aligned}$$

It is clear that $\mathcal{C}_\nu^\tau(q) \subset \mathbb{C}_\nu^\tau(q)$ but in general we do not necessarily have $\mathcal{C}_\nu^\tau(q) = \operatorname{bnd} \mathbb{C}_\nu^\tau(q)$ as shown by the following example.

Example 4.6. *We define the map*

$$\bar{\psi} : (x, y) \in \mathbb{R}^2 \longmapsto \begin{cases} \dfrac{x^2 + y^2}{2} & \text{if } x \geq 0 \\ \dfrac{(|x| + |y|)^2}{2} & \text{if } x < 0 \end{cases} \in \mathbb{R}$$

which is continuous and convex—it is immediate since both $|x|^2 + |y|^2$ and $(|x| + |y|)^2$ are convex on $\mathbb{R}^2$ and $|x|^2 + |y|^2 \leq (|x| + |y|)^2$. We define the function sign *on a vector by* $\operatorname{sign}(x, y) = (\operatorname{sign} x, \operatorname{sign} y)$*. A simple computation yields*

$$\nabla \bar{\psi}(x, y) = \begin{cases} (x, y) & \text{if } x > 0, y \neq 0 \\ (|x| + |y|) \operatorname{sign}(x, y) & \text{if } x < 0, y \neq 0 \end{cases}$$

for every (x, y) in $\mathbb{R}^2$ such that $x \neq 0$, $y \neq 0$. Moreover it is clear that $\nabla \bar{\psi}((\lambda x, \lambda y)) = \lambda \bar{\psi}(x, y)$ for each $\lambda > 0$, and, by Theorem 1.37 (b), $\partial \bar{\psi}$ is maximal cyclically monotone as the subgradient of a closed convex function.

Let $\nu = \nabla \bar{\psi}_{\#} \mu$. Such ν puts mass on the whole half space $\{(x, y) \in \mathbb{R}^2 : x > 0\}$ and on rays starting at the origin excluded and passing through $(-1, 1)$ and $(-1, -1)$ respectively. Through computations one can show, for $r > 0$,

$$\nabla \bar{\psi}(r\mathbb{S}^{d-1}) = \big(r\mathbb{S}^{d-1} \cap \{(x, y) : x > 0\}\big) \cup ((0, 0), (-1, 1)) \cup ((0, 0), (-1, -1))$$

where (a, b) denotes $\{a + \lambda(b - a) : 0 < \lambda < 1\}$ for every a, b in $\mathbb{R}^2$. As a consequence, $\nabla \bar{\psi}(r\mathbb{S}^{d-1})$ is not a closed set, hence it cannot be the boundary of any set. As a consequence, the equality $\mathcal{C}_\nu^\tau(q) = \operatorname{bnd} \mathbb{C}_\nu^\tau(q)$ does not hold.

Unfortunately, it is possible for $\mathcal{C}_\nu^\tau(q) \cap \{0\}$ to be non-empty for every $q > 0$, implying in some cases that $\nu(\mathbb{C}_\nu^\tau(q)^c)$ is strictly smaller than $\mu(\mathbb{R}^d \setminus \mathrm{B}_{0,q^{-1/\alpha}})$. It will be the case when some mass needs to be sent to zero like in Example 2.10. As a consequence, we have no guarantee in general that the tail region associated to the center-outward tail quantile has the mass we would expect it to have and introduce further assumptions. A similar discussion was conducted in Section 1.6 about center-outward quantile regions.

The next lemma shows that for everything to go as expected, we need the support of the measure ν to be large enough relative to the one of μ. Recall from Subsection 2.3.2 that we write $H(x), H_+(x)$ and $H_-(x)$ for the hyperplan, positive and negative half-spaces induced by x; $H(x) = \{y : \langle x, y \rangle = 0\}, H_{+(-)}(x) = \{y : \langle x, y \rangle > (<) 0\}$.

Lemma 4.7. *Let $P, Q \in \mathcal{P}(\mathbb{R}^d)$ be regularly varying with common index $\alpha > 0$, common auxiliary function b and limit measures $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d) \setminus \{0\}$. Assume μ vanishes on sets of Hausdorff dimension at most $d-1$, and $\operatorname{spt} \mu = \mathbb{R}^d$. We consider the set $\chi = \{x \in \mathbb{R}^d \,:\, \operatorname{spt} \nu \cap \operatorname{cl} H_+(x) = \{0\}\}$. If $\nu(\mathbb{R}^d \setminus \mathbb{C}^\tau_\nu(q)) = 1 - q$ for every q lying in $(0,1)$, then $\mu(\chi) = 0$.*

Note that since μ is a limit measure in the definition of regular variation with index $\alpha > 0$ and μ satisfies $\operatorname{spt} \mu = \mathbb{R}^d$, it is clear that μ has polar decomposition

$$\mathrm{POLAR}\, \mu = \nu_\alpha \otimes \mathrm{Unif}(\mathbb{S}^{d-1}).$$

The assumption $\operatorname{spt} \mu = \mathbb{R}^d$ is used in the proof below but one may deal without it.

Remark 4.8. *The reader should note that:*

- *although formulated as a condition on the mass given by μ to a set, the assumption only concerns ν since μ is fixed;*
- *the condition on $\operatorname{spt} \nu$ simply means that no mass from μ is sent to the origin;*

The following example give a counter-example to the sufficiency of the condition in the previous theorem and shows that more assumptions are required if one wants the condition (a) to hold, i.e., $\nu(\mathbb{R}^d \setminus \mathbb{C}^\tau(q)) = q$ for every $q \in (0,1)$.

Example 4.9. *Let f be a non-decreasing convex function from $\mathbb{R}$ to itself with continuous derivative. We define*

$$\psi : (x_1, \ldots, x_d) \in \mathbb{R}^d \longmapsto f\left(\sum_{k=1}^{d} \frac{x_k^2}{|x_k|} \right).$$

It is clear that ψ is convex and a simple computation yields

$$\nabla \psi(x_1, \ldots, x_d) = f'\left(\sum_{k=1}^{d} \frac{x_k^2}{|x_k|} \right) \left(\frac{x_1}{|x_1|}, \ldots, \frac{x_d}{|x_d|} \right).$$

Let ν denote a symmetric distribution of the form $\nu = \nu_\alpha \times \mathrm{Unif}(\mathbb{S}^{d-1})$ for some $\alpha > 0$, and set $Q = \nabla \psi_\# U$. Note that Q exactly puts mass on the rays starting from the origin and passing through some point $(\delta_1, \ldots, \delta_d)$ where $\delta_i \in \{-1, 1\}$ for each i. Therefore it is easy to see that the assumption in Theorem 4.7 is satisfied.

Let us choose the exponential function for f and $d = 2$. We claim that $\nu(\mathbb{C}^\tau_\nu(r)) < 1 - r$ for every $r > 0$. For some $s > 0$ we take $(x, y) \in \mathrm{B}_{0,s}$. Since $x^2 + y^2 \leq s$, we *have* $|x|+|y| \le s$ *whence* $[\nabla\psi]^{-1}\nabla\psi(\mathrm{B}_{0,s}) = \{(x,y)\in\mathbb{R}^2 \,:\, |x|+|y|\le r\}$ *contains strictly* $\mathrm{B}_{0,s}$ *and* $[\nabla\psi]^{-1}\nabla\psi(\mathrm{B}_{0,s})\setminus \mathrm{B}_{0,s}$ *has non-empty interior* I*. To conclude, it suffices to lower-bound* $P(I)$ *by* $P'(I)>0$ *where* P' *is some suitable rescaling of the spherical uniform distribution on* $s^2\mathbb{S}^{d-1}$*. We have just proved that* $\nu(\mathbb{C}^\tau_\nu(r)^c) < \mu(\mathrm{B}^c_{0,(1-r)^{-1/\alpha}}) = q$*.*

The following result is similar to Theorem 1.59 for center-outward quantile regions in Section 1.6. It gives a sufficient condition for the mass left in the complement of the center-outward tail quantile of order $q\in(0,1)$ to be equal to $1-q$ as required in condition (a) introduced earlier.

Theorem 4.10. *Let* $P,Q\in\mathcal{P}(\mathbb{R}^d)$ *be regularly varying with common index* $\alpha>0$*, common auxiliary function* b *and limit measures* $\mu,\nu\in\mathrm{M}_0(\mathbb{R}^d)\setminus\{0\}$*. Assume* μ *vanishes on sets of Hausdorff dimensoin at most* $d-1$*, and* $\operatorname{spt}\mu=\mathbb{R}^d$*. Let* $q>0$*. The complement of the center-outward tail quantile region* $\mathbb{C}^\tau(q)$ *in* $\mathbb{R}^d$ *has* ν*-mass content at most* $1-q$*. If we further assume that*

(a) ν *vanishes on sets of Hausdorff dimension at most* $d-1$

(b) $\mu(\chi)=0$ *where* $\chi=\left\{x\in\mathbb{R}^d \,:\, \operatorname{spt}\nu\cap\operatorname{cl}H_+(x)=\{0\}\right\}$

then there is equality in the last statement, i.e., $\nu(\mathbb{R}^d\setminus\mathbb{C}^\tau_\nu(q))=1-q$ *for every* q *lying in* $(0,1)$*. In particular, if* $\operatorname{spt}\mu\cap\chi$ *is empty, the assumption (b) is automatically satisfied.*

The following theorem gives the value of $\nu(\mathbb{C}^\tau(q)^c)$ when assumption (b) in Theorem 4.10 is not fulfilled.

Theorem 4.11. *Let* $P,Q\in\mathcal{P}(\mathbb{R}^d)$ *be regularly varying with common index* $\alpha>0$*, common auxiliary function* b *and limit measures* $\mu,\nu\in\mathrm{M}_0(\mathbb{R}^d)\setminus\{0\}$*. Assume both* μ *and* ν *vanish on sets of Hausdorff dimension at most* $d-1$*, and* $\operatorname{spt}\mu=\mathbb{R}^d$*. We write* $\mathrm{Unif}(\mathbb{S}^{d-1})$ *for the uniform distribution on the unit sphere of* $\mathbb{R}^d$ *and let* $p=\mathrm{Unif}(\mathbb{S}^{d-1})(S)$ *where* $S=\mathbb{S}^{d-1}\cap[\partial\bar{\psi}]^{-1}(\{0\})$*. Then we have*

$$\nu(\mathbb{C}^\tau(q)^c) = (1-p)(1-q).$$

Above theorem should be ckecked

Remark 4.12. *Thanks to the last theorem, when assumption (b) about the support of the measure* ν *in Theorem 4.10 is not fulfilled, we can still use center-outward tail quantile regions to study the distribution* ν *in practice. One just needs to estimate the quantity* p *in addition to* $\mathcal{C}^\tau(0)$.

In fact we can still use the center-outward quantile tail region in practice for some applications even if we don't know the quantity p*. For applications like risk-management, e.g. building a dam, having a quantile region too big might not be a problem since it will not induce a risk of failure—the dam would nevertheless be more expansive to build. However, when working on anomaly detection, we would like to have our quantile region of desired order to be as small as possible. If our region is too large, we can expect to fail to detect some anomalies which can turns out to have huge consequences, for example in structure monitoring*

The situation is comparable to the one for ordinary quantiles, $Q(p) = \inf\{x \in \mathbb{R} : F(x) \geq p\}$ *for* $0 < p < 1$ *and the cumulative distribution function* $F(x) = P(X \leq x)$ *of a random variables* X *with values in* $\mathbb{R}$*. Then always* $F(Q(p)) \geq p$ *and thus* $P[X > Q(p)] = 1 - F(Q(p)) \leq 1 - p$*. There can be strict inequality if* $Q(p)$ *is an atom of* X*, i.e.,* $P[X = Q(q)] > 0$*.*

To conclude, the following remark, which is similar to Remark 1.60 for center-outward quantile regions, propose a way to define center-outward tail quantile regions when one does not assume μ vanishes on sets of Hausdorff dimension at most $d-1$.

Remark 4.13. *When the assumption made on* ν *in the previous theorem does not hold, a similar definition of the center-outward tail quantile region can be given as follows. Let* q *live in* $\mathbb{R}_+$*. We define* $\mathbb{C}^t_{general}(q)$ *the general center-outward quantile of order* q *as*

$$\mathbb{C}^t_{general}(q) = \bigcap_{\alpha \in \mathrm{E}} \mathbb{C}^\tau(\alpha)$$

where $\mathrm{E} = \{\alpha > 0 \, : \, \nu(\mathbb{R}^d \setminus \mathbb{C}^\tau(\alpha)) \leq 1 - q\}$.

Since the family $\mathbb{C}^\tau(\alpha), \alpha \in \mathbb{R}_+ \setminus \{0\}$ *is nested, we could alternatively set* $\mathbb{C}^\tau_{general}(q) = \mathbb{C}^\tau(\alpha^*)$ *for* α^* *lying in* $\arg\min\{\alpha \, : \, \nu(\mathbb{C}^\tau(\alpha)) \leq 1 - q\}$*. It may be simpler to compute in practice but at the cost of nestedness. When the assumption of Theorem 4.7 is satisfied, all these notions are the same.*

4.2 (To keep) Empirical measures

In this subsection, we introduce empirical measures proposed in Segers (2023) for which we prove weak convergence in M_0. Let $Q \in \mathcal{P}(\mathbb{R}^d)$ be regularly varying with index $\alpha > 0$, auxiliary function b and limit measure $\nu \in \mathrm{M}_0(\mathbb{R}^d) \setminus \{0\}$. Assume

Q vanishes on sets of Hausdorff dimension at most $d-1$. As proposed above in Subsection 4.1, we choose for reference distribution P given by its polar decomposition $|\cdot|_{\#}Q \otimes \mathrm{Unif}(\mathbb{S}^{d-1})$ which is regularly varying with index $\alpha > 0$, auxiliary function b and limit measure $\mu \in \mathrm{M}_0(\mathbb{R}^d) \setminus \{0\}$ with polar decomposition $\nu_\alpha \otimes \mathrm{Unif}(\mathbb{S}^{d-1})$. Recall from Subsection 4.1 that both P and μ vanishes on sets of Hausdorff dimension at most $d-1$ and have support $\mathbb{R}^d$.

We consider $X_n, n \geq 0$ an iid sequence from the reference distribution P, and $U_n, n \geq 0$ an iid sequence from the distribution $\mathrm{Unif}(\mathbb{S}^{d-1})$ and independent of $(X_n)_{n\geq 1}$. Let $\hat{b}_n(t)$ be some estimator of $b(t)$ such that $\hat{b}(t)/b(t)$ converges to 1 in probability. For example, on may let $\hat{b}_n(t)$ be the empirical scaling function defined by $\hat{b}_n(t) = \hat{F}_n^{\leftarrow}(1-1/t)$ where $\hat{F}_n$ is the usual empirical distribution function of $\{|X_i|\}_{i=1}^n$. Thus we define the empirical measures associated to μ and ν by

$$\hat{\mu}_{n,t} = \frac{t}{n}\sum_{i=1}^n \delta_{\frac{|X_i|U_i}{\hat{b}_n(t)}}$$
$$\hat{\nu}_{n,t} = \frac{t}{n}\sum_{i=1}^n \delta_{\frac{X_i}{\hat{b}_n(t)}}.$$

It is easy to see that μ, ν satisfy the condition of Theorem 2.38(c), so $\Gamma_{0,cm}(\mu,\nu)$ is a singleton whose single element γ satisfies

$$\gamma = [\mathrm{Id} \otimes \nabla\bar{\psi}]_{\#}\mu$$

where $\bar{\psi}$ is a closed convex function such that $\operatorname{spt}\gamma \subset \partial\bar{\psi}$ whose gradient $\nabla\bar{\psi}$ is uniquely defined μ-almost everywhere.

We introduce now an empirical measure associated to the cyclically monotone zero-coupling γ. Let $\mathcal{S}_n$ denote the family of permutations of $\{1,\ldots,n\}$. For each $n \geq 1$ we choose

$$\hat{\sigma}_n \in \operatorname*{arg\,min}_{s\in\mathcal{S}_n} \sum_{i=1}^n \left|X_{s(i)} - |X_i|\, U_i\right|^2.$$

We can then define an empirical zero-coupling measure as

$$\hat{\gamma}_{n,t} = \frac{t}{n}\sum_{i=1}^n \delta\left\{\frac{|X_i|U_i}{\hat{b}_n(t)}, \frac{X_{\hat{\sigma}_n(i)}}{\hat{b}_n(t)}\right\}.$$

Both $\hat{\nu}_{n,t}$ and $\hat{\mu}_{n,t}$ give to each random point the same mass, whence by an argument relying on Remark 1.43 we see that has $\hat{\gamma}_n$ cyclically monotone random support.

Remark 4.14. *Note that defined as above, the random permutation* $\hat{\sigma}_n$ *also satisfies*

$$\hat{\sigma}_n \in \operatorname*{arg\,min}_{s \in \mathcal{S}_n} \sum_{i=1}^{n} \left| \frac{X_{\hat{\sigma}_n(i)}}{\hat{b}_n(t)} - \frac{|X_i|U_i}{\hat{b}_n(t)} \right|^2 .$$

It is clear that the random scaling $\hat{b}_n(t)$ *can be omitted in the definition of the permutation* $\hat{\sigma}_n$*, so the estimated zero-coupling measure* $\hat{\gamma}_{n,t}$ *only depends on* t *through scaling, but not through the permutation* $\hat{\sigma}_n$*. In fact,* $\hat{\gamma}_{n,t}$ *is a scaled version of the cyclically monotone coupling of the empirical distributions of* $\{X_i\}_{i=1}^n$ *and* $\{|X_i|U_i\}_{i=1}^n$*. In practice, this implies that we will only need to solve numerically the discrete Monge problem once.*

The following theorem will be useful when used together with Theorem 2.24 since it will provide us the convergence of some restriction of a sequence of maximal monotone sets T_n satisfying $\operatorname{spt} \hat{\gamma}_{n,t_n} \subset T_n$ for any n.

Theorem 4.15. *For every sequence* $t_n, n \geq 0$ *such that* $t_n/n \to 0$ *as* $n \to \infty$ *we have the weak convergences*

(a) $\hat{\mu}_{n,t_n} \xrightarrow{w} \mu$ *in* $\mathrm{M}_0(\mathbb{R}^d)$*;*

(b) $\hat{\nu}_{n,t_n} \xrightarrow{w} \nu$ *in* $\mathrm{M}_0(\mathbb{R}^d)$*;*

(c) $\hat{\gamma}_{n,t_n} \xrightarrow{w} \gamma$ *in* $\mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$*.*

The proofs of the first two convergences are largely independent of the theory we have developed so far while the one of convergence (c) relies on the convergences (a), (b) together with results from Subsection 2.2 dealing with sequences of zero-coupling.

4.3 Estimation of $\mathcal{C}^\tau(0)$

In this subsection, we propose an estimator of $\mathcal{C}^\tau(0)$ under the same assumptions as in Subsection 4.2, and provide some rationale for it. We are still working on a proof of its consistency. Thanks to the homogeneity of both $\mathcal{C}^\tau(q)$ and $\mathbb{C}^\tau(q)$ from Remark 4.4, and the simple link between $\mathcal{C}^\tau(q)$ and $\mathbb{C}^\tau(q)$ given in Lemma 4.5, it suffices to estimate $\mathcal{C}^\tau(0)$ in order to recover both $\mathcal{C}^\tau(q)$ and $\mathbb{C}^\tau(q)$, $q \in (0,1)$.

Since $\hat{\gamma}_{n,t_n}, n \geq 1$ is a sequence of cyclically monotone zero-couplings, there exists a sequence $\hat{\psi}_{n,t_n}, n \geq 1$ of random closed convex functions such that $\operatorname{spt} \hat{\gamma}_{n,t_n} \subset \partial \hat{\psi}_{n,t_n}$ for each $n \geq 1$. Let $V = \operatorname{int} \operatorname{spt} \mu = \mathbb{R}^d$. Since μ vanishes on sets of Hausdorff dimension at most $d-1$ and $\operatorname{spt}\mu = \mathbb{R}^d$, by Theorem 2.38 $\Gamma_{0,cm}(\mu,\nu)$ is a singleton and Theorem 2.24 yields

$$\partial\hat{\psi}_{n,t_n} \xrightarrow{w} \partial\bar{\psi} \text{ in } \mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d) \text{ i.e., } \partial\hat{\psi}_{n,t_n} \xrightarrow{g} \partial\bar{\psi} \text{ in probability.}$$

This result seems appealing since we used $\partial\bar{\psi}$ to define the center-outward tail quantile regions and contours. However, in practice it has no utility since we don't know $\partial\hat{\psi}_{n,t_n}$ but only the support of $\hat{\gamma}_{n,t_n}$ which is cyclically monotone but not maximal cyclically monotone and is given by

$$\operatorname{spt}\hat{\gamma}_{n,t_n} = \left\{ \left(\frac{|X_i|U_i}{\hat{b}_n(t_n)}, \frac{X_{\hat{\sigma}_n(i)}}{\hat{b}_n(t_n)} \right) \ : \ i = 1, \ldots, n \right\}.$$

Using only information contained in $\operatorname{spt}\hat{\gamma}_{n,t_n}$, we suggest to estimate $\mathcal{C}^\tau(0)$ by

$$\hat{\mathcal{C}}_n^\tau(0) = \left\{ \frac{X_{\hat{\sigma}_n(i)}}{|X_i|} \ : \ i \in \{1, \ldots, n\}, \delta_n < |X_i| < \epsilon_n \right\}.$$

where $\epsilon_n, \delta_n > 0$ are hyperparameters one should choose empirically, bearing in mind the remark below.

Remark 4.16 (Choice of the hyperparameters)**.** *In practice it may be appropriate to choose the parameters $\epsilon_n, \delta_n > 0$ according to the data we have and following the guidelines below.*

(a) One should take δ_n large enough so that only values of $|X_i|$ considered to be extremes are selected; e.g. we could take ϵ_n such that only the 10 percent of the i with the largest $|X_i|$ are used in the estimation.

(b) As suggested in Segers (2023) we should not use the most extreme values in order to reduce the variance of our estimator. In some sense, we should consider some fraction of the most extreme data to be unreliable; e.g. we could choose ϵ_n so that we ignore the 5 percent of the indices i's with the largest $|X_i|$.

The rest of this section is devoted to giving some rationale for this estimator, even if we don't have a proof of his consistency for now. To begin with, we prove the convergence, in some sense, of the support $\operatorname{spt}\hat{\gamma}_{n,t_n}$ of the empirical cyclically monotone zero-coupling between μ and ν— instead of $\partial\hat{\psi}_{n,t_n}$—, toward $\partial\bar{\psi}$.

From Theorem 1.29 and Remark 2.19, there exists a Borel set A of $\mathbb{R}^d \setminus \{0\}$ satisfying $A \subset \operatorname{proj}_1 \operatorname{spt}\gamma$ such that $\nabla\bar{\psi}$ is defined on A and $\mu(A^c) = 0$, where the complement is taken in $\mathbb{R}^d \setminus \{0\}$. One may simply take $A = D_{\bar{\psi}} \cap (\operatorname{proj}_1 \operatorname{spt}\gamma)$ for $D_{\bar{\psi}}$ the set of points where $\nabla\bar{\psi}$ is defined. In the next theorem, we will use $B = \operatorname{int} A$. One may hope that A is a μ-continuity set, in which case $\mu(B^c) = 0$.

Theorem 4.17. *Under the assumptions introduced at the beginning of the current subsection, there exists an open subset B of $\mathbb{R}^d$ on which $\nabla\bar{\psi}$ is defined, and such that*

$$\operatorname{spt}\hat{\gamma}_{n,t_n} \llcorner B \xrightarrow{w} \partial\bar{\psi} \llcorner B \quad \textit{in } \mathcal{F}(B \times \mathbb{R}^d).$$

Remark 4.18. *For the theorem above, we needed to introduce the open set $B = \operatorname{int} A$ because the map $F \in \mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d) \longmapsto F \llcorner A \in \mathcal{F}(A \times \mathbb{R}^d)$ is not continuous in general for non-open A. However, when A is open, the latter map is continuous. We refer to Proposition 12 in Segers (2022) for details.*

Let's consider the function

$$\theta : (a, b) \in (\mathbb{R}^d \setminus \{0\}) \times \mathbb{R}^d \longmapsto \left(a, \frac{b}{|a|}\right) \in (\mathbb{R}^d \setminus \{0\}) \times \mathbb{R}^d.$$

It is clearly an homeomorphism, whence $\theta(F) \in \mathcal{F}\left((\mathbb{R}^d \setminus \{0\}) \times \mathbb{R}^d\right)$ for every $F \in \mathcal{F}\left((\mathbb{R}^d \setminus \{0\}) \times \mathbb{R}^d\right)$. Seen as a map $\mathcal{F}\left((\mathbb{R}^d \setminus \{0\}) \times \mathbb{R}^d\right) \to \mathcal{F}\left((\mathbb{R}^d \setminus \{0\}) \times \mathbb{R}^d\right)$, θ is continuous since for every family of sets $F_n \in \mathcal{F}\left((\mathbb{R}^d \setminus \{0\}) \times \mathbb{R}^d\right), n \geq 1$, a sequence $y_n := \theta(x_n) \in \theta(F_n), n \geq 1$ converges along some infinite subset N of $\mathbb{N}$ if and only if $x_n = \theta^{-1}(y_n)$ converges along N as $n \to \infty$. Continuity of θ together with the homogeneity of $\partial\bar{\psi}$ allows one to prove the following corollary.

Corollary 4.19. *Under the assumptions introduced at the beginning of the current subsection, there exists an open subset B of $\mathbb{R}^d$ on which $\nabla\bar{\psi}$ is defined, and such that we have the $\mathcal{F}$-convergence*

$$\theta(\operatorname{spt}\hat{\gamma}_{n,t_n}) \llcorner B \xrightarrow{w} \theta(\partial\bar{\psi}) \llcorner B \quad \textit{in } \mathcal{F}(B \times \mathbb{R}^d)$$

where

$$\operatorname{rge}\left(\theta(\partial\bar{\psi}) \llcorner B\right) = \operatorname{proj}_2 \left\{(x, \nabla\bar{\psi}(x)) \ : \ x \in B \cap \{|x| = 1\}\right\}.$$

Using arguments similar to those used in the proof of Corollary 4.19, we derive the identity

$$\mathcal{C}^\tau(0) = \nabla\bar{\psi}(\mathbb{S}^{d-1} \cap D_{\bar{\psi}}) = \operatorname{rge}\left(\theta(\partial\bar{\psi}) \llcorner D_{\bar{\psi}}\ \right)$$

where $D_{\bar{\psi}}$ is the set of points $x \in \mathbb{R}^d$ such that $\partial\bar{\psi}(x)$ is a singleton, i.e., where $\nabla\bar{\psi}$ is defined. To use the convergence in Corollary 4.19 we introduce the set $\tilde{\mathcal{C}}^\tau(0)$ defined by

$$\tilde{\mathcal{C}}^\tau(0) = \operatorname{rge}(\theta(\partial\bar{\psi}) \llcorner B) \subset \operatorname{rge}(\theta(\partial\bar{\psi}) \llcorner D_{\bar{\psi}}) = \mathcal{C}^\tau(0)$$

and toward wich we will prove some kind of convergence of an estimator.

Remark 4.20. *Note that since $\tilde{\mathcal{C}}^\tau(0)$ is not necessarily closed, there is no reason for a convergence in a Fell space to happen. In fact, $\mathcal{C}^\tau(0)$ is not closed either. We refer to Example 4.6 for an example.*

Recall from the discussion at the beginning of the current subsubsection that $A = D_{\bar\psi} \cap (\operatorname{proj}_1 \operatorname{spt}\gamma)$ and $B = \operatorname{int} A$. Note that we do not necessarily have $\mu(\partial A) = 0$, thus we can expect in practice to have some empirical points lying in ∂A. Moreover, we will clearly have $\tilde{\mathcal{C}}^\tau(0) \subset \mathcal{C}^\tau(0)$ but not $\nu\left(\mathcal{C}^\tau(0) \setminus \tilde{\mathcal{C}}^\tau(0)\right) = 0$ in general, although it would be useful. However, when $\nabla\bar\psi$ is defined on the whole set $\operatorname{proj}_1 \operatorname{spt}\gamma$ and is also continuous, the identity is satisfied. As a consequence, we formulate the following conjecture we hope to be true—we are currently trying to refine it so that we can prove it later.

Conjecture 4.21. *For ν, $\operatorname{spt}\nu$ smooth enough and μ defined as usual, and $A = D_{\bar\psi} \cap (\operatorname{proj}_1 \operatorname{spt}\gamma)$ we have*

$$\mu(\partial A) = 0$$

and

$$\nu\left(\mathcal{C}^\tau(0) \setminus \tilde{\mathcal{C}}^\tau(0)\right) = 0.$$

Note that by Lemma 4.5, the projection $\operatorname{proj}_2(\theta(\partial\bar\psi) \llcorner D_{\bar\psi}) = \mathcal{C}^\tau(0)$ also gives $\mathbb{C}^\tau(0)$, hence also $\mathbb{C}^\tau(q)$ for every $q \in (0,1)$ by homogeneity. Assuming the result in Conjecture 4.21, it only remains to show that the weak convergence is conserved by this projection. To begin with, we prove an approximation result in the deterministic setting.

Lemma 4.22. *We consider two families of open sets B_k and $R_{\epsilon,\delta}$ and a sequence of closed sets $F_n \in \mathcal{F}\left((\mathbb{R}^d \setminus \{0\}) \times \mathbb{R}^d\right), n \geq 1$ satisfying the following assumptions,*

(a) for every $k \geq 1$, $\operatorname{cl} B_k \subset B_{k+1}$ and $\bigcup_{k\geq 1} B_k = B$.

(b) For $\epsilon, \delta > 0$, $\operatorname{cl} R_{\epsilon,\delta}$ is compact and $\operatorname{cl} R_{\epsilon,\delta} \subset R_{\epsilon+\tilde\epsilon,\delta-\tilde\delta}$ for every $\tilde\epsilon, \tilde\delta > 0$ with $\tilde\delta < \delta$.

(c) for every $k \geq 1$ and $\epsilon, \delta > 0$ we have the convergence

$$F_n \llcorner (R_{\epsilon,\delta} \cap B_k) \xrightarrow{\mathcal{F}} F \llcorner (R_{\epsilon,\delta} \cap B_k)$$

for some F in $\mathcal{F}\left((\mathbb{R}^d \setminus \{0\}) \times \mathbb{R}^d\right)$.

(d) *Let $k \geq 1$. for every $\epsilon, \tilde{\epsilon}, \delta, \tilde{\delta} > 0$ we have the equality*

$$F \llcorner (R_{\epsilon,\delta} \cap B_k) = F \llcorner (R_{\tilde{\epsilon},\tilde{\delta}} \cap B_k).$$

Under the assumptions above there exists an increasing sequence of sets Q_k satisfying $\cup_{k\geq 1} Q_k = Q$ where $Q = \operatorname{rge} F \llcorner (B \cap R_{\epsilon,\delta})$ which satisfies

$$\operatorname{cl} Q_k \subset \liminf_{n\to\infty} \operatorname{cl} \operatorname{rge} F_n \llcorner (B_k \cap R_{\epsilon,\delta}) \subset \limsup_{n\to\infty} \operatorname{cl} \operatorname{rge} F_n \llcorner (B_k \cap R_{\epsilon,\delta}) \subset Q_{k+1}.$$

Moreover we also have the convergence

$$\lim_{k\to\infty} \liminf_{n\to\infty} \operatorname{cl} \operatorname{rge} F_n \llcorner (B_k \cap R_{\epsilon,\delta}) = \operatorname{cl} Q = \lim_{k\to\infty} \limsup_{n\to\infty} \operatorname{cl} \operatorname{rge} F_n \llcorner (B_k \cap R_{\epsilon,\delta}),$$

and the equality

$$Q = \bigcup_{k\geq 1} \liminf_{n\to\infty} \operatorname{cl} \operatorname{rge} F_n \llcorner (B_k \cap R_{\epsilon,\delta}) = \bigcup_{k\geq 1} \limsup_{n\to\infty} \operatorname{cl} \operatorname{rge} F_n \llcorner (B_k \cap R_{\epsilon,\delta}).$$

We apply now this lemma to obtain some kind of convergence toward $\tilde{\mathcal{C}}^\tau(0)$. We consider the open set $B_k = \left\{ x \in \mathbb{R}^d : d(x, B^c) > 2^{-k} \right\}$. Since $\operatorname{cl} B_k = \left\{ x \in \mathbb{R}^d : |x - A^c| \geq 2^{-k} \right\}$, it is clear that $\operatorname{cl} B_k \subset B_{k+1}$. Moreover we have the inclusion $B_k \subset B$ and, since B is open, for every $x \in B$ there exist $k \geq 1$ such that $x \in B_k$. Therefore we have $\bigcup_{k\geq 1} B_k = B$, and the sequence of sets $B_k, k \geq 1$ meets assumption (a) in Lemma 4.22. We also define $R_{\epsilon,\delta} = \mathrm{B}_{0,\epsilon} \setminus \operatorname{cl} \mathrm{B}_{0,\delta}$ which is clearly an open set. Since $R_{\epsilon,\delta}$ is bounded in $\mathbb{R}^d$, its closure is compact. Finally we have $\operatorname{cl} R_{\epsilon,\delta} = \operatorname{cl} \mathrm{B}_{0,\epsilon} \setminus \mathrm{B}_{0,\delta}$, hence $\operatorname{cl} R_{\epsilon,\delta} \subset R_{\epsilon+\tilde{\epsilon},\delta-\tilde{\delta}}$ for every $\tilde{\epsilon}, \tilde{\delta} > 0$ with $\tilde{\delta} < \delta$ as required by assumption (b) in Lemma 4.22.

Theorem 4.23. *Let $F_n = \theta(\operatorname{spt} \hat{\gamma}_{n,t_n}) \llcorner B$, $F = \theta(\partial\bar{\psi}) \llcorner B$ where B is an open subset of $\mathbb{R}^d$ such that $\tilde{\mu}(\mathbb{R}^d \setminus B) = 0$ and which satisfies*

$$F_n \xrightarrow{w} F \quad \text{in } \mathcal{F}(B \times \mathbb{R}^d).$$

Then for any $k \geq 1$ the sequence $F_n \llcorner (B_k \cap R_{\epsilon,\delta}), n \geq 1$ is tight and every family of weak limit points $T_k, k \geq 1$ satisfies

$$\tilde{\mathcal{C}}^\tau(0) = \operatorname{proj}_2(\theta(\partial\bar{\psi}) \llcorner B) = \bigcup_{k\geq 1} \operatorname{cl} \operatorname{rge} T_k$$

Assuming that Conjecture 4.21 is satisfied, estimating $\tilde{\mathcal{C}}^\tau(0)$ is a good way to approach $\mathcal{C}^\tau(0)$. Thus we can use the above result to build an estimator of $\mathcal{C}^\tau(0)$. Since the sequence T_k in the statement of Theorem 4.23 is increasing, we suggest to take $k = \infty$ in the definition of the set $\operatorname{cl} \operatorname{rge} F_n \llcorner (R_{\epsilon,\delta} \cap B_k)$—recall that T_k is a limit point of this sequence as $n \to \infty$—which yields the set $\operatorname{cl} \operatorname{rge} F_n \llcorner (R_{\epsilon,\delta} \cap B)$.

Remark 4.24. *In practice we have no information either on the set B or on the sequence of sets $B_k, k \geq 1$ approximating B. However, under the assumption that $\mu(\partial A) = 0$ made in Conjecture 4.21, and assuming that the convergence of $\hat{\mu}_{n,t_n}$ toward μ is achieved, we expect that $|X_i|U_i/\hat{b}_n(t_n)$ belongs to B almost surely. As a consequence, taking $k = \infty$, we drop the restriction to the set B in the expression to build our estimator.*

As a consequence of the discussion above, we suggest to estimate $\mathcal{C}^\tau(0)$ by

$$\operatorname{cl} \circ \operatorname{rge} \left(\theta \left(\operatorname{spt} \hat{\gamma}_{n,t_n}\right) \llcorner R_{\epsilon_n,\delta_n}\right)$$

where $\epsilon_n, \delta_n > 0$ are hyperparameters one should choose empirically, bearing in mind Remark 4.16. This set can be rewritten as

$$\hat{\mathcal{C}}_n^\tau(0) = \left\{ \frac{|X_{\hat{\sigma}_n(i)}|}{|X_i|} \ : \ i \in \{1, \ldots, n\}, \delta_n < |X_i| < \epsilon_n \right\}$$

which is the estimator that we announced at the beginning of the current subsection.

Finally, to get an estimate of $\mathbb{C}^\tau(q)$ the center-outward tail quantile of order $q \in (0,1)$, we use Remark 4.4 and Lemma 4.5 to write

$$\hat{\mathbb{C}}^t(q) = (1-q)^{-1/\alpha} \bigcup_{y \in \hat{\mathcal{C}}_\nu^\tau(0)} [0, y].$$

Remark 4.25. *In the expression of $\hat{\mathbb{C}}^t(q)$ above, we use the indice of regular variation α which is not known in practice. A reasonable and usual way to address this problem is to plug in an estimated value of α—e.g. we could use the famous Hill estimator for which we refer to the monograph Resnick (2007).*

Note that the estimator $\hat{\mathbb{C}}^t(q)$ enjoys the same kind of homogeneity as $\mathbb{C}^\tau(q)$, in the sense of Remark 4.4. This property will be useful in Subsubection 4.4.2 to extrapolate center-outward quantile regions of order q near 1.

4.4 Application to the estimation of extreme quantile regions and numerical results

4.4.1 Estimation of a center-outward quantile regions of high order

In this subsubsection, we use the theory described in the previous sections to build an estimate of the center-outward quantile regions of order $q \in (0,1)$ close to one.

We consider a target measure $Q \in \mathcal{P}(\mathbb{R}^d)$ regularly varying with indice α, auxiliary function b and limit measure $\nu \in \mathrm{M}_0(\mathbb{R}^d)$ and assume that both Q and ν vanish on sets of Hausdorff dimension at most $d-1$. Let $P \in \mathcal{P}$ be defined by the polar decomposition $|\cdot|_{\#}Q \otimes \mathrm{Unif}(\mathbb{S}^{d-1})$ as suggested in Subsection 4.1. A simple computation yields that P is regularly varying with the same indice and auxiliary function as Q and limit measure $\mu \in \mathrm{M}_0(\mathbb{R}^d)$ given by the polar decomposition $\nu_\alpha \otimes \mathrm{Unif}(\mathbb{S}^{d-1})$. Moreover, it is clear that P and μ vanish on sets of Hausdorff dimension at most $d-1$, and $\operatorname{spt} \mu = \mathbb{R}^d$.

As a consequence, we are in the same framework as described in the introduction to Section 4, with the further assumption that ν vanishes on sets of Hausdorff dimension at most $d-1$. This means that (a) $P, Q \in \mathcal{P}(\mathbb{R}^d)$ are regularly varying with common index $\alpha > 0$, common auxiliary function b and limit measures $\mu, \nu \in \mathrm{M}_0(\mathbb{R}^d) \setminus \{0\}$; (b) P, Q, μ and ν all vanish on sets of Hausdorff dimension at most $d-1$, and $\operatorname{spt} \mu = \mathbb{R}^d$; (c) the restriction $\operatorname{res} \pi$ of the unique cyclically monotone $\pi \in \Pi_{cm}(P, Q)$ between P and Q is regularly varying with index α, auxiliary function b and limit measure γ where $\gamma \in \Gamma_{0,cm}(\mu, \nu)$ is the unique cyclically monotone zero-coupling between the limit measures μ and ν; (d) there exist closed convex functions ψ and $\bar{\psi}$ whose gradients are uniquely determined P and μ-almost everywhere respectively such that $\pi = [\mathrm{Id} \times \nabla\psi]_{\#}P$ and $\gamma = [\mathrm{Id} \times \nabla\bar{\psi}]_{\#}\mu$; (e) the rescaling $(b(t))^{-1}\partial\psi(b(t)\,\cdot)$ $\mathcal{F}$-converges to $\partial\bar{\psi}$ in $\mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d)$ as $t \to \infty$.

Remark 4.26. *One should notice that the assumptions of most of the results we proved in the previous sections are fulfilled, except for example Theorem 4.10 which requires an assumption on the support of ν.*

Let Y be a random vector in $\mathbb{R}^d$ with distribution Q and let $p_t = \mathsf{P}\left(Y \notin \mathbb{C}(1-1/t)\right)$. From the introduction to Section 4, for every Borel set A of $\mathbb{R}^d$ and $t > 0$ we have the decomposition

$$\mathsf{P}\left(Y \in A\right) = (1-p_t)\,\mathsf{P}\left(Y \in A \mid Y \in \mathbb{C}(1-1/t)\right) + p_t\,\mathsf{P}\left(Y \in A \mid Y \in \mathbb{C}(1-1/t)^c\right).$$

In general, there is no reason for $\mathbb{C}(s)$ to be a subset of $\mathbb{C}^\tau(s)$. One may easily think of cases where Q puts mass on the whole space $\mathbb{R}^d$ while extremes only occur in a subset of $\mathbb{R}^d$, e.g. an orthant. As a consequence, to benefit from both center-outward quantiles and tail quantiles regions, we propose to proceed as follows, taking inspiration from the decomposition introduced before. For some $\beta \in (0,1)$, let $A_t = \mathbb{C}(1-1/t) \cup b(t)\mathbb{C}^\tau(\beta)$. By construction, we have

$$(1-p_t)\,\mathsf{P}\left(Y \notin A_t \mid Y \in \mathbb{C}(1-1/t)\right) = 0,$$

hence the decomposition introduced before applied to the set A_t^c simplifies into

$$\mathsf{P}\left(Y \notin A_t\right) = p_t\,\mathsf{P}\left(Y \notin A_t \mid Y \in \mathbb{C}(1-1/t)^c\right)$$

Remark 4.27. *Note that in the definition of the set A_t, the component $b(t)\mathbb{C}^\tau(\beta)$ is homogenous in β in the sense given in Remark 4.4 while the component $\mathbb{C}(1-1/t)$ is not. The homogenous part allows us to extrapolate in the directions where there are extreme values while the other part is the usual center-outward quantile and is expected to induce small errors in directions where there is no extreme values.*

Moreover, since Q vanishes on sets of Hausdorff dimension at most $d-1$, we have

$$t\ \mathsf{P}\left(Y \in \mathbb{C}(1-1/t)^c\right)) = t\,P\left(b(t)\left(\mathbb{R}^d \setminus \mathrm{B}_{0,1}\right)\right)$$

whence $t\ \mathsf{P}\left(Y \in \mathbb{C}(1-1/t)^c\right))$ goes to $\mu(\mathbb{R}^d \setminus \mathrm{B}_{0,1}) = 1$ as $n \to \infty$. As a consequence we are interested in the convergence of

$$t\,\mathsf{P}\left(Y \in A_t^c \cap \mathbb{C}(1-1/t)^c\right) = t\,\mathsf{P}\left(Y \in b(t)\left(\mathbb{C}^\tau(\beta)^c \cap b(t)^{-1}\mathbb{C}(1-1/t)^c\right)\right)$$

As in the preceedings sections, let $D_{\bar\psi}$ be the set of points $x \in \mathbb{R}^d$ such that $\partial\bar\psi(x)$ is a singleton, i.e., where $\nabla\bar\psi$ os defined. We consider the sets $A = D_{\bar\psi} \cap (\mathrm{proj}_1 \operatorname{spt}\gamma)$ and $B = \operatorname{int} A$. Thus Remark 2.19 yields $\mu(D_{\bar\psi} \setminus A) = 0$. However, the identity $\mu(\partial A) = 0$ may not hold. As a consequence we can expect in practice to have some empirical points lying in ∂A. Moreover, we will have $\mathrm{proj}_2\left(\theta(\partial\bar\psi)\llcorner B\ \right) \subset \mathcal{C}^\tau(0)$ but not $\nu\left(\mathcal{C}^\tau(0) \setminus \mathrm{proj}_2\left(\theta(\partial\bar\psi)\llcorner B\ \right)\right) = 0$ in general.

As $(\mathrm{B}_{0,1} \cap B)$ is an open subset of $V = \operatorname{int}\operatorname{spt}\mu = \mathbb{R}^d$, the already mentioned Fell convergence from Theorem 2.38 gives

$$(b(t))^{-1}\partial\psi(b(t)\,\cdot\,)\llcorner(\mathrm{B}_{0,1} \cap B) \xrightarrow{\mathcal{F}} \partial\bar\psi\llcorner(\mathrm{B}_{0,1} \cap B).$$

Note that $\mathbb{C}^\tau(0) = \mathrm{rge}\left[\partial\bar\psi\llcorner(\mathrm{B}_{0,1} \cap D)\right]$ and

$$\begin{aligned}(b(t))^{-1}\,\mathbb{C}(1-1/t) &= (b(t))^{-1}\,\nabla\bar\psi(b(t)\mathrm{B}_{0,1}) \\ &= \mathrm{rge}\left[(b(t))^{-1}\partial\psi(b(t)\,\cdot\,)\llcorner(D \cap \mathrm{B}_{0,1})\right].\end{aligned}$$

This gives some motivation to formulate the following conjecture.

Conjecture 4.28. *For $\beta \in (0,1)$ we have the convergence*

$$t\,\mathsf{P}\left(Y \in b(t)\left(\mathbb{C}^{\tau}(\beta)^{c} \cap b(t)^{-1}\mathbb{C}(1-1/t)^{c}\right)\right) \longrightarrow \nu\left(\mathbb{C}^{\tau}(\beta)^{c}\right) = (1-p)(1-\beta)$$

as $t \to \infty$, where $p = \mu\left([\nabla\psi]^{-1}(\{0\}) \cap \mathrm{B}_{0,1}^{c}\right)$ as in Theorem 4.11. Let $\chi = \{x \in \mathbb{R}^{d} : \operatorname{spt}\nu \cap \operatorname{cl} H_{+}(x) = \{0\}\}$ like in Theorem 4.10. If we further assume that $\mu(\chi) = 0$, then $p = 0$ in the expression above.

In practice we assume that the convergence in Conjecture 4.28 is achieved for some t^{*} large enough. This yields the approximation

$$\mathsf{P}\left(Y \in A_{t}^{c}\right) \approx p_{t^{*}}\,(1-p)(1-\beta)$$

where $p = \mu\left([\nabla\psi]^{-1}(\{0\}) \cap \mathrm{B}_{0,1}^{c}\right)$ is the share of directions on which Q has heavy tails. To build a region of probability $q \in (0,1)$ near 1, the idea is to choose β such that $p_{t^{*}}\,(1-p)(1-\beta) = 1-q$, i.e.,

$$\beta = 1 - \frac{1-q}{p_{t^{*}}(1-p)}.$$

In the end of this subsubsection, we propose to use this approximation to deal with real data. Let $n \in \mathbb{N} \setminus \{0\}$ denote the size of the dataset, and let $\{Y_i\}_{i=1,\ldots,n}$, $\{U_i\}_{i=1,\ldots,n}$ be two iid sample with distribution Q, $\operatorname{Unif}(\mathbb{S}^{d-1})$ respectively. We define $\{X_i\}_{i=1,\ldots,n}$ by $X_i = U_i|Y_i|$. We call $\{Y_i\}_{i=1,\ldots,n}$ the target sample, and $\{X_i\}_{i=1,\ldots,n}$ the reference sample.

Since the definition of A_t involves unknown sets, it is natural to substitute them with empirical counterpart built from the data. Thus replace the set $A_{t^{*}}$ by the approximation $\hat{A}_{t^{*}}$ defined by

$$\hat{A}_{t^{*}} = \hat{\mathbb{C}}(1-1/t^{*}) \;\cup\; \hat{\mathbb{C}}^{t}(\beta)$$

which can be expressed, thanks to homogeneity, as

$$\hat{A}_{t^{*}} = \hat{\mathbb{C}}(1-1/t^{*}) \;\cup\; \left[(1-\beta)^{-1/\alpha}\hat{\mathbb{C}}^{t}(0)\right]$$

where $\hat{\mathbb{C}}^{t}(0)$ is defined as at the end of Subsection 4.3, using the estimator of the center-outward tail quantile contour of order 0

$$\hat{\mathcal{C}}_{n}^{\tau}(0) = \left\{ \frac{|X_{\hat{\sigma}_n(i)}|}{|X_i|} \;:\; i \in \{1,\ldots,n\}, \delta_n < |X_i| < \epsilon_n \right\}.$$

where $\epsilon_n, \delta_n > 0$ are hyperparameters one should choose empirically, bearing in mind Remark 4.16, and

$$\hat{\mathbb{C}}(1 - 1/t^*) = \left\{ y \ : \ \exists x \text{ such that } (x, y) \in \operatorname{spt} \hat{\gamma}_{n,t^*} \text{ and } |x| \leq \hat{b}_n(t^*) \right\}$$

for $\hat{b}_n(t) = \hat{F}_n^{\leftarrow}(1 - 1/t)$—recall that $\hat{F}_n$ is the usual empirical distribution function of the reference sample $\{|X_i|\}_{i=1}^n$.

Remark 4.29. *In the expression of $\hat{\mathbb{C}}^t(\beta)$ above, we use the indice of regular variation α which is not known in practice. A reasonable and usual way to address this problem is to plug in an estimated value of α—e.g. we could use the famous Hill estimator for which we refer to the monograph Resnick (2007). For our experiments in Subsubsection 4.4.2, we will assume that the indice is known.*

In practice, we choose δ_n so that the $100(1 - p_{t^*})$ percent of the couples in $\operatorname{spt} \hat{\gamma}_{n,t_n}$ such that the element given by the first d coordinates have the smallest norm are ignored, and ϵ_n such that the $100(1 - p_{t^*}/2)$ percent—the link between ϵ_n and δ_n has been chosen arbitrarily—of the couples in $\operatorname{spt} \hat{\gamma}_{n,t_n}$ such that the element given by the first d coordinates have the largest norm are ignored too.

Note that p_{t^*}, the share of the data considered to be extreme, is choose empirically as a small number in $(0, 1)$ near 0 for which visually the shape of the empirical contour $\hat{\mathcal{C}}_\nu^\tau(0)$ does not change much when p_{t^*} changes by a small amount, and we also require p_{t^*} to be large enough to have a meaningful contour $\hat{\mathcal{C}}_\nu^\tau(0)$—we need enough extremal points to build a meaningful estimator. This classical bias-variance tradeoff in statistics with heavy-tailed distributions will be discuss in Subsubsection 4.4.2.

4.4.2 Numerics

In this last subsubsection, we carry out numerical experiments on the approximation proposed in Subsubection 4.4.1. All experiments below will be based on two-dimensional distributions in order to make it easier to represent. The target sample will always be composed of $n = 3000$ points. Let N denote the set $\{1, \ldots, n\}$.

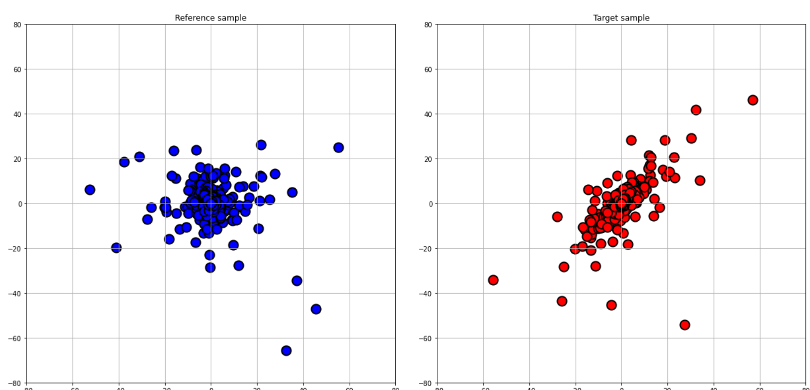
Let $g_\alpha : t \in \mathbb{R}_+ \longmapsto \mathbb{1}_{[1,\infty)}(t)\alpha t^{-(\alpha+1)}$ be the density of a Pareto distribution with parameter $\alpha > 0$. We recall from Subsection 1.6 the definition of the family $\mathcal{P}_{ell}^2$ of elliptical distributions on $\mathbb{R}^2$. Let $\mathrm{P}_{0,I,g}$ denote the distribution which is uniform on the unit sphere and has radial density g (a density over $\mathbb{R}_+$). We say that X has distribution $\mathrm{P}_{\mu,\Sigma,g}$, where μ is a location parameter and Σ is a symmetric positive definite real matrix, iff $Y = \Sigma^{-1/2}(X - \mu)$ has spherical distribution $\mathrm{P}_{0,I,g}$. We can finally write $\mathcal{P}_{ell}^2 = \{ \mathrm{P}_{\mu,\Sigma,g} \}$ for the family of all full-rank elliptical distributions over $\mathbb{R}^2$ with radial densities over elliptical support sets.

Let $\mathcal{N}(\mu, \Sigma)$ denote the normal distribution with mean $\mu \in \mathbb{R}^d$ and covariance matrix Σ. Since Σ is positive semi-definite, we can write the Cholesky decomposition of Σ as $\Sigma = \Sigma^{1/2}(\Sigma^{1/2})^T$. Recall the identity in law $\mathcal{N}(\mu, \Sigma) = \Sigma^{1/2}\mathcal{N}(0_{\mathbb{R}^2}, I_2) + \mu$, where I_2 is the 2×2 matrix with 1 on the diagonal and 0 everywhere else. It is easy to find a function g such that the normal distribution $\mathcal{N}(0_{\mathbb{R}^2}, I_2)$ can be written $\mathrm{P}_{0,I_2,g}$. Thus for such g, we also have the identity in law $\mathcal{N}(\mu, \Sigma) = \mathrm{P}_{\mu,\Sigma,g}$. As a consequence, elliptic distribution with Pareto density for radial density is a natural way to construct heavy-tailed counterpart to general Gaussian distribution in $\mathbb{R}^2$. Following Hallin et al. (2021), we will conduct experiments with mixtures of elliptic distribution with Pareto density for radial density instead of multivariate gaussian distributions in the later reference.

To begin with, we consider the simplest case: an elliptic distribution for which the center-outward quantile region of every order $q \in (0, 1)$ is known. Note that we have already encountered elliptic distribution in Subsection 1.6. Let $\alpha = 1.5$—the choice of α will be the same for all examples—and

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} 0.75 & 0 \\ 0 & 0.75 \end{bmatrix}.$$

Let $Y_i, i \in N$ be a iid sample from the distribution $\mathrm{P}_{\mu,\Sigma,g_\alpha}$. We write $X_i, i \in N$ for the reference sample defined as at the end of Subsubsection 4.4.1, i.e., $X_i = U_i|Y_i|$ where $U_i, i \in N$ is a iid sample from the uniform distribution on the sphere which is also independent of the target sample $Y_i, i \in N$. The two samples are illustrated in Figure 1.



(a) Reference sample (b) Target sample

Figure 1: Empirical samples

Following the discussion in Subsection 1.6, it is clear that the center-outward quantile region of order $q \in (0,1)$ can be simply expressed as

$$\mathbb{C}_{\mathrm{P}_{0,I_2,g_\alpha}:\mathrm{P}_{0,\Sigma,g_\alpha}}(q) = \nabla\psi(G_\alpha^{\leftarrow}(q)\mathrm{B}_{0,1}) = \Sigma^{1/2}(G_\alpha^{\leftarrow}(q)\mathrm{B}_{0,1}) + \mu,$$

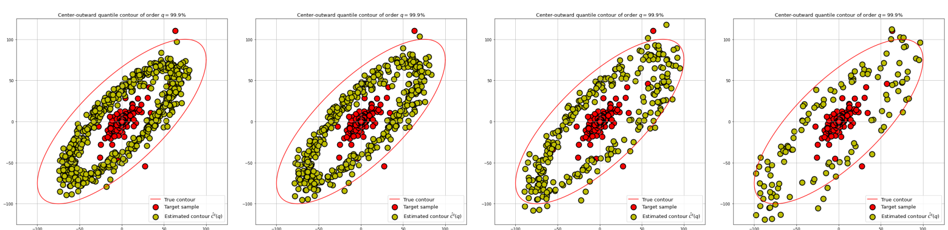
i.e.,

$$\mathbb{C}_{\mathrm{P}_{0,I_2,g_\alpha}:\mathrm{P}_{0,\Sigma,g_\alpha}}(q) = \Sigma^{1/2}\mathrm{B}_{0,(1-q)^{-\alpha}},$$

where G_α is the cumulative distribution function of the Pareto distribution with parameter $\alpha > 0$ and $\nabla\psi$ is the P-almost everywhere unique gradient of a closed convex function pushing P forward to Q. The same arguments yields a simple expression of the center-outward quantile contour of order $q \in (0,1)$ as

$$\mathcal{C}_{\mathrm{P}_{0,I_2,g_\alpha}:\mathrm{P}_{0,\Sigma,g_\alpha}}(q) = \Sigma^{1/2}\left\{x \in \mathbb{R}^2 \quad : \ |x| = (1-q)^{-\alpha}\right\}.$$

In Figure 2 we represent, for $q = 99.9\%$ and different values of p_{t^*}, the estimate $\hat{\mathcal{C}}(q)$ of $\mathcal{C}_{\mathrm{P}_{0,I_2,g_\alpha}:\mathrm{P}_{0,\Sigma,g_\alpha}}(q)$ defined in Subsection 4.3, together with the true contour $\mathcal{C}_{\mathrm{P}_{0,I_2,g_\alpha}:\mathrm{P}_{0,\Sigma,g_\alpha}}(q)$ and the target sample. When it is clear from the context we will simply write $\mathcal{C}(q)$ and $\mathbb{C}(q)$ for the center-outward quantile contours and regions of order $q \in (0,1)$ of the target distribution with respect to the reference distribution.



(a) $p_{t^*} = 0.20$ (b) $p_{t^*} = 0.15$ (c) $p_{t^*} = 0.10$ (d) $p_{t^*} = 0.05$

Figure 2: Empirical estimate of the center-outward quantile contour of order $q = 99.9\%$

We remark that as the share p_{t^*} of the data deemed extreme decreased, the estimate $\hat{\mathcal{C}}(q)$ of the set $\mathcal{C}_{\mathrm{P}_{0,I_2,g_\alpha}:\mathrm{P}_{0,\Sigma,g_\alpha}}(q)$ becomes "closer" to the true contour. However, as p_{t^*} increases, we have less points to build the estimate and the points of the set $\hat{\mathcal{C}}(q)$ seem to be more widely scattered; as a consequence, the estimate becomes less easy to use. This is the usual bias/variance trade-off in statistics with extreme values. From now on, we will work with $p_{t^*} = 0.10$ which seems to be a good choice according to the above results in Figure 2. From now on, we will choose $p_{t^*} = 0.10$ in all our examples. As suggested by Segers (2023) in a talk that he gave at Joint Statistical

Meeting in Toronto, Ontario—see Remark 4.16—we choose $p_{t^*} = 0.10$ but don't use the $100 * p_{t^*}/2 = 5$ percent of the couples in spt $\hat{\gamma}_{n,t_n}$ such that the element given by the first d coordinates have the largest norm. Figure 3 shows that doing so we reduce in some sense the variance of our estimator, hence we will continue to use this method in the subsequent examples.

Remark 4.30. *Applying the method described above, our dataset of $n = 3000$ points only gives $3000 * 0.05 = 150$ couples that are actually used to build the estimator, which is really few. This illustrates a major difficulty in statistics with heavy-tailed distributions: you need to have access to really large quantities of data to see enough extreme values that you can use.*

Recall from Section 4.1 that the center-outward tail quantile regions and contours are homogenous. In our example, the support of the spectral measure in the polar decomposition of the limit measure in the definition of regular variation is the whole sphere, i.e., there is extremes values in all directions. As a consequence, it is reasonable to approximate the center-outward quantile contour of high order by a homogeneous set. In fact, since we use g_α for radial measure, the center-outward quantile contour is homogenous here. In Figure 4, we represent, for p_{t^*} and different values of q, the estimate $\hat{\mathcal{C}}(q)$ of $\mathcal{C}_{\mathrm{P}_{0,I_2,g_\alpha}:\mathrm{P}_{0,\Sigma,g_\alpha}}(q)$ defined in Subsection 4.3, together with the true contour $\mathcal{C}_{\mathrm{P}_{0,I_2,g_\alpha}:\mathrm{P}_{0,\Sigma,g_\alpha}}(q)$ and the target sample.

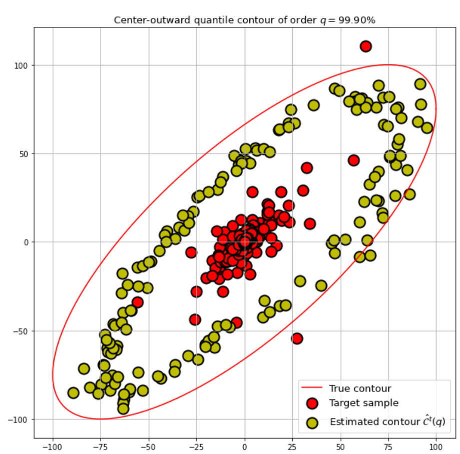
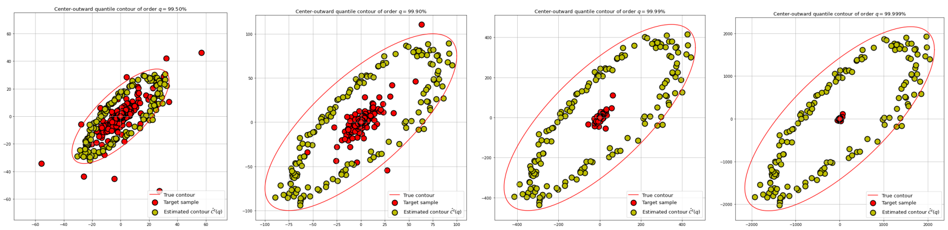


Figure 3: Empirical estimate of the center-outward quantile contour of order $q = 99.9\%$



(a) $q = 99.50\%$ (b) $q = 99.90\%$ (c) $q = 99.99\%$ (d) $q = 99.999\%$

Figure 4: Empirical estimate of the center-outward quantile contour of order q

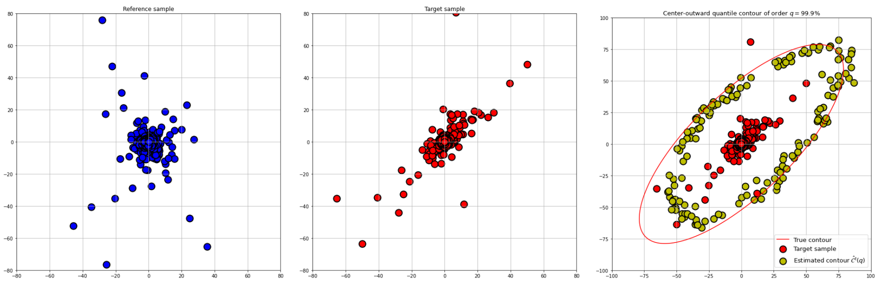
Note that the axes are different on each plot. Visually, the difference between the estimator and the true contour is the same. However, the area inside the true contour which is not "inside" the estimator of the contour—since we did not produce an interpolation of the sets of points that make up the estimator set $\hat{\mathcal{C}}(q)$, the last statement is not clear but we hope it is visually clear—has a mass which decreases at a power rate as q increases. As a consequence, the error made when using the estimated contour $\hat{\mathcal{C}}(q)$ goes to zero as q goes to one.

We now consider a slightly different case by simply adding a gaussian white noise to the distribution of the above example. We replace the distribution $\mathrm{P}_{0,\Sigma,g_\alpha}$ by the mixture $0.30\mathcal{N}(0_{\mathbb{R}^2}, I_2) + 0.70\mathrm{P}_{0,\Sigma,g_\alpha}$. In this case, we have a simple approximation of the true contour of order $q \in (0,1)$ close to 0. Indeed, for such q, the probability that a random variable with distribution $\mathcal{N}(0_{\mathbb{R}^2}, I_2)$ belongs to the center-outward quantile region of order q is close to one. Moreover, the quantile region of order q for $\mathcal{N}(0_{\mathbb{R}^2}, I_2)$ is included quantile region of order q for $\mathrm{P}_{0,\Sigma,g_\alpha}$. As a consequence, we have the approximation

$$\mathcal{C}(q) \approx \Sigma^{-1/2}\partial \mathrm{B}_{0,(1-\frac{q-0.30}{1-0.30})^{-\alpha}}.$$

In Figure 5 we plot the results for this mixture distribution in the same way as before.

Surprisingly, it seems that our method seems perform better to estimate the center-outward quantile contour of order $q = 99.9\%$ with this mixture distribution than it does with the elliptic distribution used in the first example, except for the "low left corner" of the contour.



(a) Reference sample (b) Target sample (c) Empirical estimate of the center-outward quantile contour of order $q = 99.9\%$

Figure 5: Empirical samples

From here to the end of this section we will reproduce outputs of Figure 3 in Hallin et al. (2021)—which is reproduced as Figure 6 below—with mixtures of mixtures of elliptic distribution with Pareto density for radial density instead of multivariate gaussian distributions in the later reference. To do so we introduce the positive-definite covariance matrices

$$\Sigma_1 = \begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix}, \qquad \Sigma_2 = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}, \qquad \Sigma_3 = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

and the mean matrices

$$\mu_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \mu_h = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \mu_v = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

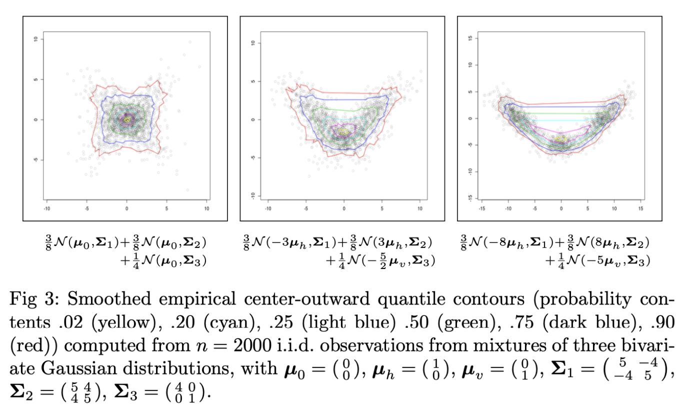
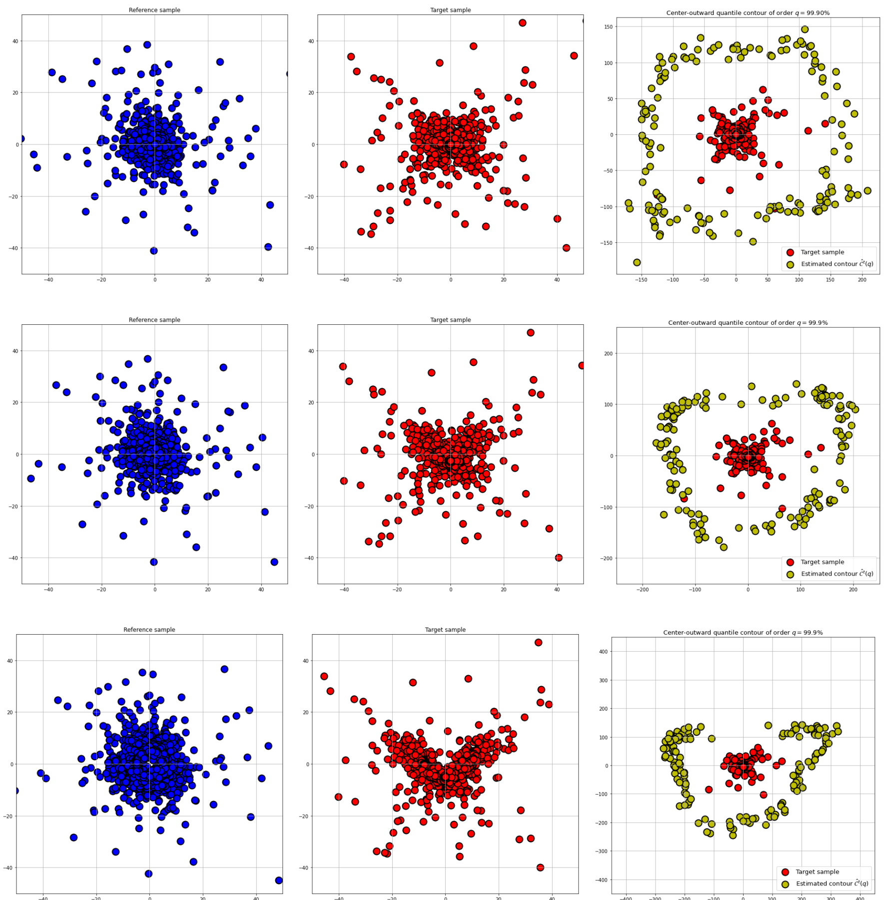


Figure 6: Figure 3 from Hallin et al. (2021)

We replace the distributions used in Hallin et al. (2021) by

$$\frac{3}{8}\mathrm{P}_{\mu_0,\Sigma_1,g_\alpha}+\frac{3}{8}\mathrm{P}_{\mu_0,\Sigma_2,g_\alpha}+\frac{1}{4}\mathrm{P}_{\mu_0,\Sigma_3,g_\alpha}$$
$$\frac{3}{8}\mathrm{P}_{-3\mu_h,\Sigma_1,g_\alpha}+\frac{3}{8}\mathrm{P}_{3\mu_h,\Sigma_2,g_\alpha}+\frac{1}{4}\mathrm{P}_{-\frac{5}{2}\mu_v,\Sigma_3,g_\alpha}$$
$$\frac{3}{8}\mathrm{P}_{-8\mu_h,\Sigma_1,g_\alpha}+\frac{3}{8}\mathrm{P}_{8\mu_h,\Sigma_2,g_\alpha}+\frac{1}{4}\mathrm{P}_{-5\mu_v,\Sigma_3,g_\alpha}$$

respectively (from left to right). In Figure 7 we represent on each row the reference sample, the target sample and the estimated center-outward quantile contour of order $q = 99.9\%$. The distributions are in the same order (top to bottom) as in Figure 6.



(g) Reference sample

(h) Target sample

(i) Empirical estimate of the center-outward quantile contour of order $q = 99.9\%$

Figure 7: Empirical samples

It turns out that our method manage to catch the general "form" of the target distributions but it struggles with non-convexity more than the usual center-outward quantile methods in Figure 6 taken from Hallin et al. (2021). This is especially visible for the first distribution. Moreover, for the last distribution our method fails to recover the extremal directions visually on the top of the distribution.

It is important to note that due to lack of time we used samples of relatively small sizes here ($n = 3000$), so further experiments are required to confirm our results. However, one may say that our results look promising.

5 Conclusion

What has been achieved

To deal with a multivariate regularly varying target distribution Q, following Segers (2023) we proposed to use a regularly varying distribution as reference distribution—instead of the spherical uniform distribution—in the construction of the center-outward quantile regions $\mathbb{C}(q)$ and contours $\mathcal{C}(q)$, $q \in (0,1)$, introduced by Hallin et al. (2021) using optimal transport with quadratic distance as a cost. In order to have simple quantile regions for the reference distribution—like for the uniform spherical distribution—we also chose the reference distribution P to admit a polar decomposition involving the uniform distribution on the unit sphere, $|\cdot|_{\#}Q \otimes \mathrm{Unif}(\mathbb{S}^{d-1})$. This reference distribution P is regularly varying with the same indexe and auxiliary function as Q. Moreover, the limit measure ν in the definition of regular variation for P has polar decomposition $\nu_\alpha \otimes \mathrm{Unif}(\mathbb{S}^{d-1})$, whence it only depends on the indice of regular variation.

In our main result (Theorem 2.38) we proved that the unique cyclically monotone coupling π between the reference and target distributions is regularly varying with the same indice and auxiliary function, and has a limit measure γ whose support is cyclically monotone too. Moreover, we proved that the latter measure can be written $\gamma = [\mathrm{Id} \otimes \nabla\bar{\psi}]_{\#}\mu$ where $\bar{\psi}$ is a closed convex function satisfying $\operatorname{spt}\gamma \subset \partial\bar{\psi}$ and $\nabla\bar{\psi}$ is uniquely defined μ-almost everywhere. Noticing the similarity with π, which can be written $\pi = [\mathrm{Id} \otimes \nabla\psi]_{\#}P$ where ψ is a closed convex function satisfying $\operatorname{spt}\gamma \subset \partial\psi$ and $\nabla\psi$ is uniquely defined P-almost everywhere —we refer to McCann (1995), we introduced center-outward tail quantile regions $\mathbb{C}^\tau(q)$ and contours $\mathcal{C}^\tau(q)$, $q \in (0,1)$, extending the work of Hallin et al. (2021) to measures with possibly infinite total mass in the set $\mathrm{M}_0(\mathbb{R}^d)$ and proved some basic properties about them.

Building on this and on usual approximations in Extreme Values Theory, we proposed a simple approximation allowing one to construct quantile regions of order close to 1 using: $\mathbb{C}(1 - 1/t^*)$ for some large t^* and $\mathbb{C}^\tau(\beta)$ for some $\beta \in (0,1)$ to be determined. To conclude the theoretical part of the text, we introduced an estimator $\hat{\mathbb{C}}^\tau(0)$ of $\mathbb{C}^\tau(0)$ and gave some rationale for it. Using the homogeneity of $\mathbb{C}^\tau(q)$ and the link between $\mathbb{C}^\tau(q)$ and $\mathcal{C}^\tau(q)$, it provides an estimator of $\mathbb{C}^\tau(\beta)$ for every $\beta \in (0,1)$ too. Finally, we tested our method on simple examples built from mixtures of elliptical distributions with Pareto distributions for the radial distributions and obtained promising results.

What remains to do

In the construction of the approximation that allowed us to construct quantile regions of order close to 1, we let a convergence we believe to be true as Conjecture 4.28. A second result, giving a useful approximation of $\mathcal{C}^\tau(q)$, was left as Conjecture 4.21 in the construction of the estimator $\hat{\mathbb{C}}^\tau(0)$. Moreover, we didn't manage to provide a proof of the consistency of the latter estimator. The difficulty arises from the fact that throughout the text we have used an approach based on the convergence of closed sets in the Fell topology, whereas it turns out that $\mathbb{C}^\tau(q)$ is not closed in general. As a consequence another approach should be considered. Since the results of the experiments are promising, it is not impossible that our conjectures are true and that our estimator is indeed consistent in some sense under some assumptions; this gives some motivation to prove these results.

Possible applications and future works

To begin with, we didn't address the problem of interpolating center-outward tail quantile contours from real data points like it is done in Hallin et al. (2021) and Beirlant et al. (2020) for center-outward quantile contours; this is left to future work. In Beirlant et al. (2020) it has been proposed to use center-outward quantile to build a multivariate extension of the univariate theory of risk measurement. For this application, being able to estimate quantiles of order close to one for heavy-tailed distributions is of obvious interest. Moreover, we also see applications in anomaly detection when multivariate heavy tailed distributions are involved. Finally, we believe that the present work could also lead to applications in sampling from an unknown heavy-tailed distribution learned from a dataset using the decomposition introduced at the beginning of Section 4 to build a mixture distribution.

A Proofs of Section 2

Proofs of Subsection 2.1

Proof of Lemma 2.2. We treat the case $i = 1$ but the other one may be treated exactly in the same way. It suffices to check that for every sequence $\mu_n \in \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d), n \geq 1$ which M_0-converges to $\mu \in \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$, we have that $\varphi_1(\gamma_n)$ M_0-converges to $\varphi_1(\mu) \in \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$. To do so, we consider the Portmanteau Theorem 2.1(iii) in Lindskog et al. (2014) (or equivalently Lemma 4.1 in Kallenberg (2017)) and apply it twice. Let $A \subset \mathbb{R}^d \setminus \{0\}$ be bounded away from the origin, then $\mathrm{int}(A)$ ($\mathrm{int}(A) \times \mathbb{R}^d$) and $cl(A)$ ($cl(A) \times \mathbb{R}^d$) are respectively open and closed set bounded away from zero in $\mathbb{R}^d$ ($\mathbb{R}^d \times \mathbb{R}^d$). Since $\mu_n \in \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d), n \geq 1$ M_0-converges to $\mu \in \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$, we have as $n \to \infty$

$$\mu(\mathrm{int}(A) \times \mathbb{R}^d) \leq \liminf \mu_n(\mathrm{int}(A) \times \mathbb{R}^d) \leq \limsup \mu_n(\mathrm{cl}(A) \times \mathbb{R}^d) \leq \mu(\mathrm{cl}(A) \times \mathbb{R}^d)$$

what we can rewrite

$$\varphi_1(\mu)(\mathrm{int}(A)) \leq \liminf \varphi_1(\mu_n)(\mathrm{int}(A)) \leq \limsup \varphi_1(\mu_n)(\mathrm{cl}(A)) \leq \varphi_1(\mu)(\mathrm{cl}(A)).$$

Whence we have that $\varphi_1(\mu_n)$ M_0-converges to $\varphi_1(\mu) \in \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$ and φ_1 is continuous. $\square$

Proof of Lemma 2.5. Let $r > 0$. To avoid ambiguities, we write $\mathrm{B}_{0,s}(\mathbb{R}^d)$ and $\mathrm{B}_{0,s}(\mathbb{R}^d \times \mathbb{R}^d)$ for the balls centered at 0 and of radius $s > 0$ in the spaces $\mathbb{R}^d$ and $\mathbb{R}^d \times \mathbb{R}^d$ respectively. For some $r' > 0$ small enough we have $\mathrm{B}_{0,r'}(\mathbb{R}^d) \times \mathrm{B}_{0,r'}(\mathbb{R}^d) \subset \mathrm{B}_{0,r}(\mathbb{R}^d \times \mathbb{R}^d)$ and

$$(\mathbb{R}^d \times \mathbb{R}^d) \setminus \mathrm{B}_{0,r}(\mathbb{R}^d \times \mathbb{R}^d) \subset \left(\left(\mathbb{R}^d \setminus \mathrm{B}_{0,r'}(\mathbb{R}^d)\right) \times \mathbb{R}^d\right) \cup \left(\mathbb{R}^d \times \left(\mathbb{R}^d \setminus \mathrm{B}_{0,r'}(\mathbb{R}^d)\right)\right).$$

Therefore, for any $\gamma \in \Gamma_0(\mu, \nu)$ we have

$$\gamma((\mathbb{R}^d \times \mathbb{R}^d) \setminus \mathrm{B}_{0,r}) \leq \mu(\mathbb{R}^d \setminus \mathrm{B}_{0,r'}) + \nu(\mathbb{R}^d \setminus \mathrm{B}_{0,r'}) < \infty. \qquad \square$$

Proof of Lemma 2.7. Let $\pi \in \Pi(\alpha, \beta)$ for some $\alpha \in \overline{\mu}$ and $\beta \in \overline{\nu}$. For every $A \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$ we have

$$\pi(A \times \mathbb{R}^d) = \alpha(A) = \mu(A)$$

and

$$\pi(\mathbb{R}^d \times A) = \beta(A) = \nu(A)$$

so $\varphi_1\pi = \mu$ and $\varphi_2\pi = \nu$, hence $\gamma = \mathrm{res}\,\pi \in \Gamma_0(\mu, \nu)$.

Conversely, let $\gamma \in \Gamma_0(\mu, \nu)$. If we choose $\gamma' \in \bar{\gamma}$, $\alpha \in \overline{\mu}$ and $\beta \in \overline{\nu}$ such that $z := \gamma'(\{0\}) \geq 0$,

$$\alpha(\{0\}) = z + \gamma\left(\{0\} \times (\mathbb{R}^d \setminus \{0\})\right)$$

and

$$\beta(\{0\}) = z + \gamma\left((\mathbb{R}^d \setminus \{0\}) \times \{0\}\right)$$

then a simple computation gives $\gamma' \in \Pi(\alpha, \beta)$.

In particular, the choice $\gamma'(\{0\}) = 0$ gives the simplest representation with $\alpha(\{0\}) = \gamma(\{0\} \times (\mathbb{R}^d \setminus \{0\}))$ and $\beta(\{0\}) = \gamma((\mathbb{R}^d \setminus \{0\}) \times \{0\})$. $\square$

Proofs of Subsection 2.2

Proof of Lemma 2.14. Let $G \subset \mathbb{E}$ be open and suppose that $\operatorname{spt} \mu \cap G \neq \varnothing$. We need to show that $\operatorname{spt}(\mu_n) \cap G \neq \varnothing$ for all large n. For $\epsilon > 0$ small enough, $\tilde{G} = G \cap \operatorname{int}\left(\mathbb{R}^d \setminus \mathrm{B}_{0,\epsilon}\right) \neq \varnothing$ and $\tilde{G}$ is an open set bounded away from 0. Since $\mu \in \mathrm{M}_0(\mathbb{E}) \subset \mathrm{M}_0(\mathbb{R}^d)$ we have $\operatorname{spt} \mu \neq \{0\}$ according to Remark 1.7, so we can also suppose $\tilde{G} \cap \operatorname{spt} \mu \neq \varnothing$. It follows from the Portmanteau lemma for M_0-convergence (see Theorem 2.1 in Lindskog et al. (2014)) that $\liminf_{n\to\infty} \mu_n(G) \geq \liminf_{n\to\infty} \mu_n(\tilde{G}) \geq \mu(\tilde{G}) > 0$. Since $\operatorname{spt}(\mu_{(n)}) \cap G \neq \varnothing$ if and only if $\mu_{(n)}(G) > 0$, the first part of the Lemma is proved. The set $\{(\mu, F) : \operatorname{spt} \mu \subset F\}$ is closed in view of Lemma D.8(iv) in Segers (2022) since $\mathbb{E}$ is a LCHS space. $\square$

Proof of Lemma 2.15. We deal with the different assertions separately:

(a) As $\Gamma_{0,cm}(\mu, \nu) = \Gamma_0(\mu, \nu) \cap \mathrm{M}_{0,cm}$, it suffices to show that $\Gamma_0(\mu, \nu)$ and $\mathrm{M}_{0,cm}$ are closed.

First we show that the set $\Gamma_0(\mu, \nu)$ is closed. The argument used in Segers (2022) - continuous mapping theorem with the usual projection proj_i - cannot be used directly since $(\mathrm{proj}_1)_\# \gamma \neq \mu$, $(\mathrm{proj}_2)_\# \gamma \neq \nu$ in general.

According to Lemma 2.2, the map $\varphi_i : \gamma \in \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d) \mapsto (\mathrm{proj}_{i\#}\, \gamma)_{|\mathbb{R}^d \setminus \{0\}}$ is continuous for $i = 1, 2$. Let $\gamma_n \in \Gamma_0(\mu, \nu), n \geq 1$ be a sequence M_0-converging to γ, then $\varphi_1(\gamma) = \lim_{n\to\infty} \varphi_1(\gamma_n) = \mu$, $\varphi_2(\gamma) = \lim_{n\to\infty} \varphi_2(\gamma_n) = \nu$ by continuity, so $\gamma \in \Gamma_0(\mu, \nu)$ and $\Gamma_0(\mu, \nu)$ is closed in $\mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$.

Next, we have $\mathrm{M}_{0,cm} = \operatorname{spt}^{-1}(\mathcal{F}_{\mathrm{cm}})$ where the map $\operatorname{spt} : \mathrm{M}_0(\mathbb{R}^d) \to \mathcal{F}(\mathbb{R}^d)$ sends a measure in M_0 to its support. The set $\operatorname{spt}^{-1}(\mathcal{F}_{\mathrm{cm}})$ is closed by lower semi-continuity of spt (Lemma 2.14) and the fact that the complement of $\mathcal{F}_{\mathrm{cm}}$ can be written as a union over finite intersections of sets of the form $\mathcal{F}_G = \{F \in \mathcal{F}(\mathbb{R}^d) : F \cap G \neq \varnothing\}$ for open $G \subset \mathbb{R}^d \times \mathbb{R}^d$ (Lemma 3.2 in Segers (2022)).

(b) If $\gamma_n \in \Gamma_0(\mu_n, \nu_n)$ for all n and if $\mu_n \xrightarrow{0} \mu$, $\nu_n \xrightarrow{0} \nu$ and $\gamma_n \xrightarrow{0} \gamma$, then also $\gamma \in \Gamma_0(\mu, \nu)$ using arguments similar to those used for (a) and relying on the continuity of φ_i. It follows that $\{(\mu, \nu, \gamma) : \mu, \nu \in \mathrm{M}_0, \gamma \in \Gamma_0(\mu, \nu)\}$ is closed in $\mathrm{M}_0(\mathbb{R}^d) \times \mathrm{M}_0(\mathbb{R}^d) \times \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$. The intersection of the former set with the closed set $\mathrm{M}_0(\mathbb{R}^d) \times \mathrm{M}_0(\mathbb{R}^d) \times \mathrm{M}_{0,cm}$ is the set

$$\{(\mu, \nu, \gamma) : \mu, \nu \in \mathrm{M}_0, \gamma \in \Gamma_{0,cm}(\mu, \nu)\}$$

which is thus closed as well.

(c) Let $\epsilon > 0$. Since K and L are compact in $\mathrm{M}_0(\mathbb{R}^d)$, Prohorov's characterization of relative compactness in $\mathrm{M}_0(\mathbb{R}^d)$ in Theorem 1.9 yields the existence of two sequences $r_i^K, i \geq 1$, $r_i^L, i \geq 1$ decreasing to zero, and for each $i \geq 1$ there exist compacts sets C_i^K, C_i^L in $\mathbb{R}^d \setminus \mathrm{B}_{0,r_i^K}$ and $\mathbb{R}^d \setminus \mathrm{B}_{0,r_i^L}$ respectively, such that $\forall (\mu, \nu) \in K \times L$ we have

$$\mu\left(\mathbb{R}^d \setminus (\mathrm{B}_{0,r_i^K} \cup C_i^K)\right) \leq \epsilon/2 \,,\; \nu\left(\mathbb{R}^d \setminus (\mathrm{B}_{0,r_i^L} \cup C_i^L)\right) \leq \epsilon/2$$

and

$$\sup_{\mu \in K} \mu\left(\mathbb{R}^d \setminus \mathrm{B}_{0,r_i^K}\right) < \infty \,,\; \sup_{\nu \in L} \nu\left(\mathbb{R}^d \setminus \mathrm{B}_{0,r_i^L}\right) < \infty.$$

It is easy to see that we can choose a common sequence by taking $r_i = \max(r_i^K, r_i^L), i \geq 1$ and intersecting both C_i^L and C_i^K with $\mathbb{R}^d \setminus \mathrm{B}_{0,r_i}$, which is closed and smaller than both $\mathbb{R}^d \setminus \mathrm{B}_{0,r_i^L}$ and $\mathbb{R}^d \setminus \mathrm{B}_{0,r_i^K}$, to get two new compacts satisfying the characterization of compactness in $\mathrm{M}_0(\mathbb{R}^d)$ recalled above.

Since we are working with the Euclidean distance, it is clear that for every $s > \sqrt{2}$ we have

$$\begin{aligned}\mathbb{R}^{2d} \setminus \mathrm{B}_{0,s\,r_i} &\subset \left[\mathbb{R}^{2d} \setminus (\mathrm{B}_{0,r_i} \times \mathbb{R}^d)\right] \cup \left[\mathbb{R}^{2d} \setminus (\mathbb{R}^d \times \mathrm{B}_{0,r_i})\right] \\ &= \left[(\mathbb{R}^d \setminus \mathrm{B}_{0,r_i}) \times \mathbb{R}^d\right] \cup \left[\mathbb{R}^d \times (\mathbb{R}^d \setminus \mathrm{B}_{0,r_i})\right]\end{aligned}$$

We choose $s = 2$, whence for every $\gamma \in \Gamma_0(\mu, \nu)$ we can bound from above, using the fact that γ is a zero-coupling,

$$\gamma(\mathbb{R}^{2d} \setminus \mathrm{B}_{0,2r_i}) \leq \sup_{\mu \in K} \mu\left(\mathbb{R}^d \setminus \mathrm{B}_{0,r_i}\right) + \sup_{\nu \in L} \nu\left(\mathbb{R}^d \setminus \mathrm{B}_{0,r_i}\right) < \infty.$$

Let $\gamma \in \Gamma_0(\mu, \nu)$, $\tilde{C}_i = (C_i^K \cup \mathrm{cl}\,\mathrm{B}_{0,r_i}) \times (C_i^L \cup \mathrm{cl}\,\mathrm{B}_{0,r_i})$ and $C_i = \tilde{C}_i \setminus \mathrm{B}_{0,2r_i}$. Note that C_i is compact in $\mathbb{R}^d$ and included in $\mathbb{R}^{2d} \setminus \mathrm{B}_{0,2r_i}$. A simple computation yields $\mathbb{R}^{2d} \setminus (\mathrm{B}_{0,2r_i} \cup C_i) = \mathbb{R}^{2d} \setminus (\mathrm{B}_{0,2r_i} \cup \tilde{C}_i)$ and using the inclusion above involving $\mathbb{R}^{2d} \setminus \mathrm{B}_{0,2\,r_i}$ we get

$$\mathbb{R}^{2d} \setminus (\mathrm{B}_{0,2r_i} \cup C_i) \subset \left[\tilde{C}_i \cup \mathrm{B}_{0,r_i} \times \mathbb{R}^d\right]^c \cup \left[\tilde{C}_i \cup \mathbb{R}^d \times \mathrm{B}_{0,r_i}\right]^c.$$

Writing $\tilde{C}_i = \left(C_i^K \cup \left[\operatorname{cl} \mathrm{B}_{0,r_i} \times \mathbb{R}^d\right]\right) \cap \left(\mathbb{R}^d \times \left[C_i^L \cup \operatorname{cl} \mathrm{B}_{0,r_i}\right]\right)$ yields

$$\begin{aligned}\mathbb{R}^{2d} \setminus (\mathrm{B}_{0,2r_i} \cup C_i) \subset & \left(\left[C_i^K \cup \operatorname{cl} \mathrm{B}_{0,r_i}\right]^c \times \mathbb{R}^d\right) \ \cup \ \left(\mathbb{R}^d \times \left[C_i^L \cup \operatorname{cl} \mathrm{B}_{0,r_i})\right]^c\right) \\ & \cup \left[\left((C_i^K \cup \operatorname{cl} \mathrm{B}_{0,r_i}) \times \mathbb{R}^d\right) \cup \left(\mathbb{R}^d \times \mathrm{B}_{0,r_i})\right]^c \\ & \cup \left[\left(\mathbb{R}^d \times (C_i^L \cup \operatorname{cl} \mathrm{B}_{0,r_i})\right) \cup \left(\mathrm{B}_{0,r_i} \times \mathbb{R}^d\right)\right]^c\end{aligned}$$

whence

$$\mathbb{R}^{2d} \setminus (\mathrm{B}_{0,2r_i} \cup C_i) \subset \left(\left[C_i^K \cup \mathrm{B}_{0,r_i}\right]^c \times \mathbb{R}^d\right) \cup \left(\mathbb{R}^d \times \left[C_i^L \cup \mathrm{B}_{0,r_i}\right]^c\right).$$

The last inclusion allows one to write the upper bound

$$\gamma(\mathbb{R}^{2d} \setminus (\mathrm{B}_{0,2r_i} \cup C_i)) \leq \mu\left(\mathbb{R}^d \setminus (\mathrm{B}_{0,r_i} \cup C_i^K)\right) + \nu\left(\mathbb{R}^d \setminus (\mathrm{B}_{0,r_i} \cup C_i^L)\right) \leq \epsilon$$

where we used that γ is a zero-coupling between μ and ν.

Therefore, by Prohorov's characterization already used above, the set $M = \bigcup_{\mu \in K, \nu \in L} \Gamma_0(\mu, \nu)$ is relatively compact in $\mathrm{M}_0(\mathbb{R}^{2d})$.

The set

$$\begin{aligned}&\{(\mu, \nu, \gamma) \ : \ \mu \in K, \nu \in L, \gamma \in \Gamma_0(\mu, \nu)\} \\ &\qquad = \{(\mu, \nu, \gamma) \ : \ \gamma \in \Gamma_0(\mu, \nu)\} \cap \left(K \times L \times \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)\right)\end{aligned}$$

is closed by (2) and contained in the relatively compact set $K \times L \times M$, and is therefore compact. □

Proof of Lemma 2.16. We apply Lemma A.5 in de Valk and Segers (2018) to $\tilde{\gamma}$ which is a positive Borel measure on $\mathbb{R}^d \times \mathbb{R}^d$ to get that $\operatorname{spt}(\operatorname{proj}_{1\#} \tilde{\gamma}) = \operatorname{cl}(\operatorname{proj}_1(\operatorname{spt} \tilde{\gamma}))$. Remember from Section 1.1 that $\operatorname{spt} \gamma$ is defined as $\operatorname{spt} \tilde{\gamma}$ and that $\operatorname{proj}_{1\#} \tilde{\gamma} = \operatorname{proj}_{1\#} \gamma$. Thus the first equality can be rewritten as $\operatorname{spt}(\operatorname{proj}_{1\#} \gamma) = \operatorname{cl}(\operatorname{proj}_1(\operatorname{spt} \gamma))$. Since $\mu = \operatorname{proj}_{1\#} \gamma|_{\mathbb{R}^d \setminus \{0\}}$ we have $\operatorname{spt} \mu \subset \operatorname{spt} \operatorname{proj}_{1\#} \gamma$. As a consequence, if $\operatorname{spt} \gamma \subset T$ for some $T \subset \mathbb{R}^d \times \mathbb{R}^d$, then $\operatorname{spt} \mu \subset \operatorname{spt} \operatorname{proj}_{1\#} \gamma \subset \operatorname{cl}(\operatorname{dom} T)$.

Assume T is maximal monotone. By Theorem 12.41 (near convexity of domains and ranges) in Rockafellar and Wets (1998), there exists a convex subset C of $\mathbb{R}^d$

such that $C \subset \operatorname{dom} T \subset \operatorname{cl} C$. As $\operatorname{spt} \mu \subset \operatorname{cl}(\operatorname{dom} T)$, we have $\operatorname{spt} \mu \subset \operatorname{cl} C$, hence $\operatorname{int}(\operatorname{spt} \mu) \subset \operatorname{int} \operatorname{cl} C$. Moreover, the convexity of C yields $\operatorname{ri} C = \operatorname{ri} \operatorname{cl} C$. As a consequence, we can write

$$\begin{aligned} \operatorname{int}(\operatorname{spt} \mu) &\subset \operatorname{int} \operatorname{cl} C \subset \operatorname{ri} \operatorname{cl} C \\ &= \operatorname{ri} C \subset C \subset \operatorname{dom} T, \end{aligned}$$

thus the statement is proved.

□

Proof of Lemma 2.20. Let $\hat{\mu} = \operatorname{proj}_{1\#} \tilde{\gamma}$ and $\hat{\nu} = \operatorname{proj}_{2\#} \tilde{\gamma}$. According to Lemma 2.16, it is immediate that both $\tilde{V} = \operatorname{int} \operatorname{spt} \hat{\mu}$ and V are contained in both $\operatorname{dom} S$ and $\operatorname{dom} T$. The rest of the proof in Segers (2022) can then be reused directly since it exclusively relies on properties of maximal monotone maps.

Let $A \subset \tilde{V}$ be a non-empty open ball. It is sufficient to show that $S(x) = T(x)$ for all $x \in A$. To do so, we apply the criterion in Corollary 1.5 in Alberti and Ambrosio (1999): it is sufficient to show that $S(x) \cap T(x) \neq \varnothing$ for all x in a dense subset of A. Let $U \subset A$ be open and non-empty. By the said criterion, we are done if we can find $x \in U$ such that $S(x) \cap T(x) \neq \varnothing$. We have $\tilde{\gamma}(U \times \mathbb{R}^d) = \hat{\mu}(U) > 0$, since otherwise $\operatorname{spt} \hat{\mu}$ would be disjoint with U, contradicting $U \subset A \subset \operatorname{int}(\operatorname{spt} \hat{\mu})$. Since $U \times \mathbb{R}^d$ is open, $\operatorname{spt} \tilde{\gamma}$ intersects $U \times \mathbb{R}^d$. Moreover, $\operatorname{spt} \tilde{\gamma} \cap (U \times \mathbb{R}^d) \subset \operatorname{spt} \tilde{\gamma} \subset S \cap T$. For $(x, y) \in \operatorname{spt} \tilde{\gamma} \cap (U \times \mathbb{R}^d)$, we thus have $x \in U$ together with $y \in S(x) \cap T(x)$. The latter intersection is thus not empty, as required. Since $\operatorname{spt} \mu \subset \operatorname{spt} \hat{\mu}$, we have $S(x) = T(x)$ for all $x \in V$ too.

The statements about W follow by switching the roles of $\tilde{\mu}$ and $\tilde{\nu}$, noting that $\operatorname{dom} S^{-1} = \operatorname{rge} S$. □

Remark A.1. *In the proof of the above lemma, note that* $\tilde{V} = V$. *Indeed, from Remark 2.17 we have* $\operatorname{spt} \phi_1(\gamma) \subset \operatorname{spt} \varphi_1(\gamma) \cup \{0\}$ *and* $\operatorname{spt} \varphi_1(\gamma) \subset \operatorname{spt} \phi_1(\gamma)$. *If* 0 *lies in* $\operatorname{spt} \phi_1(\gamma)$ *but not in* $\operatorname{spt} \varphi_1(\gamma)$ *which is closed in* $\mathbb{R}^d$*, then there exists some* $\epsilon > 0$ *such that* $\mathrm{B}_{0,\epsilon} \subset \operatorname{spt} \varphi_1(\gamma)^c$. *Thus we can write* $0 \notin \tilde{V} = \operatorname{int} \operatorname{spt} \phi_1(\gamma)$ *to get* $V = \tilde{V}$. *Otherwise, we have* $\operatorname{spt} \phi_1(\gamma) = \operatorname{spt} \varphi_1(\gamma)$ *and the result is immediate.*

Proof of Theorem 2.21. We deal with the assertions (a), (b) and (c) separately:

(a) The first assertion is a direct application of Lemma 2.15(3). If μ_n, ν_n M_0-converge to $\mu, \nu \in \mathrm{M}_0$ respectively, then the sets $K = \{\mu_n : n \geq 1\} \cup \{\mu\}$ and $L = \{\nu_n : n \geq 1\} \cup \{\nu\}$ are both compact as is $Z = \{(a, b, \zeta) : a \in K, b \in L, \zeta \in \Gamma_0(a, b)\}$. Whence $\{\gamma_n : n \geq 1\}$ is relatively compact in $\mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$ and by Lemma 2.15(2) any limit point lies in $\Gamma_0(\mu, \nu)$ $(\Gamma_{0,cm}(\mu, \nu))$ since for each $n \geq 1$, we have $\gamma_n \in \Gamma_0(\mu, \nu)$ $(\gamma_n \in \Gamma_{0,cm}(\mu, \nu))$.

(b) Since $\mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d)$ is compact, there exist an infinite subset N of $\mathbb{N}$ and some T in $\mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d)$ such that T_n $\mathcal{F}$-converges to T as $n \to \infty$ in N. Thanks to Lemma 3.2 in Segers (2022) we know that $\mathcal{F}_{\mathrm{mcm}}(\mathbb{R}^d \times \mathbb{R}^d) \cup \{\varnothing\}$ is closed. As a consequence, $T_n \in \mathcal{F}_{\mathrm{mcm}}(\mathbb{R}^d \times \mathbb{R}^d), n \geq 1$ implies $T \in \mathcal{F}_{\mathrm{mcm}}(\mathbb{R}^d \times \mathbb{R}^d) \cup \{\varnothing\}$. Since μ and ν are nonzero measures and $\gamma \in \Gamma_{0,cm}(\mu, \nu)$, we have $\operatorname{spt}\gamma \neq \varnothing$. From (c), whose proof is given below, $\operatorname{spt}\gamma \subset T$. Therefore T is non empty, hence T lies in $\mathcal{F}_{\mathrm{mcm}}(\mathbb{R}^d \times \mathbb{R}^d)$.

(c) Applying Lemma 2.14 to the sequences $\gamma_n, n \geq 1$ and $T_n, n \geq 1$, we get $\operatorname{spt}\gamma \subset T$. □

Proof of Corollary 2.23. The first assertion is a direct consequence of of Theorem 2.21 since if all limits points of $\gamma_n, n \geq 1$ are equal to γ, then γ_n M_0-converges to γ in $\mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$ along the whole sequence. According to Theorem 2.21, any limit point T_∞ of $T_n, n \geq 1$ belongs to $\mathcal{F}_{\mathrm{mcm}}(\mathbb{R}^d \times \mathbb{R}^d) \subset \mathcal{F}_{\mathrm{mm}}(\mathbb{R}^d \times \mathbb{R}^d)$ since $\gamma_n \in \Gamma_{0,cm}(\mu_n, \nu_n)$, and by point (c) of the same theorem we have $\operatorname{spt}\gamma \subset T_\infty$. According to Lemma 2.20, $T \llcorner V = F \llcorner V$ for every other $F \in \mathcal{F}_{\mathrm{mm}}(\mathbb{R}^d \times \mathbb{R}^d)$ such that $\operatorname{spt}\gamma \subset F$. Whence $T_\infty \llcorner V = T \llcorner V$ and $T_n \llcorner V$ $\mathcal{F}$-converges to $T \llcorner V$ along the whole sequence. □

Proof of Theorem 2.24. One just needs to substitute $\mathcal{P}(\mathbb{R}^d)$ ($\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$) with $\mathrm{M}_0(\mathbb{R}^d)$ ($\mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$), Lemma D.11 with Lemma 2.14, Lemma 4.2 with Lemma 2.20 and Lemma 4.1 with Lemma 2.15 in the proof of Theorem 1.2.a) in Segers (2022). We reproduce the proof here.

We first show that the sequence of laws of the random triples (μ_n, ν_n, γ_n) in $\mathrm{M}_0(\mathbb{R}^d) \times \mathrm{M}_0(\mathbb{R}^d) \times \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$ is uniformly tight. Let $\epsilon > 0$. Since the laws of the random measures μ_n and ν_n converge in distribution in $\mathcal{P}(\mathrm{M}_0(\mathbb{R}^d))$, there exist compact sets $K, L \subset \mathrm{M}_0(\mathbb{R}^d)$ such that $\mathsf{P}(\mu_n \in K) \geq 1 - \epsilon/2$ and $\mathsf{P}(\nu_n \in L) \geq 1 - \epsilon/2$ for all integer n. By construction, for all $n \geq 1$ we have $\gamma_n \in \Gamma_{0,cm}(\mu_n, \nu_n)$, hence the probability that the random triple (μ_n, ν_n, γ_n) belongs to the set $C = \{(\mu', \nu', \gamma') : \mu' \in K, \nu' \in L, \gamma' \in \Gamma_0(\mu', \nu')\}$ is at least $1 - \epsilon$. Moreover, Lemma 2.15(c) implies that the set C is compact. As a consequence the sequence of laws of the triple (μ_n, ν_n, γ_n) is uniformly tight.

Suppose that along some subsequence, (μ_n, ν_n, γ_n) converges in distribution to the possibly random triple (μ', ν', γ'). By assumption μ_n, ν_n converge in distribution to μ, ν respectively, thus we necessarily have $\mu' = \mu$ and $\nu' = \nu$ almost surely. The laws of (μ_n, ν_n, γ_n) are concentrated on the set $\{(\mu_0, \nu_0, \gamma_0) : \gamma_0 \in \Gamma_{0,cm}(\mu_0, \nu_0)\}$, which is closed in $\mathrm{M}_0(\mathbb{R}^d) \times \mathrm{M}_0(\mathbb{R}^d) \times \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$ by Lemma 2.15(b). The usual Portmanteau theorem for convergence in distribution over a Polish space yields $\gamma' \in \Gamma_{0,cm}(\mu, \nu)$ with probability one. Now $\Gamma_{0,cm}(\mu, \nu)$ is equal to the singleton $\{\gamma\}$ by assumption, so that actually $(\mu', \nu', \gamma') = (\mu, \nu, \gamma)$ with probability one. Since this is true for every convergent subsequence of (μ_n, ν_n, γ_n), we obtain that the whole sequence (μ_n, ν_n, γ_n) converges in distribution to the degenerate law at (μ, ν, γ).

Next, let $V = \operatorname{int}(\operatorname{spt}\mu)$ and recall that the Fell space $\mathcal{F}(V \times \mathbb{R}^d)$ is compact and Polish since $\mathbb{E} = V \times \mathbb{R}^d$ is an open subset of $\mathbb{R}^d$. To show that $T_n \llcorner V \xrightarrow{w} T \llcorner V$ in $\mathcal{F}(V \times \mathbb{R}^d)$, we apply Proposition 2.5 in Segers (2022). Let $\hat{T}$ be the weak limit in $\mathcal{F}$ of T_n as $n \to \infty$ in some infinite subset N of $\mathbb{N}$. Since $\gamma_n \xrightarrow{w} \gamma$ in $\mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$, the limit being deterministic, we also have $(\gamma_n, T_n) \to (\gamma, \hat{T})$ in distribution as $n \to \infty$ in N. In view of Lemma 2.14 applied to $\mathbb{E} = V \times \mathbb{R}^d$ and thanks to the Portmanteau theorem, we have $\mathsf{P}[\operatorname{spt}\gamma \subset \hat{T}] = 1$. By Lemma 2.20, on an event with probability one, $\hat{T}$ coincides on V with any other $T' \in \mathcal{F}_{\mathrm{mcm}}$ such that $\operatorname{spt}\gamma \subset T'$, in particular with the given T. Since this is true for the limit in distribution of any converging subsequence, criterion (ii) in Proposition 2.5 in Segers (2022) is fulfilled. We conclude that $T_n \llcorner V \xrightarrow{w} T \llcorner V$ in $\mathcal{F}(V \times \mathbb{R}^d)$. □

Proofs of Subsection 2.3

Proof of Theorem 2.26. We borrow from the proof of Theorem 4.4 in de Valk and Segers (2018) the construction of two sequences $\mu_n, \nu_n \in \mathrm{M}_0(\mathbb{R}^d), n \geq 1$ of finite and equal mass approximating respectively μ and ν in $\mathrm{M}_0(\mathbb{R}^d)$ and of a sequence $\tilde{\pi}_n \in \Pi_{cm}(\tilde{\mu}_n, \tilde{\nu}_n), n \geq 1$ of cyclically monotone couplings between $\tilde{\mu}_n$ and $\tilde{\nu}_n$.

If the common value of $\tilde{\mu}(\mathbb{R}^d)$ and $\tilde{\nu}(\mathbb{R}^d)$ is finite, we can normalize $\tilde{\mu}$ and $\tilde{\nu}$ to become probability measures on $\mathbb{R}^d$— the same idea was used in Remark 1.47—and apply the Main Theorem in McCann (1995). So assume that $\tilde{\mu}(\mathbb{R}^d) = \tilde{\nu}(\mathbb{R}^d) = +\infty$.

For $n \in \mathbb{N}$, let ν_n be the restriction of ν to the complement of $n^{-1}\overline{\mathrm{B}}_{0,1}$ the closed ball with center 0 and radius n^{-1}. Then $a_n = \nu_n(\mathbb{R}^d) = \nu(\mathbb{R}^d \setminus n^{-1}\overline{\mathrm{B}}_{0,1})$ is finite but grows to infinity as $n \to \infty$. Further, $\nu_n \xrightarrow{0} \nu$ as $n \to \infty$, since for every $f \in \mathrm{C}_{b,0}^+$, we have $\nu_n(f) = \nu(f)$ for every $n \in \mathbb{N}$ that is sufficiently large such that f vanishes on $n^{-1}\overline{\mathrm{B}}_{0,1}$. Note that ν_n lies in $\mathrm{M}_0(\mathbb{R}^d)$ since it is a restriction of $\nu \in \mathrm{M}_0(\mathbb{R}^d)$.

Let $\epsilon_n = \inf\{\epsilon > 0 : \mu(\mathbb{R}^d \setminus \epsilon\overline{\mathrm{B}}_{0,1}) \leq a_n\}$, which is well defined since $a_n = \nu(\mathbb{R}^d \setminus n^{-1}\overline{\mathrm{B}}_{0,1}) > 0$ for n large enough, and $\mu(\mathbb{R}^d \setminus \epsilon\overline{\mathrm{B}}_{0,1}) \to 0$ as $\epsilon \to \infty$. Necessarily $\mu(\mathbb{R}^d \setminus \epsilon_n\overline{\mathrm{B}}_{0,1}) \leq a_n$. Let $n \geq 1$. As $\mu \in \mathrm{M}_0(\mathbb{R}^d)$ and $\mu(\epsilon_n\overline{\mathrm{B}}_{0,1}) = +\infty$, there exists some positive δ such that $\delta < \epsilon_n$ and $\mu(\mathbb{R}^d \setminus \delta\overline{\mathrm{B}}_{0,1}) \geq a_n$. Moreover, the convergence $a_n \to \infty$ as $n \to \infty$ of the increasing sequence $a_n, n \geq 1$ implies the existence of some $N \geq n$ such that $\mu(\mathbb{R}^d \setminus \delta\overline{\mathrm{B}}_{0,1}) \leq a_N$. Thus, $\delta \geq \epsilon_N$ and we finally get $\mu(\mathbb{R}^d \setminus \epsilon_N\overline{\mathrm{B}}_{0,1}) \geq a_n$, which implies $\mu(\mathbb{R}^d \setminus \epsilon_n\overline{\mathrm{B}}_{0,1}) \to +\infty$ as $n \to \infty$ since the sequence ϵ_n is clearly decreasing. As a consequence, we necessarily have $\epsilon_n \to 0$ as $n \to \infty$. Put $m_n = a_n - \mu(\mathbb{R}^d \setminus \epsilon_n\overline{\mathrm{B}}_{0,1})$, let κ_n denote the Lebesgue-uniform distribution on $\epsilon_n \overline{\mathrm{B}}_{0,1}$ and let μ_n be the sum of $\mu(\,\cdot \setminus \epsilon_n \overline{\mathrm{B}}_{0,1})$ and $m_n \operatorname{res} \kappa_n$—we take $\operatorname{res} \kappa_n$ here to ensure that μ_n lies in $\mathrm{M}_0(\mathbb{R}^d)$. Then $\mu_n(\mathbb{R}^d) = a_n$ by construction and $\mu_n \xrightarrow{0} \mu$ as $n \to \infty$, again because for every $f \in \mathrm{C}^+_{b,0}$ we have $\mu_n(f) = \mu(f)$ for all sufficiently large n.

Recall from Section 1.1 that $\tilde{\mu}_n$, $\tilde{\nu}_n$ denote the Borel measures on $\mathbb{R}^d$ defined by $\operatorname{res} \tilde{\mu}_n = \mu_n$, $\operatorname{res} \tilde{\nu}_n = \nu_n$ and $\tilde{\mu}_n(\{0\}) = 0$, $\tilde{\nu}_n(\{0\}) = 0$ respectively. Apply Theorem 6 in McCann (1995) to the probability measures $\hat{\mu}_n = a_n^{-1} \tilde{\mu}_n$ and $\hat{\nu}_n = a_n^{-1} \tilde{\nu}_n$ to find $\hat{\pi}_n \in \Pi(\hat{\mu}_n, \hat{\nu}_n)$ with cyclically monotone support. Then $\tilde{\pi}_n = a_n \hat{\pi}_n$ belongs to $\Pi(\tilde{\mu}_n, \tilde{\nu}_n)$ and has the same, cyclically monotone support as $\hat{\pi}_n$. By Rockafellar and Wets' Theorem 1.37, there exists a closed convex function ψ_n on $\mathbb{R}^d$ such that $\operatorname{spt} \tilde{\pi}_n$ is contained in $T_n = \partial \psi_n \in \mathcal{F}_{\mathrm{mcm}}(\mathbb{R}^d \times \mathbb{R}^d)$.

One may notice that $\tilde{\pi}_n$ is a real coupling between μ_n and ν_n seen as Borel measures on $\mathbb{R}^d$ putting no mass on $\{0\}$. By construction $\tilde{\pi}_n$ puts no mass on $\{0\}, \{0\} \times \mathbb{R}^d$ and $\mathbb{R}^d \times \{0\}$ and its restriction $\gamma_n = \operatorname{res} \tilde{\pi}_n$ to $(\mathbb{R}^d \times \mathbb{R}^d) \setminus \{0\}$ lies in $\Gamma_{0,cm}(\mu_n, \nu_n)$ and $\operatorname{spt} \gamma_n = \operatorname{spt} \tilde{\pi}_n$.

Applying Corollary 2.21, there exist an infinite subset N of $\mathbb{N}$ and γ in $\Gamma_{0,cm}(\mu, \nu)$ such that γ_n M_0-converges to γ as $n \to \infty$ in N. $\square$

Proof of Theorem 2.27. The proof is largely inspired by the one of Proposition 10 in McCann (1995). We split $(\mathbb{R}^d \times \mathbb{R}^d) \setminus \{0\}$ into $(\mathbb{R}^d \setminus \{0\}) \times \mathbb{R}^d$ and $\{0\} \times (\mathbb{R}^d \setminus \{0\})$.

We first consider γ on subsets of $(\mathbb{R}^d \setminus \{0\}) \times \mathbb{R}^d$. Let

$$\begin{aligned} S &= \{(x, \nabla\psi) : x \in (\operatorname{dom} \nabla\psi) \cap (\mathbb{R}^d \setminus \{0\})\} \\ &= (\operatorname{dom} \nabla\psi \times \mathbb{R}^d) \cap \partial\psi \cap ((\mathbb{R}^d \setminus \{0\}) \times \mathbb{R}^d). \end{aligned}$$

Taking argument from the proof of Proposition 10 in McCann (1995), the measurability of $\nabla\psi$ is manifest since it coincides with the pointwise limit of a sequence of continuous approximants

$$\langle \nabla\psi(x), z \rangle = \lim_{n\to\infty} n\,(\psi(x + z/n) - \psi(x))$$

hence $\operatorname{dom} \nabla\psi$ belongs to $\mathcal{B}(\mathbb{R}^d \setminus \{0\})$. Since $\partial\psi$ is closed and contains $\operatorname{spt} \gamma$, the set S lies in $\mathcal{B}((\mathbb{R}^d \times \mathbb{R}^d) \setminus \{0\})$ and contains the set

$$\tilde{S} = ((\operatorname{dom} \nabla\psi \cap (\mathbb{R}^d \setminus \{0\})) \times \mathbb{R}^d) \cap \operatorname{spt} \gamma,$$

which is such that $\gamma(\tilde{S}^c) = \mu((\operatorname{dom} \nabla\psi \cap (\mathbb{R}^d \setminus \{0\}))^c) = 0$ as $\nabla\psi$ is defined μ-almost everywhere. Thus, for every $A \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$ and $B \in \mathcal{B}(\mathbb{R}^d)$ we have

$$\begin{aligned} \gamma(A \times B) = \gamma((A \times B) \cap S) &= \gamma((A \cap (\nabla\psi^{-1})B) \times \mathbb{R}^d) \\ &= \mu(A \cap (\nabla\psi^{-1})B) = [\mathrm{Id} \otimes \nabla\psi]_\# \mu(A \times B), \end{aligned}$$

so that the restriction of γ to $(\mathbb{R}^d \setminus \{0\}) \times \mathbb{R}^d$ is indeed $[\mathrm{Id} \otimes \nabla\psi]_{\#}\mu$.

We now deal with subsets of $\{0\} \times (\mathbb{R}^d \setminus \{0\})$. Let $B \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$ be bounded away from the origin in $\mathbb{R}^d$. We can write successively

$$\begin{aligned} \nu(B) &= \gamma(\mathbb{R}^d \times B) \\ &= \gamma((\mathbb{R}^d \setminus \{0\}) \times B) + \gamma(\{0\} \times B) \\ &= [\mathrm{Id} \otimes \nabla\psi]_{\#}\mu((\mathbb{R}^d \setminus \{0\}) \times B) + \gamma(\{0\} \times B), \end{aligned}$$

and since B, $(\mathbb{R}^d \setminus \{0\}) \times B$ are bounded away from the origin in $\mathbb{R}^d$, $\mathbb{R}^d \times \mathbb{R}^d$ respectively, it is clear that $\nu(B), \gamma((\mathbb{R}^d \setminus \{0\}) \times B) < +\infty$, whence we have $\gamma(\{0\} \times B) = \nu(B) - [\mathrm{Id} \otimes \nabla\psi]_{\#}\mu((\mathbb{R}^d \setminus \{0\}) \times B)$.

Let $A, B \in \mathcal{B}(\mathbb{R}^d)$ be such that at least one of them is bounded away from zero. From the discussion above, we can write

$$\begin{aligned} \gamma\left((A \setminus \{0\}) \times B\right) &= [\mathrm{Id} \otimes \nabla\psi]_{\#}\mu\left((A \setminus \{0\}) \times B\right) \\ \gamma\left((A \cap \{0\}) \times B\right) &= \delta_0(A)\left[\nu(B) - [\mathrm{Id} \otimes \nabla\psi]_{\#}\mu((\mathbb{R}^d \setminus \{0\}) \times B)\right] \end{aligned}$$

where we used in the last identity that $0 \in A$ implies by construction that B is bounded away from the origin. Therefore, as soon as either A or B is bounded away from the origin, we have

$$\begin{aligned} \gamma\left(A \times B\right) &= \gamma\left((A \setminus \{0\}) \times B\right) + \gamma\left((A \cap \{0\}) \times B\right) \\ &= [\mathrm{Id} \otimes \nabla\psi]_{\#}\mu\left((A \setminus \{0\}) \times B\right) + \delta_0(A)\left(\nu(B) - \nabla\psi_{\#}\mu(B)\right) \\ &= [\mathrm{Id} \ \otimes \nabla\psi]_{\#}\mu(A \times B) + \delta_0 \otimes [\nu - \nabla\psi_{\#}\mu](A \times B). \end{aligned}$$

As an element of $\mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$, γ is σ-finite and the products $A \times B$ considered generate the Borel sets on $(\mathbb{R}^d \times \mathbb{R}^d) \setminus \{0\}$, thus γ has the desired expression. $\square$

Proof of Theorem 2.29. Assume μ and ν have equal finite mass. The result is a direct application Main Theorem in McCann (1995) together with a simple rescaling argument. For the rest of the proof here we will suppose that μ and ν have infinite mass.

Since both μ and ν have infinite mass, it is clear that $0 \in \operatorname{spt}\mu \cap \operatorname{spt}\nu$. From Lemma 2.16, we have $\operatorname{spt}\operatorname{proj}_1\gamma = \operatorname{cl}(\operatorname{proj}_1(\operatorname{spt}\gamma))$, so we can write $\operatorname{spt}\mu = \operatorname{cl}(\operatorname{proj}_1(\operatorname{spt}\gamma))$ and $\operatorname{spt}\nu = \operatorname{cl}(\operatorname{proj}_2(\operatorname{spt}\gamma))$. Recall from Remark 2.28 that if $\gamma(\{0\} \times (\mathbb{R}^d \setminus \{0\})) = 0$ for some $\gamma \in \Gamma_{0,cm}(\mu, \nu)$, we have $\gamma = [\mathrm{Id} \ \otimes \nabla\psi]_{\#}\mu$ for every lsc convex function ψ such that $\operatorname{spt}\gamma \subset \partial\psi$. Let $\gamma, \zeta \in \Gamma_{h,cm}(\mu, \nu)$ be zero-couplings associated with the closed convex functions ψ, ϕ respectively. It is clear that $\tilde{\gamma} = [\mathrm{Id} \otimes \nabla\psi]_{\#}\tilde{\mu}$ and

$\tilde{\zeta} = [\mathrm{Id} \otimes \nabla\phi]_{\#}\tilde{\mu}$ so the associated $\bar{\nu}$ for these representations satisfy $\bar{\nu}_\gamma(\{0\}) = \gamma\left((\mathbb{R}^d \setminus \{0\}) \times \{0\}\right)$ and $\bar{\nu}_\zeta(\{0\}) = \zeta\left((\mathbb{R}^d \setminus \{0\}) \times \{0\}\right)$.

To prove $\nabla\psi = \nabla\phi$ μ-almost everywhere, it is sufficient to show that $\nabla\psi_{\#}\tilde{\mu} = \nabla\phi_{\#}\tilde{\mu}$ and then apply Theorem 3.1(Uniqueness) in de Valk and Segers (2018). Since by definition of a zero-coupling we have $\nu = \varphi_2\gamma = \varphi_2\zeta$, i.e., $\nu = \nabla\psi_{\#}\mu|_{\mathbb{R}^d\setminus\{0\}} = \nabla\phi_{\#}\mu|_{\mathbb{R}^d\setminus\{0\}}$, it suffices to prove that $\gamma((\mathbb{R}^d \setminus \{0\}) \times \{0\}) = \zeta((\mathbb{R}^d \setminus \{0\}) \times \{0\})$, i.e., $\tilde{\mu}([\nabla\psi]^{-1}(\{0\})) = \tilde{\mu}([\nabla\phi]^{-1}(\{0\}))$.

By Theorem 1.29 and Remark 1.30 there exists sets A_ψ, A_ϕ such that $\mu(A_\psi^c) = \mu(A_\phi^c) = 0$ and $\nabla\psi$, $\nabla\phi$ are defined on A_ψ, A_ϕ respectively. Moreover, as $\gamma \in \Gamma_0(\mu,\nu)$, γ has infinite mass too, hence $0 \in \operatorname{spt}\gamma$ and $0 \in \operatorname{proj}_1 \operatorname{spt}\gamma$. Thus $(\operatorname{proj}_1 \operatorname{spt}\gamma)^c \subset \mathbb{R}^d \setminus \{0\}$ and we can write successively

$$\begin{aligned}
\mu\left((\operatorname{proj}_1 \operatorname{spt}\gamma)^c\right) &= \gamma\left(\operatorname{spt}\gamma \cap \left[(\operatorname{proj}_1 \operatorname{spt}\gamma)^c \times \mathbb{R}^d\right]\right) \\
&\leq \gamma\left([\operatorname{proj}_1 \operatorname{spt}\gamma \times \operatorname{proj}_2 \operatorname{spt}\gamma] \cap \left[(\operatorname{proj}_1 \operatorname{spt}\gamma)^c \times \mathbb{R}^d\right]\right) \\
&= 0.
\end{aligned}$$

A similar argument yields $\mu\left((\operatorname{proj}_1 \operatorname{spt}\zeta)^c\right) = 0$ too. As a consequence $\mu(A^c) = 0$ for

$$A = [\operatorname{proj}_1 \operatorname{spt}\gamma] \cap [\operatorname{proj}_2 \operatorname{spt}\zeta] \cap A_\psi \cap A_\phi.$$

Let us consider the sets

$$\begin{aligned}
Z_\psi &:= A \cap [\partial\psi]^{-1}(\{0\}), \\
Z_\phi &:= A \cap [\partial\phi]^{-1}(\{0\}), \\
\chi &:= \{x \in A \,:\, \operatorname{spt}\nu \cap \operatorname{cl} H_+(x) = \{0\}\}, \\
\xi &:= \{x \in A \,:\, \operatorname{spt}\nu \cap H_+(x) = \varnothing\}.
\end{aligned}$$

We claim

$$\chi \subset Z_\psi \cap Z_\phi \subset Z_\psi \cup Z_\phi \subset \xi.$$

and $\mu(\xi \setminus \chi) = 0$, implying $\mu(\xi \setminus (Z_\psi \cup Z_\phi)) = 0$. This will allow us te substitute both Z_ψ and Z_ϕ by the same set ξ or χ.

To begin with, suppose that we have $(x,0), (a,b) \in \operatorname{spt}\gamma$, then by homogeneity and monotonicity

$$\langle \lambda x - a, -b \rangle \geq 0, \qquad \forall \lambda > 0.$$

Dividing by λ and taking the limit as $\lambda \to \infty$ we get $\langle x, b \rangle \leq 0$. Since $y \in \mathbb{R}^d \mapsto \langle x, y \rangle$ is continuous, this result extends to any b lying in $\operatorname{spt}\nu = \operatorname{cl}\operatorname{proj}_2 \operatorname{spt}\gamma$. We just proved that $(x,0) \in \operatorname{spt}\gamma$ implies that $\operatorname{spt}\nu \cap H_+(x)$ is empty. In particular, this yields $Z_\psi \subset \xi$. Using the same arguments with minor changes, we get $Z_\phi \subset \xi$ too.

On the other hand, if (x,y) lies in $\operatorname{spt}\gamma$, we have $\langle x,y\rangle \geq 0$ since the monotonicity of $\operatorname{spt}\gamma$ implies $(\lambda-1)^2\langle x,y\rangle \geq 0$ for every $\lambda>0$. If x belongs to $\chi \subset A$, we have $x \in \operatorname{proj}_1 \operatorname{spt}\gamma$ and $\operatorname{spt}\nu \cap \operatorname{cl} H_+(x) = \{0\}$. Thus we necessarily have the identity $[\operatorname{spt}\gamma](x) = \{0\}$, where $\operatorname{spt}\gamma$ is seen as a multivalued map, hence the inclusion $\chi \subset Z_\psi$ is proved. Again, the same argument yields $\chi \subset Z_\phi$ too.

We prove now that the set $\xi \setminus \chi$ is contained in the boundary of a convex set. As a consequence $\xi\setminus\chi$ will be of Hausdorff dimension at most $d-1$ and have a null μ-mass by assumption. First, note that by construction it is clear that ξ is a convex set. Indeed, if $x, a \in \xi$ and $\lambda > 0$, then $\langle \lambda a + (1-\lambda)x, y\rangle = \lambda\langle a,y\rangle + (1-\lambda)\langle x,y\rangle \leq 0$ for every $y \in \operatorname{spt}\nu$, hence $\lambda a + (1-\lambda)x$ lies in ξ which is convex. Finally, we prove by contradiction that $\xi\setminus\chi$ is a subset of $\operatorname{bnd}\xi$. Let $x \in (\xi\setminus\chi)\cap \operatorname{int}\xi$, then there exists some $\epsilon>0$ such that $\mathrm{B}_{x,\epsilon} \subset \xi$. As $x \in \xi\setminus\chi$, there exists some y lying in $\operatorname{spt}\nu$ such that $y \neq 0$ and $\langle x,y\rangle = 0$. Let $\tilde{x}$ belong to $\mathrm{B}_{x,\epsilon}$. Writing $\langle \tilde{x}, y\rangle = \langle \tilde{x}-x+x, y\rangle$ we get $\langle \tilde{x}-x, y\rangle \leq 0$ for every $\tilde{x}$ in $\mathrm{B}_{x,\epsilon}$ which is absurd since $y \neq 0$. Therefore $\xi\setminus\chi \subset \operatorname{bnd}\xi$ where ξ is convex as announced.

To conclude, note that $\tilde{\mu}\left([\nabla\psi]^{-1}(\{0\})\right) = \mu\left(A \cap [\partial\psi]^{-1}(\{0\})\right)$, and likewise $\tilde{\mu}\left([\nabla\phi]^{-1}(\{0\})\right) = \mu\left(A \cap [\partial\psi]^{-1}(\{0\})\right)$. As a consequence, using the approximation of both Z_ψ and Z_ϕ by the same set ξ or χ from the above discussion, we can write

$$\tilde{\mu}\left([\nabla\psi]^{-1}(\{0\})\right) = \mu(\chi) = \mu(\xi) = \tilde{\mu}\left([\nabla\phi]^{-1}(\{0\})\right).$$

Applying Theorem 3.1 (Uniqueness) in de Valk and Segers (2018) in the way announced at the beginning of the proof, we get $\nabla\psi = \nabla\phi$ μ-almost everywhere for every closed convex functions ψ, ϕ satisfying $\operatorname{spt}\gamma \subset \partial\psi$, $\operatorname{spt}\zeta \subset \partial\phi$ for some $\gamma,\zeta \in \Gamma_{h,cm}(\mu,\nu)$. By Theorem 2.27 and since $\gamma(\{0\}\times(\mathbb{R}^d\setminus\{0\})) = 0$, $\zeta(\{0\}\times(\mathbb{R}^d\setminus\{0\})) = 0$ we have

$$\gamma = [\mathrm{Id}\ \otimes \nabla\psi]_{\#}\mu = [\mathrm{Id}\ \otimes\nabla\phi]_{\#}\mu = \zeta,$$

whence $\gamma = \zeta$ and $\Gamma_{h,cm}(\mu,\nu)$ contains at most one element. If γ lies in $\Gamma_{h,cm}(\mu,\nu)$, there exists some closed convex function ψ such that $\operatorname{spt}\gamma \subset \partial\psi$. Let $T = \nabla\psi$ which is defined μ-almost everywhere by Theorem 1.29 and Remark 1.30. From the discussion above it is clear that such function T satisfies the statement of the lemma we are proving.

□

Proof of Lemma 2.30. Let $(a,b) \in \{0\}\times\mathbb{R}^d$ and $(x,y) \in A$. Since A is homogenous, for $\lambda>0$ we have $(\lambda x, \lambda y) \in A$. By monotonicity of A it is therefore clear that as soon as $(a,b) \in A$ we have

$$\begin{aligned} \langle \lambda x - 0, \lambda y - b\rangle &\geq 0 \qquad \forall\lambda>0 \\ \text{i.e., } \lambda\left[\lambda\langle x,y\rangle - \langle x,b\rangle\right] &\geq 0 \qquad \forall\lambda>0 \end{aligned}$$

The last inequality fails in the following cases :

- $\langle x, y\rangle = 0$ and $\langle x, b\rangle > 0$
- $\langle x, y\rangle > 0$, $\langle x, b\rangle > 0$ and $\lambda < \langle x, b\rangle/\langle x, y\rangle$
- $\langle x, y\rangle < 0$ and $\lambda > \max(\langle x, b\rangle/\langle x, y\rangle, 0)$

One should notice that the case $\langle x, y\rangle < 0$ is impossible due to the monotonicity of A. Indeed, using same ideas as above we should have, for every $\lambda > 0$ and $(x, y) \in A$, $(\lambda - 1)^2\langle x, y\rangle = \langle \lambda x - x, \lambda y - y\rangle \geq 0$. Therefore, we just proved that for every $b \in H_+(x)$, $(0, b)$ cannot belong to A. $\square$

Proof of Theorem 2.32. One just needs to apply Lemma 2.30 to $A = \operatorname{spt}\gamma$. Since μ vanishes on sets of Hausdorff dimension at most $d-1$, there exists $x \in \operatorname{int}\operatorname{spt}\mu \subset \operatorname{int} E$, and thanks to Lemma 2.16 $\operatorname{int}\operatorname{spt}\mu \subset \operatorname{dom}\operatorname{spt}\gamma$ so we have $x \in \operatorname{dom}(\operatorname{spt}\gamma)$. We claim that for every $y \in \operatorname{spt}\nu \setminus \{0\} \subset E \setminus \{0\}$ we have $\langle x, y\rangle > 0$. Indeed, $E \subset \operatorname{cl} H_+(y)$ so $x \in \operatorname{int} E \subset \operatorname{int}\operatorname{cl} H_+(y) = H_+(y)$ whence $E \setminus \{0\} \subset H_+(x)$.

According to Lemma 2.30, $\operatorname{spt}\gamma \cap (\{0\} \times H_+(x)) = \varnothing$, so $\operatorname{spt}\gamma \cap (\{0\} \times \operatorname{spt}\nu) = \varnothing$. Therefore $\gamma(\{0\} \times (\mathbb{R}^d \setminus \{0\})) = \gamma(\{0\} \times \operatorname{spt}\nu \setminus \{0\}) = 0$ and $\Gamma_{0,cm}(\mu, \nu) \subset \Gamma_{h,cm}(\mu, \nu)$. $\square$

Proof of Theorem 2.34. We apply Lemma 2.30 to $A = \operatorname{spt}\gamma$. For each $1 \leq i \leq k$, we get that $\operatorname{spt}\gamma \cap (\{0\} \times H_+(x_i))$ is empty. As a consequence, $\operatorname{spt}\gamma \cap (\{0\} \times \mathbb{R}^d \setminus \{0\}) = \operatorname{spt}\gamma \cap (\{0\} \times \cup_{i=1}^k H_+(x_i))$ is empty too, whence the conclusion. $\square$

The first part of the following proof is taken from the one of Theorem 5.1 in de Valk and Segers (2018) with minor changes. We give all the details.

Proof of Theorem 2.38. (a) Let a sequence $t_n > 0, n \geq 1$ satisfying $t_n \to \infty$. We consider the Borel measures (on $\mathbb{R}^d \setminus \{0\}$) $\mu_n = t_n \operatorname{res} P(b(t_n)\,\cdot\,)$ and $\nu_n = t_n \operatorname{res} Q(b(t_n)\cdot)$, both with mass t_n and M_0-converging respectively to μ and ν. In view of Lemma 5.3 in de Valk and Segers (2018), the function $\psi_n = (b(t_n)b(t_n))^{-1}\psi(b(t_n)\cdot)$ is closed and convex, and the graph of $\partial\psi_n$ contains the support of the measure $\gamma_n = t_n \operatorname{res}\pi(B(t_n)\cdot)$ that zero-couples μ_n and ν_n. By Theorem 2.24, there exists an infinite $\mathrm{N} \subset \mathbb{N}$ such that γ_n M_0-converges to some $\gamma \in \mathrm{M}_0(\mathbb{R}^d \times \mathbb{R}^d)$ as $n \to \infty$ in N and $\partial\psi_n = (b(t_n))^{-1}\partial\psi(b(t_n)\,\cdot\,)$ $\mathcal{F}$-converges to some T in $\mathcal{F}_{\mathrm{mcm}}(\mathbb{R}^d \times \mathbb{R}^d)$ such that $\operatorname{spt}\gamma \subset T$. Since $T \in \mathcal{F}_{\mathrm{mcm}}(\mathbb{R}^d \times \mathbb{R}^d)$, Rockafellar and Wets' Theorem 1.37 gives the existence of some closed convex function $\bar{\psi}$ such that $T = \partial\bar{\psi}$.

Relabelling t_n, we now suppose that the convergences take place along the whole sequence. Since $b(t)$ is invertible for every $t > 0$ and $b(\lambda t)^{-1}b(t) \to \lambda^{-1/\alpha}$ as $t \to \infty$, the homogeneity follows from the theory of operator regular variation Proposition 6.1.2 in Meerschaert and Scheffler (2001), as M_0-convergence is equivalent to vague convergence in compactified and punctured Euclidean space, at least for measures that do not charge the artificially added points at infinity (Proposition 4.4 in Lindskog et al. (2014)). The homogeneity of γ implies that of it support,

$$\lambda^{1/\alpha} \operatorname{spt} \gamma = \operatorname{spt} \gamma.$$

(b) Using Lemma 2.29 and Corollary 2.23, under the hypothesis of (b), γ_{t_n} converges to some $\gamma \in \mathrm{M}_{0,cm}(\mathbb{R}^d \times \mathbb{R}^d)$ and $\partial\psi_{t_n} \llcorner V$ $\mathcal{F}$-converges to $\partial\bar{\psi} \llcorner V$ in $\mathcal{F}(V \times \mathbb{R}^d)$ along the whole sequence, where V denotes $\operatorname{int} \operatorname{spt} \mu$ and $\nabla\bar{\psi}$ is uniquely defined μ-almost everywhere. We also have (see Remark 2.28) $\gamma = [\mathrm{Id} \otimes \nabla\bar{\psi}]_{\#}\mu$.

Let $T = \operatorname{spt} \gamma \in \mathcal{F}_{\mathrm{cm}}(\mathbb{R}^d \times \mathbb{R}^d)$, then for $x \in \mathbb{R}^d$ we have $y \in T(x)$ if and only if $\lambda^{1/\alpha} y \in T(\lambda^{1/\alpha} x)$, whence

$$T(\lambda x) = \lambda T(x), \qquad \lambda > 0, x \in \mathbb{R}^d$$

where possibly both sides can be equal to the empty set.

Let $\lambda > 0$. Recall that for μ-almost all x in V, $\nabla\bar{\psi}(\lambda x)$ and $\nabla\bar{\psi}(x)$ are defined and uniquely determined. A simple computation yields $\mu\left((\operatorname{proj}_1 \operatorname{spt} \gamma)^c\right) = 0$ (we refer to the proof of Theorem 2.29 for details). As a consequence, for μ-almost all x, $T(x)$ is non-empty and by homogeneity $T(x)$ is non-empty too. Thus we have

$$\nabla\bar{\psi}(\lambda x) = T(\lambda x) = \lambda T(x) = \lambda \nabla\bar{\psi}(x), \qquad \forall \lambda > 0,$$

μ-almost everywhere as announced.

(c) Assume $\operatorname{spt} \mu = \mathbb{R}^d$. Thanks to (a) we know that any limit point of $t_n \operatorname{res} \pi(b(t_n) \cdot), n \geq 1$ has homogenous support, then following Remark 2.37 it is clear that $\Gamma_{0,cm}(\mu, \nu) = \Gamma_{h,cm}(\mu, \nu)$ which is therefore a singleton $\{\gamma\}$. We can thus apply Theorem 2.23 with $V = \operatorname{int} \operatorname{spt} \mu = \mathbb{R}^d$ to get the convergence in (a) along the whole sequences. $\square$

B Proofs of Section 4

Proofs of Subsection 4.1

Proof of Lemma 4.7. Applying (c) from Theorem 2.38 we get that $\Gamma_{0,cm}(\mu,\nu)$ is a singleton whose single element γ can be written $\gamma = [\mathrm{Id} \times \nabla\bar{\psi}]_{\#}\mu$ for every closed convex function $\bar{\psi}$ satisfying $\gamma \subset \partial\bar{\psi}$. The gradient $\nabla\bar{\psi}$ is the μ-almost everywhere unique cyclically monotone map such that $\nu = \operatorname{res}\nabla\bar{\psi}_{\#}\mu$. From Theorem 2.38 we also have that $\partial\bar{\psi}$ is homogenous.

By assumption μ vanishes on sets of Hausdorff dimension at most $d-1$, whence there exists a Borel set $A_{\bar{\psi}}$ of $\mathbb{R}^d \setminus \{0\}$ such that $\nabla\bar{\psi}$ is defined on $A_{\bar{\psi}}$ and $\mu(A^c) = 0$. From the proof of Lemma 2.29, we have $\mu\left((\mathrm{proj}_1 \operatorname{spt}\gamma)^c\right) = 0$. As a consequence $\mu(A^c) = 0$ for

$$A = [\mathrm{proj}_1 \operatorname{spt}\gamma] \cap A_{\bar{\psi}}.$$

Let us consider the sets

$$\begin{aligned}
Z &:= A \cap \left[\partial\bar{\psi}\right]^{-1}(\{0\}),\\
\chi &:= \{x \in A \;:\; \operatorname{spt}\nu \cap \operatorname{cl} H_+(x) = \{0\}\},\\
\xi &:= \{x \in A \;:\; \operatorname{spt}\nu \cap H_+(x) = \varnothing\}.
\end{aligned}$$

From Lemma 2.29 we have $\chi \subset Z \subset \xi$ and $\mu(\xi \setminus \chi) = 0$ which implies

$$\mu\left([\nabla\bar{\psi}]^{-1}(\{0\})\right) = \mu(\chi) = \mu(\xi).$$

Since $\partial\bar{\psi}$ is homogenous, and $\bar{\psi}$ is convex, $[\partial\bar{\psi}]^{-1}(\{0\})$ is a convex cone. Therefore there is a connected subset S of $\mathbb{S}^{d-1}$ such that the restriction of μ to $[\partial\bar{\psi}]^{-1}(\{0\})$, when it is nonzero, has polar decomposition

$$\mathrm{POLAR}\,\mu_{|[\partial\bar{\psi}]^{-1}(\{0\})} = \nu_\alpha \otimes \frac{\mathrm{Unif}(S)}{\mathrm{Unif}(\mathbb{S}^{d-1})(S)}$$

where $\mathrm{Unif}(\mathbb{S}^{d-1})(S)$ must be read as the measure $\mathrm{Unif}(\mathbb{S}^{d-1})$ applied to the set S.

Assume $m := \mu(\chi) > 0$, we claim that $\nu(\mathbb{R}^d \setminus \mathbb{C}^{\tau}_{\nu}(q)) = 1 - q$ cannot hold for every q lying in $(0,1)$. Since $\mu([\nabla\bar{\psi}]^{-1}(\{0\})) > 0$, $\mu_{|[\partial\bar{\psi}]^{-1}(\{0\})}$ is non-zero, whence we necessarily have $p := \mathrm{Unif}(\mathbb{S}^{d-1})(S) > 0$. As a consequence, for every $r > 0$ there is some x living in $\mathrm{B}_{0,r} \cap A$ such that $\nabla\bar{\psi}(x) = 0$, whence 0 lies in $\nabla\bar{\psi}(\mathrm{B}_{0,r})$. It is clear that

$$[\nabla\bar{\psi}]^{-1}(\{0\}) \cup \mathrm{B}_{0,r} \subset [\nabla\bar{\psi}]^{-1}\nabla\bar{\psi}(\mathrm{B}_{0,r})$$

for each $r > 0$, and this inclusion implies

$$\nu(\nabla\bar{\psi}(\mathrm{B}_{0,r})^c) = \mu(([\nabla\bar{\psi}]^{-1}\nabla\bar{\psi}(\mathrm{B}_{0,r}))^c) \le \mu([\nabla\bar{\psi}]^{-1}(\{0\})^c \cap \mathrm{B}_{0,r}^c).$$

Using polar decomposition of μ as $\nu_\alpha \otimes \mathrm{Unif}(\mathbb{S}^{d-1})$, a simple computation yields

$$\mu([\nabla\bar{\psi}]^{-1}(\{0\})^c \cap \mathrm{B}_{0,r}^c) = r^{-\alpha}(1-p)$$

for every $r > 0$. Choosing $r = (1-q)^{-1/\alpha}$ we finally get, for every q in $(0,1)$,

$$\nu(\mathbb{R}^d \setminus \mathbb{C}_\nu^\tau(q)) \le (1-q)(1-p) < (1-q)$$

whence a contradiction. As a consequence, $\mu(\chi) = 0$ as announced. □

Proof of Theorem 4.10. The proof relies on arguments similar to those used in Section 1.6. Note that the assumptions of Theorem 2.38(c) are satisifed. Let $\bar{\psi}$ be a closed convex function whose gradient $\nabla\bar{\psi}$ is the μ-almost everywhere unique cyclically monotone function such that $\nu = \operatorname{res}\nabla\bar{\psi}_{\#}\mu$.

The first assertion is immediate since for every q in $(0,1)$ it is clear that

$$\mathrm{B}_{0,(1-q)^{-1/\alpha}} \subset [\nabla\bar{\psi}]^{-1}\left(\nabla\bar{\psi}\big(\mathrm{B}_{0,(1-q)^{-1/\alpha}}\big)\right) = [\nabla\bar{\psi}]^{-1}(\mathbb{C}^\tau(q))$$

which implies $\tilde{\nu}(\mathbb{R}^d \setminus \mathbb{C}^\tau(q)) \le \mu(\mathbb{R}^d \setminus \mathrm{B}_{0,(1-q)^{-1/\alpha}}) = 1-q$.

Let $\bar{\phi} := \bar{\psi}^*$ denote the Legendre transform of $\bar{\psi}$. Applying Theorem 1.33 to the proper closed convex function $\bar{\psi}$ yields $\partial\bar{\psi} = (\partial\bar{\psi}^*)^{-1}$ where both sides of the equality are seen as multivalued maps. Since both μ and ν vanishes on sets of Hausdorff dimension at most $d-1$, there exist Borel sets A,B of $\mathbb{R}^d$ such that $\nabla\bar{\psi}$,$\nabla\bar{\phi}$ are defined on A,B respectively and $\tilde{\mu}(A^c) = \tilde{\nu}(B^c) = 0$. We may choose A, B to be subsets of $\mathbb{R}^d \setminus \{0\}$, then $\mu(A^c) = \nu(B^c) = 0$ where the complements are taken in $\mathbb{R}^d \setminus \{0\}$. For every x in $A \cap [\nabla\bar{\psi}]^{-1}(B)$, we have $\nabla\bar{\phi} \circ \nabla\bar{\psi}(x) = x$ because $\partial\bar{\psi} = (\partial\bar{\phi})^{-1}$.

Here the proof differs from the one for probability measures since 0 lies in B^c. As a consequence, $[\nabla\bar{\psi}]^{-1}(B)^c = [\nabla\bar{\psi}]^{-1}(B^c)$ may contain a Borel sets of $\mathbb{R}^d \setminus \{0\}$ of infinite mass for μ. Using assumption (b), we can show that $\tilde{\mu}([A \cap [\nabla\bar{\psi}]^{-1}(B)]^c) = 0$. Indeed, since both μ and ν are limit measures in the definition of regular variation, we have as consequences of Theorem 2.38 that $\partial\bar{\psi}$ is homogenous and $\nabla\bar{\psi}$ is associated with the unique cyclically monotone coupling γ lying in $\Gamma_{h,cm}(\mu,\nu) = \Gamma_{0,cm}(\mu,\nu)$. From the proof of Lemma 2.29, $\mu(\chi) = 0$ is equivalent to $\mu(A \cap [\partial\psi]^{-1}(\{0\})) = 0$ which implies $\tilde{\mu}([\nabla\bar{\psi}]^{-1}(\{0\})) = 0$. Therefore we can write successively

$$\begin{aligned}
\tilde{\mu}([A \cap [\nabla\bar{\psi}]^{-1}(B)]^c) &= \tilde{\mu}([\nabla\bar{\psi}]^{-1}(B^c)) \\
&= \tilde{\mu}([\nabla\bar{\psi}]^{-1}(B^c \setminus \{0\})) + \tilde{\mu}([\nabla\bar{\psi}]^{-1}(\{0\})) \\
&= \tilde{\mu}([\nabla\bar{\psi}]^{-1}(B^c \setminus \{0\})) \\
&= \nu(B^c \setminus \{0\}) = 0.
\end{aligned}$$

We just proved that the equality $\nabla\bar{\phi} \circ \nabla\bar{\psi} = \mathrm{Id}$ holds μ-almost everywhere. As a consequence, we finally have the wanted result

$$\nu(\mathbb{R}^d \setminus \mathbb{C}^\tau_\nu(q)) = \mu(\mathbb{R}^d \setminus \mathrm{B}_{0,(1-q)^{-1/\alpha}}) = 1 - q$$

for every q lying in $(0, 1)$. $\square$

Proof of Theorem 4.11. Applying (c) from Theorem 2.38 yields $\Gamma_{0,cm}(\mu, \nu) = \{\gamma\}$ where γ can be written $\gamma = [\mathrm{Id} \times \nabla\bar{\psi}]_\# \mu$ for every closed convex function $\bar{\psi}$ satisfying $\gamma \subset \partial\bar{\psi}$. The gradient $\nabla\bar{\psi}$ is the μ-almost everywhere unique cyclically monotone map such that $\nu = \mathrm{res}\, \nabla\bar{\psi}_\# \mu$. From Theorem 2.38 we also have that $\partial\bar{\psi}$ is homogenous.

Since $\partial\bar{\psi}$ is homogenous, and $\bar{\psi}$ is convex, $[\partial\bar{\psi}]^{-1}(\{0\})$ is a convex cone. Therefore there is a connected subset S of $\mathbb{S}^{d-1}$ such that the restriction of μ to $[\partial\bar{\psi}]^{-1}(\{0\})$, when it is nonzero, has polar decomposition

$$\mathrm{POLAR}\, \mu_{|[\partial\bar{\psi}]^{-1}(\{0\})} = \nu_\alpha \times \frac{\mathrm{Unif}(S)}{p}$$

where $\mathrm{Unif}(S)$ is the uniform measure on the set S and $p = \mathrm{Unif}(\mathbb{S}^{d-1})(S)$ must be read as the measure $\mathrm{Unif}(\mathbb{S}^{d-1})$ applied to the set S. It is clear that $S = \mathbb{S}^{d-1} \cap [\partial\bar{\psi}]^{-1}(\{0\})$ since $\mathrm{spt}\, \mu = \mathbb{R}^d$ implies $\mathrm{spt}\, \mu_{|[\partial\bar{\psi}]^{-1}(\{0\})} = [\partial\bar{\psi}]^{-1}(\{0\})$. Note that the case $p = 0$ is treated in Theorem 4.10, so from now on we assume $p > 0$.

Let μ' denote the measure such that its restriction $\mu'_{|[\partial\bar{\psi}]^{-1}(\{0\})}$ to $[\partial\bar{\psi}]^{-1}(\{0\})$ is the null measure and its restriction $\mu'_{|[\partial\bar{\psi}]^{-1}(\{0\})^c}$ to $[\partial\bar{\psi}]^{-1}(\{0\})^c$ matches $\mu_{|[\partial\bar{\psi}]^{-1}(\{0\})^c}$.

Let $\bar{\phi} := \bar{\psi}^*$ denote the Legendre transform of $\bar{\psi}$. Applying Theorem 1.33 to the proper closed convex function $\bar{\psi}$ yields $\partial\bar{\psi} = (\partial\bar{\psi}^*)^{-1}$ where both sides of the equality are seen as multivalued maps. Since both μ and ν vanish on sets of Hausdorff dimension at most $d - 1$, there exist Borel sets A, B of $\mathbb{R}^d$ such that $\nabla\bar{\psi}$, $\nabla\bar{\phi}$ are defined on A, B respectively and $\tilde{\mu}(A^c) = \tilde{\nu}(B^c) = 0$. We may choose A,B to be subsets of $\mathbb{R}^d \setminus \{0\}$, then $\mu(A^c) = \nu(B^c) = 0$ where the complements are taken in $\mathbb{R}^d \setminus \{0\}$. For every x in $A \cap [\nabla\bar{\psi}]^{-1}(B)$, we have $\nabla\bar{\phi} \circ \nabla\bar{\psi}(x) = x$ because $\partial\bar{\psi} = (\partial\bar{\phi})^{-1}$. A simple computation yields successively

$$\begin{aligned}
\mu'([A \cap [\nabla\bar{\psi}]^{-1}(B)]^c) &= \mu'([\nabla\bar{\psi}]^{-1}(B^c)) \\
&= \mu'([\nabla\bar{\psi}]^{-1}(B^c \setminus \{0\})) + \mu'([\nabla\bar{\psi}]^{-1}(\{0\})) \\
&= \mu'([\nabla\bar{\psi}]^{-1}(B^c \setminus \{0\})) + 0 \\
&= \mu([\nabla\bar{\psi}]^{-1}(B^c \setminus \{0\})) \\
&= \nu(B^c \setminus \{0\}) = 0.
\end{aligned}$$

As a consequence $\nabla\bar{\phi} \circ \nabla\bar{\psi} = \mathrm{Id}$ holds μ'-almost everywhere. Moreover, since 0 lies in $\nabla\bar{\psi}(\mathrm{B}_{0,r})$ for any $r > 0$ we have $[\nabla\bar{\psi}]^{-1}\left(\mathbb{C}^\tau(q)\right)^c \subset [\nabla\bar{\psi}]^{-1}(\{0\})^c$, hence

$$\begin{aligned}\nu(\mathbb{C}^\tau(q)^c) &= \mu([\nabla\bar{\psi}]^{-1}\nabla\bar{\psi}(\mathrm{B}_{0,(1-q)^{-1/\alpha}})^c) \\ &= \mu'([\nabla\bar{\psi}]^{-1}\nabla\bar{\psi}(\mathrm{B}_{0,(1-q)^{-1/\alpha}})^c).\end{aligned}$$

As a consequence we can write

$$\nu(\mathbb{C}^\tau(q)^c) = \mu'(\mathrm{B}_{0,(1-q)^{-1/\alpha}})^c)$$

and using polar decomposition we have

$$\nu(\mathbb{C}^\tau(q)^c) = (1-q)\mathrm{Unif}(\mathbb{S}^{d-1})(S^c)$$

where $\mathrm{Unif}(\mathbb{S}^{d-1})(S)$ must be read as the measure $\mathrm{Unif}(\mathbb{S}^{d-1})$ applied to the set S^c. This yields the desired identity

$$\nu(\mathbb{C}^\tau(q)^c) = (1-p)(1-q).$$

□

Proof of Theorem 4.15

In this subsection, we prove the M_0-convergence of the empirical measures introduced at the beginning of Subsection 4.2 and that will be used in Subsection 4.3 to build estimators of center-outward quantile regions and contours. For $\mu \in \mathrm{M}_0(\mathbb{R}^d)$ and $a > 0$, let $\mu(a\cdot)$ denote the measure defined by $B \mapsto m(\{ax : x \in B\})$.

Lemma B.1. *Let $(m_n)_{n\in\mathbb{N}} \subset \mathrm{M}_0(\mathbb{R}^d)$ be a sequence of random measures converging weakly to $m \in \mathrm{M}_0(\mathrm{S})$, and $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ a sequence of real values converging in probability to $a \in \mathbb{R} \setminus \{0\}$ in $\mathbb{R}$. Then we can state that*

$$m_n(a_n\cdot) \xrightarrow{w} m(a\cdot) \text{ in } \mathrm{M}_0(\mathbb{R}^d).$$

Proof. Let

$$\phi : \mathrm{M}_0(\mathbb{R}^d) \times \mathbb{R} \to \mathrm{M}_0(\mathbb{R}^d) \tag{7}$$
$$(m, a) \to m(a\cdot) \tag{8}$$

1. According to Slutsky's lemma, we have $(m_n, a_n) \xrightarrow{w} (m, a)$ in $\mathrm{M}_0(\mathbb{R}^d) \times \mathbb{R}$.

2. We then check that ϕ is continuous at (m, a). Let $(m_n)_{n\in\mathbb{N}}$ and $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ be (deterministic) sequences such that m_n M_0-converges to μ and $a_n \to a$ in $\mathbb{R}$ as $n \to \infty$.

   Let $f \in \mathrm{C}^+_{b,0}(\mathbb{R}^d)$, using triangular inequality we split the euclidian distance in two.

$$
\begin{aligned}
|m(a\cdot)(f) - m_n(a_n\cdot)(f)| \leq & \left| m\left(f\left(\frac{\cdot}{a}\right)\right) - m_n\left(f\left(\frac{\cdot}{a}\right)\right)\right| && (9)\\
& + \left| m_n\left(f\left(\frac{\cdot}{a}\right)\right) - m_n\left(f\left(\frac{\cdot}{a_n}\right)\right)\right| && (10)
\end{aligned}
$$

   It is immediate that $f(\cdot/a) \in \mathrm{C}^+_{b,0}$ so the first term goes to zero as n goes to ∞ by definition of the M_0-convergence. For the second term we use that since $(m_n)_{n\in\mathbb{N}}$ is relatively compact in $\mathrm{M}_0(\mathbb{R}^d)$, according to Theorem 2.7 in Hult and Lindskog (2006), we have a sequence $(r_i)_{n\in\mathbb{N}}$ satisfying $r_i \searrow 0$ such that for each i, for every positive ϵ there exists a compact set $C_i \subset \mathrm{S} \setminus \mathrm{B}_{0,r_i}$ such that $\sup_{n\in\mathbb{N}} m_n(\mathrm{S} \setminus (C_i \cup \mathrm{B}_{0,r_i})) \leq \epsilon$. We choose i large enough to have $dom(f) \subset \mathrm{S} \setminus \mathrm{B}_{0,r_i}$. Since f is uniformly continuous on C_i, we can find N such that $\sup_{x\in C_i} |f(x/a_n) - x/a| < \epsilon$ for every $n \geq \mathrm{N}$. We then split in two the integral.

$$
\begin{aligned}
\left| m_n\left(f\left(\frac{\cdot}{a}\right)\right) - m_n\left(f\left(\frac{\cdot}{a_n}\right)\right)\right| \leq & \left| \int_{C_i \setminus \mathrm{B}_{0,r_i}} f\left(\frac{x}{a_n}\right) - f\left(\frac{x}{a}\right) m_n(dx)\right| && (11)\\
& + \left| \int_{\mathrm{S}\setminus(C_i \cup \mathrm{B}_{0,r_i})} f\left(\frac{x}{a_n}\right) - f\left(\frac{x}{a}\right) m_n(dx)\right| && (12)
\end{aligned}
$$

   To deal with $m_n(C_i)$, we define $f_k(x) = \max(1 - k\, d(x, C_i), 0)$ which is in $\mathrm{C}^+_{b,0}$ for k large enough. So, for such k, we have $m_n(C_i) \leq m_n(f_k)$ and, as $n \to \infty$, $m_n(f_k) \to \mu(f_k) < \infty$ and there exists $M \in \mathbb{R}^+$ such that $m_n(C_i) \leq M$ holds for every n and M depends only on i which in turn depends on f.

   So, for n large enough, we have

$$
|m(a\cdot)(f) - m_n(a_n\cdot)(f)| \leq \epsilon(1 + M + 2||f||_\infty).
$$

   Hence ϕ is continuous at point (μ, a).

3. Continuous mapping theorem finally concludes.

□

Theorem B.2. *Let X_n, $n \geq 0$ be an iid sequence from P, regularly varying with scaling function b and limit measure ν. Let $\overline{\nu}_{n,t} = \frac{t}{n} \sum_{i=1}^{n} \delta_{X_i/b(t)}$. If $t = t_n$ is such that $t_n \to \infty$ and $t_n/n \to 0$ as $n \to \infty$, then we have the convergence*

$$\overline{\nu}_{n,t} \xrightarrow{w} \nu \qquad \text{in } \mathrm{M}_0(\mathbb{R}^d).$$

Proof. Thanks to Theorem 1.12, it is sufficient to show the pointwise convergence of the Laplace transforms on $\mathrm{C}^+_{b,0}(\mathbb{R}^d)$.

Let $t_n \in \mathbb{N}, n \geq 0$ be a sequence such that $t_n \to \infty$ and $t_n/n \to 0$ as $n \to \infty$, $f \in \mathrm{C}^+_{b,0}(\mathbb{R}^d)$, and $r_f > 0$ be such that $f(\mathrm{B}_{s_0,r_f}) = \{0\}$. Since X_i, $i \geq 0$ are iid we have,

$$\begin{aligned}
\Psi_{\overline{\nu}_{n,t_n}}(f) &= \mathbb{E}\left[\exp\left(-\int f d\overline{\nu}_{n,t}\right)\right] &(13)\\
&= \mathbb{E}\left[\exp\left(-\frac{t_n}{n}\sum_{i=1}^{n} f\left(\frac{X_i}{b(t_n)}\right)\right)\right] &(14)\\
&= \mathbb{E}\left[\exp\left(-\frac{t_n}{n} f\left(\frac{X_1}{b(t_n)}\right)\right)\right]^n &(15)\\
&= \exp\left(n\,\log\left(1-\mathbb{E}\left[1-\exp\left(-\frac{t_n}{n} f\left(\frac{X_1}{b(t_n)}\right)\right)\right]\right)\right) &(16)
\end{aligned}$$

We then study the asymptotic behavior of the term in the $\log(1-\cdot)$.

$$\begin{aligned}
\epsilon_{n,t_n} &= \mathbb{E}\big[1-\exp(-\frac{t_n}{n} f(\frac{X_1}{b(t_n)}))\big] &(17)\\
&= \mathbb{E}\big[\left(1-\exp(-\frac{t_n}{n} f(\frac{X_1}{b(t_n)}))\right) \mathbb{1}_{X_1 \in b(t_n)(\mathrm{S}\setminus \mathrm{B}_{s_0,r_f})}\big] &(18)
\end{aligned}$$

Since f is bounded we may use a Taylor expansion of the exponential function, uniformly over x, as $n \to \infty$,

$$\sup_{x\in\mathrm{S}} \left|1 - e^{-\frac{t_n}{n}f(x)} - \frac{t_n}{n} f(x)\right| = o\left(\frac{t_n}{n}\right).$$

We get

$$n\epsilon_{n,t_n} = n\mathbb{E}\left[\left(\frac{t_n}{n}f(\frac{X_1}{b(t_n)}) + o\left(\frac{t_n}{n}\right)\right)\mathbb{1}_{X_1\in b(t_n)(\mathrm{S}\setminus \mathrm{B}_{s_0,r_f})}\right] \tag{19}$$

$$= t_n\mathbb{E}\left[f(\frac{X_1}{b(t_n)}) + o(1)\mathbb{1}_{X_1\in b(t_n)(\mathrm{S}\setminus \mathrm{B}_{s_0,r_f})}\right] \tag{20}$$

Since $t\,\mathsf{P}\left(X_1/b(t)\in\cdot\right)\xrightarrow{0}\nu(\cdot)$ as $t\to\infty$ we have, as $n\to\infty$: $t_n\mathbb{E}\left[f\left(X_1/b(t_n)\right)\right]\to\nu(f)$ and, since $\mathrm{S}\setminus\mathrm{B}_{s_0,r_f}$ is closed and bounded away from the origin, we can define f_k as in the proof of lemma 2.1 to upperbound $t_n\mathbb{E}\big[o(1)\mathbb{1}_{X_1\in b(t_n)(\mathrm{S}\setminus\mathrm{B}_{s_0,r_f})}\big]$ uniformly in n. Whence we have $n\epsilon_{n,t_n}=\nu(f)+o(1)$ as $n\to\infty$.
Finally, as $n\to\infty$, we have $\forall f\in \mathrm{C}^+_{b,0}(\mathbb{R}^d)$, and as $n\to\infty$

$$\Psi_{\overline{\nu}_{n,t_n}}(f) = \exp\left(n\,\log\left(1-\frac{n\epsilon_{n,t_n}}{n}\right)\right)\to e^{-\nu(f)} = \Psi_\nu(f).$$

□

Corollary B.3. *Let X_n, $n\geq 0$ be an iid sequence from P, regularly varying with scaling function b and limit measure ν. We denote $\hat{b}_n(t)$ the empirical scaling function defined by $\hat{b}_n(t)=\hat{F}_n^{\leftarrow}(1-1/t)$ where $\hat{F}_n$ is the usual empirical distribution function of $\{|X_i|\}_{i=1}^n$. Let $\hat{\nu}_{n,t}=\frac{t}{n}\sum_{i=1}^n\delta_{X_i/\hat{b}_n(t)}$. If $t=t_n$ is such that $t_n\to\infty$ and $t_n/n\to 0$ as $n\to\infty$, then we have the convergence*

$$\hat{\nu}_{n,t}\xrightarrow{w}\nu \qquad \text{in } \mathrm{M}_0(\mathbb{R}^d).$$

Proof. It is a direct application of Lemma B.1 and Theorem B.2. Indeed, under the hypothesis of the corollary, we have that $\overline{\nu}_{n,t}\xrightarrow{w}\nu$ in $\mathrm{M}_0(\mathbb{R}^d)$. Moreover, the previous convergence implies $\hat{b}_n(t)/b(t)\xrightarrow{\mathsf{P}}1$. Indeed, taking arguments from Theorem 4.2 in Resnick (2007), for each $\epsilon>0$ we have

$$\mathsf{P}\left(\left|\hat{b}_n(t)/b(t_n)-1\right|>\epsilon\right) = \mathsf{P}\left(\hat{b}_n(t_n)>(1+\epsilon)b(t_n)\right)+\mathsf{P}\left(\hat{b}_n(t_n)<(1-\epsilon)b(t_n)\right) \tag{21}$$

Thanks to the usual properties of the generalized inverse function, we have

$$\left\{\hat{b}_n(t_n)>(1+\epsilon)b(t_n)\right\} = \left\{1-1/t_n>\hat{F}_n((1+\epsilon)b(t_n))\right\}$$

and

$$\left\{\hat{b}_n(t_n)<(1-\epsilon)b(t_n)\right\}\subset\left\{1-1/t_n\leq\hat{F}_n((1-\epsilon)b(t_n))\right\}.$$

We remark that $1 - \hat{F}_n(s\, b(t_n)) = t_n^{-1}\overline{\nu}_{n,t_n}(\{x : |x| > s\})$ to get

$$\mathsf{P}\left(\left|\hat{b}_n(t)/b(t_n) - 1\right| > \epsilon\right) \leq \mathsf{P}\left(\overline{\nu}_{n,t_n}(\{x : |x| > 1 + \epsilon\}) > 1\right) \tag{22}$$
$$+ \mathsf{P}\left(\overline{\nu}_{n,t_n}(\{x : |x| > 1 - \epsilon\}) \leq 1\right) \tag{23}$$

Since $\forall s \in \mathbb{R}^+, \nu(\{x : |x| = s\}) = 0$, the Portmanteau theorem in $\mathrm{M}_0(\mathbb{R}^d)$ gives that $\mu \in \mathrm{M}_0(\mathbb{R}^d) \mapsto \mu(\{x : |x| > s\}) \in \mathbb{R} \setminus \{0\}$ is continuous at point ν, so by continuous mapping we have, as $n \to \infty$, $\overline{\nu}_{n,t_n}(\{x : |x| > s\}) \to \nu(\{x : |x| > s\}) = s^{-a}$ in distribution (thanks to Corollary B.2). Whence $\hat{b}_n(t)/\hat{b}(t) \xrightarrow{\mathbb{P}} 1$.

Since $\hat{\nu}_{n,t} = \phi\big(\overline{\nu}_{n,t}, \hat{b}_n(t)/\hat{b}(t)\big)$ where ϕ is the function defined in the proof of Lemma B.1, the same lemma concludes. □

Corollary B.4. *Let $X_n, n \geq 0$ be an iid sequence from P, regularly varying with scaling function b and limit measure ν, and $U_n, n \geq 0$ an iid sequence from* $\mathrm{Unif}(\mathbb{S}^{d-1})$*. We suppose that $X_n, n \geq 0$ is independent of $U_n, n \geq 0$. We denote $\hat{b}_n(t)$ the empirical scaling function defined by $\hat{b}_n(t) = \hat{F}_n^{\leftarrow}(1 - 1/t)$ where $\hat{F}_n$ is the usual empirical distribution function of $\{|X_i|\}_{i=1}^n$. Let $\hat{\mu}_{n,t} = \frac{t}{n}\sum_{i=1}^n \delta_{|X_i|U_i/\hat{b}_n(t)}$. If $t = t_n$ is such that $t_n \to \infty$ and $t_n/n \to 0$ then $\hat{\mu}_{n,t_n} \xrightarrow{w} \mu$ in $\mathrm{M}_0(\mathbb{R}^d)$ as $n \to \infty$ where μ is defined as $\mu = |\cdot|_{\#}\nu \times \mathrm{Unif}_{\mathbb{S}^{d-1}}$.*

Proof. It is a particular case of the previous corollary, we just need to check that $|X_i|U_i$ is regularly varying with the right scaling function b and limit measure. Since $|X_iU_i| = |X_i|$ it is immediate that $|X_i|U_i$ has scaling function b. To check the limit measure, let $f \in \mathrm{C}^+_{b,0}(\mathrm{S})$, and $r_f > 0$ be such that $f(\mathrm{B}_{s_0,r_f}) = \{0\}$.

Since f is bounded, we can apply Fubini's theorem, so

$$\mathbb{E}\left[f\left(\frac{|X_1|U_1}{b_n(t)}\right)\right] = \iint f\left(\frac{|x|u}{b_n(t)}\right) \mathrm{Unif}_{\mathbb{S}^{d-1}}(du)\, P_{\frac{|X_1|U_1}{b_n(t)}}(dx) \tag{24}$$

We consider $\varphi : (x, u) \in (\mathrm{S} \setminus s_0) \times \mathbb{S}^{d-1} \longmapsto f(xu) \in \mathbb{R}^+$ and $F : x \in \mathrm{S} \setminus s_0 \longmapsto \int \varphi(x, u)\mathrm{Unif}_{\mathbb{S}^{d-1}}(du) \in \mathbb{R}^+$. We claim $F \in \mathrm{C}^+_{b,0}(\mathrm{S})$.
Indeed, we have

1. $\forall x \in (\mathrm{S} \setminus s_0), u \mapsto \varphi(x, u)$ is $\mathrm{Unif}_{\mathbb{S}}$-integrable
2. $\forall u, x_0, \lim_{x \to x_0} \varphi(x, u) = \varphi(x_o, u)$
3. $|\varphi(x, u)| < |f|_\infty < \infty$ and $\mathrm{Unif}_{\mathbb{S}}(\mathbb{S}) < \infty$

So F is continuous, nonnegative, and bounded thanks to the third point. Since $|U_1| = 1$, $\forall u \in \mathbb{S}, \varphi(\mathrm{B}_{s_0,r_f}, u) = \{0\}$ and $F(\mathrm{B}_{s_0,r_f}) = \{0\}$. So $F \in \mathrm{C}^+_{b,0}(\mathrm{S})$ and by definition of the convergence in $\mathrm{M}_0(\mathrm{S})$, we have, since $tP(x : \frac{x}{b(t)} \in \cdot) \xrightarrow{0} \nu(\cdot)$ as $t \to \infty$,

$$t\mathbb{E}\left[f\left(\frac{|X_1|U_1}{b(t)}\right)\right] \to \int f\left(|x|u\right) \mathrm{Unif}_{\mathbb{S}^{d-1}}(du)\, \nu(dx).$$

Therefore $|X_1|U_1$ is regularly varying with scaling function b and limit measure $|\cdot|_{\#}\nu \times \mathrm{Unif}_{\mathbb{S}^{d-1}}$

Remark B.5. *The proof can be simplified a lot with the polar decomposition introduced for Theorem 1.15, see Segers et al. (2017) Proposition 3.1 or simply (we are in $\mathbb{R}^d$) Resnick (2007) Theorem 6.1. We didn't do it for lack of time.*

□

Let $\mathcal{S}_n$ denote the family of permutations of $\{1, .., n\}$. The following corollary gives the convergence of the empirical zero-coupling.

Corollary B.6. *Let $X_n, n \geq 0$ be an iid sequence from P, regularly varying with scaling function b and limit measure ν, and $U_n, n \geq 0$ an iid sequence from* $\mathrm{Unif}(\mathbb{S}^{d-1})$. *We suppose that $X_n, n \geq 0$ is independent of $U_n, n \geq 0$. We denote $\hat{b}_n(t)$ the empirical scaling function defined by $\hat{b}_n(t) = \hat{F}_n^{\leftarrow}(1 - 1/t)$ where $\hat{F}_n$ is the usual empirical distribution function. Let $\hat{\mu}_{n,t}$ and $\hat{\nu}_{n,t}$ be defined as above. We define an empirical zero-coupling measure as*

$$\hat{\gamma}_{n,t} = \frac{t}{n} \sum_{i=1}^{n} \delta \left\{ \frac{|X_i|U_i}{\hat{b}_n(t)}, \frac{X_{\hat{\sigma}_n(i)}}{\hat{b}_n(t)} \right\}.$$

where $\hat{\sigma}_n \in \mathcal{S}_n$ satisfies

$$\hat{\sigma}_n \in \operatorname*{arg\,min}_{s \in \mathcal{S}_n} \sum_{i=1}^{n} \left| \frac{X_{\hat{\sigma}_i}}{\hat{b}_n(t)} - \frac{|X_i|U_i}{\hat{b}_n(t)} \right|^2.$$

If $t = t_n$ is such that $t_n \to \infty$ and $t_n/n \to 0$ then $\hat{\gamma}_{n,t_n} \xrightarrow{w} \gamma$ in $\mathrm{M}_0(\mathrm{S})$ as $n \to \infty$ where γ is the only cyclically monotone zero-coupling between μ and ν defined as in the previous results.

Proof. Since ν has infinite mass as a non degenerate limit measure of regular variation, it is clear that $\operatorname{spt} \mu$ is the whole set $\mathbb{R}^d$. Indeed, μ is defined as $|\cdot|_{\#}\nu \times \mathrm{Unif}_{\mathbb{S}^{d-1}}$ so it has a radiale measure whose support is $\mathbb{R}_+$ and a spectral measure putting mass on the whole unit sphere. We can therefore apply (c) from Theorem 2.38 to get the uniqueness of the cyclically monotone zero-coupling γ. To conclude, we apply Theorem 2.24 to the sequences of random measures $\hat{\mu}_{n,t_n}$ and $\hat{\nu}_{n,t_n}$ weakly converging in $\mathrm{M}_0(\mathbb{R}^d)$ to the deterministic measures μ and ν respectively. $\square$

Proofs of Subsection 4.3

Proof of Theorem 4.17. Recall that $\mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d)$ is Polish and compact so $\operatorname{spt} \hat{\gamma}_{n,t_n}$ is tight and relatively compact. Consider an infinite subset N of $\mathbb{N}$ such that $\operatorname{spt} \hat{\gamma}_{n,t_n}$ converges along N to some $T \in \mathcal{P}(\mathcal{F}(\mathbb{R}^d \times \mathbb{R}^d))$.

Since $\hat{\gamma}_{n,t_n}$ weakly M_0-converge to γ which is deterministic, applying Slutsky's lemma we get the weak convergence of $(\hat{\gamma}_{n,t_n}, \operatorname{spt} \hat{\gamma}_{n,t_n})$ toward (γ, T) along the subsequence N. Using Lemma 2.14, it is therefore clear that $\operatorname{spt} \gamma \subset T$ holds almost surely. Moreover, we have $\operatorname{spt} \hat{\gamma}_{n,t_n} \subset \partial \hat{\psi}_{n,t_n}$ almost surely for every n in N. Using the equivalence between convergence in the Fell topology and Painlevé-Kuratowksi for closed sets (we refer to Theorem 4.5 in Rockafellar and Wets (1998)), it is immediate that the set

$$\{(A, B) \in \mathcal{F}(\mathbb{R}^d)^2 \ : \ A \subset B\}$$

is closed in the product topology. Therefore, using the usual Portmanteau Theorem for weak convergence, we get $T \subset \partial \bar{\psi}$ almost surely. We have proved that almost surely

$$\operatorname{spt} \gamma \subset T \subset \partial \bar{\psi}.$$

To conclude, one just needs to notice that seen as multivalued maps $\mathbb{R}^d \rightrightarrows \mathbb{R}^d$, the three sets are equals when restricted to a well-chosen open set. Indeed, Theorem 1.29 yields a Borel set A of $\mathbb{R}^d$ such that $\nabla \psi$ is defined on A and $\mu(A^c) = 0$. Clearly, we can choose A to be a subset of $\operatorname{proj}_1 \operatorname{spt} \gamma$ since the later set is a μ-continuity set—we refer to Remark 2.19. Let $B = \operatorname{int} A$ which satisfies $B \subset \operatorname{proj}_1 \operatorname{spt} \gamma$ too.

For every $b \in B$, $\partial \psi(b)$ is single valued and $\operatorname{spt} \gamma(b)$ is non-empty since $B \subset \operatorname{proj}_1 \operatorname{spt} \gamma$. Necessarily we have $[\operatorname{spt} \gamma](b) = \partial \psi(b)$. As a consequence $\operatorname{spt} \gamma \llcorner B = T \llcorner B = \partial \psi \llcorner B$, hence $T \llcorner B$ is deterministic. Since any limit point of $\partial \hat{\psi}_{n,t_n} \llcorner B$ is equal to $\operatorname{spt} \gamma \llcorner B$, the convergence takes place along the whole sequence and the desired result is proved. $\square$

Proof of Corollary 4.19. Note that for every subset S of $\mathbb{R}^d \setminus \{0\}$ and closed set F in $\mathcal{F}\left((\mathbb{R}^d \setminus \{0\}) \times \mathbb{R}^d\right)$ we have $\theta(F \llcorner S) = \theta(F) \llcorner S$. As a consequence, by continuous mapping—we refer to Theorem 5.27 in Kallenberg (2021)—, we get from Theorem 4.17 the convergence

$$\theta(\operatorname{spt} \hat{\gamma}_{n,t_n}) \llcorner B \xrightarrow{w} \theta(\partial \bar{\psi}) \llcorner B \ \text{ in } \mathcal{F}(B \times \mathbb{R}^d).$$

where B is the open set introduced in Theorem 4.17. To prove the last identity, we use the homogeneity of $\nabla\bar{\psi}$ to write

$$\theta(\partial\bar{\psi})\llcorner B = \left\{(x, \nabla\bar{\psi}(x/|x|)) \,:\, x \in B\right\} = \left\{(x, \nabla\bar{\psi}(x)) \,:\, x \in B \cap \{|x| = 1\}\right\}.$$

$\square$

Proof of Lemma 4.22. From the sequence $B_k, k \geq 1$ we build an increasing sequence Q_k satisfying $\cup_{k\geq 1} Q_k = Q$ where $Q = \operatorname{rge} F\llcorner(B \cap R_{\epsilon,\delta})$. One may set $Q_k = \operatorname{rge} F\llcorner(B_k \cap R_{\epsilon,\delta})$. Since $B_k \subset B_{k+1}$ for every $k \geq 1$ and $\cup_{k\geq 1} B_k = B$, it is clear that $Q_k \subset Q_{k+1}$ for every $k \geq 1$ and $\cup_{k\geq 1} Q_k = Q$.

To begin with, we show for every $k \geq 1$ the inclusion

$$\operatorname{cl} Q_k \subset \liminf_{n\to\infty} \operatorname{cl}\operatorname{rge} F_n\llcorner(B_k \cap R_{\epsilon,\delta}) \subset \limsup_{n\to\infty} \operatorname{cl}\operatorname{rge} F_n\llcorner(B_k \cap R_{\epsilon,\delta}) \subset Q_{k+1}.$$

The first inclusion is trivial since from Lemma D.10 in Segers (2022), $\operatorname{cl}\operatorname{rge}$ is lower semi-continuous from $\mathcal{F}\left((B_k \cap R_{\epsilon,\delta}) \times \mathbb{R}^d\right)$ to $\mathcal{F}(\mathbb{R}^d)$. Indeed, by lower semi-continuity we have

$$\operatorname{cl} Q_k = \operatorname{cl}\operatorname{rge} F_n\llcorner(B_k \cap R_{\epsilon,\delta}) \subset \liminf_{n\to\infty} \operatorname{cl}\operatorname{rge} F_n\llcorner(B_k \cap R_{\epsilon,\delta}).$$

For the last inclusion, we consider y lying in $\limsup_{n\to\infty} \operatorname{cl}\operatorname{rge} F_n\llcorner(B_k \cap R_{\epsilon,\delta})$. By definition, there is a sequence $y_n \in \operatorname{cl}\operatorname{rge} F_n\llcorner(B_k \cap R_{\epsilon,\delta}), n \geq 1$ and an infinite subset N of $\mathbb{N}$ such that $y_n \to y$ as $n \to \infty$ in N. It is easy to see that we can choose y_n in $\operatorname{rge} F_n\llcorner(B_k \cap R_{\epsilon,\delta})$ instead of its closure. Since rge is the projection on the last d coordinates, we have a sequence $x_n \in B_k \cap R_{\epsilon,\delta}, n \geq 1$ such that (x_n, y_n) lies in $F_n \cap \left((B_k \cap R_{\epsilon,\delta}) \times \mathbb{R}^d\right)$. By assumption, $S_{k,\epsilon,\delta} := \operatorname{cl}(B_k \cap R_{\epsilon,\delta})$ is closed bounded subset of $\mathbb{R}^d$, hence it is compact. Moreover, by assumptions (a),(b) we have $S_{k,\epsilon,\delta} \subset \operatorname{cl} B_k \cap \operatorname{cl} R_{\epsilon,\delta} \subset B_{k+1} \cap R_{\epsilon+\tilde{\epsilon},\delta+\tilde{\epsilon}}$ for every $\tilde{\epsilon}, \tilde{\delta} > 0$. As a consequence, there is an infinite subset N' of N and some x lying in

$$\limsup_{n\to\infty} \operatorname{rge} F_n\llcorner(B_{k+1} \cap R_{\epsilon+\tilde{\epsilon},\delta+\tilde{\epsilon}}) = \operatorname{rge} F\llcorner(B_{k+1} \cap R_{\epsilon+\tilde{\epsilon},\delta+\tilde{\epsilon}})$$

such that $(x_n, y_n) \to (x, y)$ as $n \to \infty$ in N'. Here we used assumption (c). Whence y lies in $\operatorname{rge} F\llcorner(B_{k+1} \cap R_{\epsilon+\tilde{\epsilon},\delta+\tilde{\epsilon}})$ which is equal to Q_{k+1} by hypothesis (d).

Since $Q_k \subset Q_{k+1}$ for every $k \geq 1$ and $\cup_{k\geq 1} Q_k = Q$, one can easily prove, using the definition of lim inf and lim sup in term of limits of subsequences, that $\operatorname{cl} Q_k$ converges to $\operatorname{cl} Q$ in $\mathcal{F}(\mathbb{R}^d)$. The latter convergence yields the announced convergence

$$\lim_{k\to\infty} \liminf_{n\to\infty} \operatorname{cl}\operatorname{rge} F_n\llcorner(B_k \cap R_{\epsilon,\delta}) = \operatorname{cl} Q = \lim_{k\to\infty} \limsup_{n\to\infty} \operatorname{cl}\operatorname{rge} F_n\llcorner(B_k \cap R_{\epsilon,\delta}).$$

In view of the construction of the sequence $Q_k, k \geq 1$, we can also write

$$Q = \bigcup_{k\geq 1} \liminf_{n\to\infty} \operatorname{cl}\operatorname{rge} F_n \llcorner (B_k \cap R_{\epsilon,\delta}) = \bigcup_{k\geq 1} \limsup_{n\to\infty} \operatorname{cl}\operatorname{rge} F_n \llcorner (B_k \cap R_{\epsilon,\delta}).$$

$\square$

Proof of Theorem 4.23. Since both $R_{\epsilon,\delta}$ and B_k are an open sets, we have from the discussion at the beginning of the current subsubsection,

$$F_n \llcorner (R_{\epsilon,\delta} \cap B_k) \xrightarrow{w} F \llcorner (R_{\epsilon,\delta} \cap B_k) \quad \text{in } \mathcal{F}\left((R_{\epsilon,\delta} \cap B_k) \times \mathbb{R}^d\right)$$

as $n \to \infty$ for every $\epsilon, \delta > 0$.

An application of the Skorohod representation theorem (Theorem 5.31 in Kallenberg (2021)) yields the existence of some random elements $\xi_1, \xi_2, \ldots$ in $\mathcal{F}(B \times \mathbb{R}^d)$ satisfying

$$\xi_n \overset{\mathrm{d}}{=} F_n, \quad \xi_n \xrightarrow{\mathcal{F}} F \text{ a.s in } \mathcal{F}(B \times \mathbb{R}^d).$$

Assumptions (c),(d) in Lemma 4.22 holds almost surely, whence applying Lemma 4.22, we get the inclusion

$$\operatorname{cl} Q_k \subset \liminf_{n\to\infty} \operatorname{cl}\operatorname{rge} \xi_n \llcorner (B_k \cap R_{\epsilon,\delta}) \subset \limsup_{n\to\infty} \operatorname{cl}\operatorname{rge} \xi_n \llcorner (B_k \cap R_{\epsilon,\delta}) \subset Q_{k+1}.$$

for every $k \geq 1$ and

$$\mathcal{C}^\tau(0) = \bigcup_{k\geq 1} \liminf_{n\to\infty} \operatorname{cl}\operatorname{rge} \xi_n \llcorner (B_k \cap R_{\epsilon,\delta}) = \bigcup_{k\geq 1} \limsup_{n\to\infty} \operatorname{cl}\operatorname{rge} \xi_n \llcorner (B_k \cap R_{\epsilon,\delta}).$$

Recall from Subsection 1.3 that the Fell space $\mathcal{F}(\mathbb{E})$ is compact and Polish for any open subset $\mathbb{E}$ of $\mathbb{R}^d$. Here we use the set $\mathbb{E} = \mathbb{R}^d$ which contains the sequence of sets $\operatorname{cl}\operatorname{rge}\left(F_n \llcorner (R_{\epsilon,\delta} \cap B_k)\right), n \geq 1$. Therefore there exists an infinite subset N of $\mathbb{N}$ such that $\operatorname{cl}\operatorname{rge} F_n \llcorner (R_{\epsilon,\delta} \cap B_k)$ weakly $\mathcal{F}$-converges to some T_k in $\mathcal{P}(\mathcal{F}(\mathbb{R}^d))$ as $n \to \infty$ in N. Then the discussion above yields

$$\operatorname{cl} Q_k \subset T_k \subset Q_{k+1}.$$

for every $k \geq 1$ and finally the desired equality

$$\mathcal{C}^\tau(0) \cap B = \bigcup_{k\geq 1} T_k.$$

$\square$

References

Alberti, G. and Ambrosio, L. (1999). A geometrical approach to monotone functions in $\mathbb{R}^n$. *Mathematische Zeitschrift*, 230:259–316.

Anderson, R. D. and Klee, V. L. (1952). Convex functions and upper semi-continuous collections. *Duke Math. J.*, 19:349–357.

Basrak, B. and Planinić, H. (2019). A note on vague convergence of measures. *Statistics & Probability Letters*, 153:180–186.

Beirlant, J., Buitendag, S., Del Barrio, E., Hallin, M., and Kamper, F. (2020). Center-outward quantiles and the measurement of multivariate risk. *Insurance: Mathematics and Economics*, 95:79–100.

Catalano, M., Lavenant, H., Lijoi, A., and Prünster, I. (2021). A wasserstein index of dependence for random measures. *arXiv preprint arXiv:2109.06646*.

Chernozhukov, V., Galichon, A., HALLIN, M., and HENRY, M. (2017). Monge–kantorovich depth, quantiles, ranks and signs. *The Annals of Statistics*, 45:223–256.

Clémençon, S., Jalalzai, H., Lhaut, S., Sabourin, A., and Segers, J. (2022). Concentration bounds for the empirical angular measure with statistical learning applications. *Bernoulli, to appear (arXiv:2104.03966).*

de Valk, C. and Segers, J. (2018). Tails of optimal transport plans for regularly varying probability measures. *arXiv preprint arXiv:1811.12061.*

Goix, N., Sabourin, A., and Clémençon, S. (2017). Sparse representation of multivariate extremes with applications to anomaly detection. *J. Multivariate Anal.*, 161:12–31.

Guillen, N., Mou, C., and Swiech, S. (2019). Coupling lévy measures and comparison principles for viscosity solutions. *Transactions of the American Mathematical Society*, 372(10):7327–7370.

Guyader, A. (2023). *Statistique.* https://perso.lpsm.paris/ aguyader/files/teaching/M1.

Hallin, M., Del Barrio, E., Cuesta-Albertos, J., and Matrán, C. (2021). Distribution and quantile functions, ranks and signs in dimension d: a measure transportation approach. *Annals of Statistics*, 49:1139–1165.

Hult, H. and Lindskog, F. (2006). Regular Variation for Measures on Metric Spaces. *Publ. Inst. Math.*, 80(94):121–140.

Kallenberg, O. (2017). *Random Measures, Theory and Applications*. Springer Cham.

Kallenberg, O. (2021). *Foundations of Modern Probability*. Springer Cham.

Knott, M. and Smith, C. (1984). On the optimal mapping of distributions. *Journal of Optimization Theory and Applications*, 43:39–49.

Lindskog, F., Resnick, S. I., and Roy, J. (2014). Regularly varying measures on metric spaces: Hidden regular variation and hidden jumps. *Probability Surveys*, 11:270–314.

McCann, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80(2):309–323.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative risk management*. Series in Finance. Princeton University Press.

Meerschaert, M. M. and Scheffler, H.-P. (2001). *Limit Distributions for Sums of Independent Random Vectors: Heavy Tails in Theory and Practice*. New York: Wiley.

Molchanov, I. (2017). *Theory of Random Sets*. Springer Cham.

Resnick, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer Science & Business Media.

Resnick, S. I. (2008). *Extreme Values, Regular Variation and Point Processes*. Springer Science & Business Media.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton: Princeton University Press.

Rockafellar, R. T. and Wets, R. J.-B. (1998). *Variational Analysis*. New York : Springer.

Segers, J. (2022). Graphical and uniform consistency of estimated optimal transport plans. *arXiv preprint arXiv:2208.02508*.

Segers, J. (2023). Optimal Transport and Multivariate Extremes. Oral presentation at the Joint Statistical Meeting in Toronto, Ontario.

Segers, J., Zhao, Y., and Meinguet, T. (2017). Polar decomposition of regularly varying time series in star-shaped metric spaces. *Extremes*, 20:539–566.