Faculté des Sciences

Topics in multivariate spatial quantiles

Thesis presented by Dimitri KONEN

with a view to obtaining the PhD Degree in Sciences ("Docteur en Sciences")

Academic year 2022-2023

Supervisor: Professor Davy PAINDAVEINE

Département de Mathématique, ECARES

Thesis jury:

Thomas VERDEBOUT (Université Libre de Bruxelles, Chair)
Yvik SWAN (Université Libre de Bruxelles, Secretary)
Catherine DEHON (Université Libre de Bruxelles)
Marc HALLIN (Université Libre de Bruxelles)
Stanislav NAGY (Charles University)
Davy PAINDAVEINE (Université Libre de Bruxelles)
Bruno PREMOSELLI (Université Libre de Bruxelles)
Germain VAN BEVER (Université de Namur)

FREEDOM TO RESEARCH

Topics in multivariate geometric quantiles

Dimitri Konen

Thèse présentée en vue de l'obtention

du titre de Docteur en Sciences de l'Université Libre de Bruxelles

Réalisée sous la direction de Prof. Davy Paindaveine

Octobre 2022

Foreword

This thesis was carried out from October 2020 onwards to October 2022, as part of a four-year PhD program in the Département de Mathématique and in the European Center for Advanced Research in Economics and Statistics (ECARES) at the Université Libre de Bruxelles (ULB, Belgium) under the supervision of Prof. Davy Paindaveine. It was supported financially by the Fonds de la Recherche Scientifique F.R.S.-FNRS through a Research Fellowship.

Remerciements

Trouver ma voie dans les mathématiques n'a pas été simple. De ma cinquième secondaire à la fin de mon bachelier, je me suis considéré comme analyste. Un penchant pour les probabilités s'est dessiné en BA3. Par un hasard des choses, j'ai atterri dans le master de statistiques, en me considérant alors toujours comme analyste à tendance probabiliste. Dans ces circonstances, difficile de décider quoi entreprendre après le master... Entamer une thèse de doctorat ? Sur quel sujet ? Travailler dans le privé ? Où donc ? Ces considérations m'ont troublé l'esprit pendant un moment. Jusqu'à ma rencontre avec Davy. Nous avons collaboré un peu lors de ma première année de master. Ce projet n'a finalement rien donné. Mais Davy a été séduit par mon travail. J'ai été conquis par Davy. Dès ce moment, il est devenu chaque jour un peu plus clair que, finalement, le sujet que je pourrais choisir pour la réalisation d'une thèse importerait peu. Si je pouvais travailler avec lui, alors go ! Collaborer avec Davy durant mon mémoire et ces deux années de thèse a été la plus grande source de bonheur et d'épanouissement intellectuel que j'ai pu avoir. Sa disponibilité, la pertinence de ses réflexions et ses encouragements en ont fait le superviseur de thèse rêvé. La complémentarité de nos approches en a fait un collaborateur idéal. C'est donc naturellement à Davy que vont mes premiers et plus sincères remerciements.

Je souhaite remercier ma maman qui, durant ces années, a été d'une présence et d'un soutien inconditionnels, toujours à l'écoute, jamais dans le jugement, chaleureuse et profondément bienveillante. Elle a toujours su m'épauler lorsque j'en avais besoin... et c'était bien nécessaire par moments ! C'est irremplaçable, une maman comme ça.

Je remercie Alexia du fond du coeur, pour son dynamisme et son enthousiasme à toute épreuve. Elle m'a redonné le goût de l'exploration mathématique. Sans elle, je ne me serais certainement pas lancé dans la recherche de postdocs. Je la remercie également pour l'aide inestimable qu'elle m'a apportée pour améliorer mes rédactions en anglais.

A mes amis - Gérard, Maximilien, Thibaut, Christine, Julie, Gaspard - avec qui j'ai partagé des moments si précieux au cours des dernières années. Ils ont illuminé mes journées (ainsi que mes soirées, du reste) et mes papilles !

Je remercie mes amis de guindaille - Guillaume, Cédric, Sylvain, Morgane, Kévin, Nathan, Flore - pour leur amitié si précieuse et tout ce que j'ai pu faire avec eux durant ces années de folie. Merci aussi à tous les copains et toutes les copines de la Guilde Améthyste qui, quelque part, constitue une seconde famille.

J'aimerais remercier mes collègues du NO9, qui ont rendu le couloir tellement plus vivant. Plus particulièrement, merci Yvik, Thomas, Vivien, Gaspard, Julien et Julie !

Je remercie enfin les sept personnes qui, en plus de Davy, ont accepté de composer le jury de cette thèse: Thomas Verdebout, Yvik Swan, Catherine Dehon, Marc Hallin, Stanislav Nagy, Bruno Premoselli et Germain Van Bever. Leurs commentaires et suggestions très pertinentes ont mené à une amélioration substantielle de ce manuscrit.

Contents

Chapter 1

Introduction

The aim of this thesis is to better understand and extend the concept of *geometric quantiles, ranks and depth*. Along with many other companion *depth* notions, it is concerned with finding relevant ways of doing geometry *with respect to a dataset*. The main question is the following: for any $d \geq 1$, letting $\mathcal{X} = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ be a dataset, how far is a given location $x \in \mathbb{R}^d$ from $\mathcal{X}$? As we will explain below, this question is strongly related to the cumulative distribution function (cdf) and the associated quantiles of a probability measure. In this introduction, we start by recalling elementary facts about the cdf and quantiles of probability measures over $\mathbb{R}$. We explain how these concepts allow for a geometric description of $\mathbb{R}$ that is suited to a given probability measure P. We then discuss some extensions of these notions in $\mathbb{R}^d$ when $d > 1$, before introducing our main subject of interest, namely *geometric quantiles, ranks, and depth*.

1.1 Univariate quantiles and ranks

For any probability measure P over $\mathbb{R}$, we define the *cumulative distribution function* $F_P : \mathbb{R} \to [0, 1]$ of P by letting

$$F_P(x) = P\big[(-\infty, x]\big]$$

for any $x \in \mathbb{R}$. Let us mention important properties of the cdf:

1. F_P is non-decreasing, with

$$\lim_{x \to -\infty} F_P(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} F_P(x) = 1;$$

2. F_P is right-continuous by continuity of probability measures from above ;

3. F_P is continuous at $x \in \mathbb{R}$ if and only if $P[\{x\}] = 0$;

4. F_P is injective if and only if $P\big[(a, b]\big] > 0$ for any $-\infty < a < b < \infty$. This readily follows from the fact that

$$P\big[(a, b]\big] = F_P(b) - F_P(a).$$

Provided P is non-atomic and satisfies the last non-vanishing property, F_P is a continuous bijection between $\mathbb{R}$ and $(0,1)$. In particular, F_P is invertible. The inverse $Q_P : (0,1) \to \mathbb{R}$ of F_P is called the *quantile map*. It is non-decreasing, with

$$\lim_{\tau \to 0} Q_P(\tau) = -\infty, \quad \text{and} \quad \lim_{\tau \to 1} Q_P(\tau) = +\infty.$$

For any $\tau \in (0,1)$, the number $Q_P(\tau)$ is called the *quantile of order* τ of P. Taking $\tau = 1/2$ gives the *median* of P; it separates the probability content in equal proportion on both sides.

When F_P is not invertible, we say that a point $x \in \mathbb{R}$ is a quantile of order $\tau \in (0,1)$ if it satisfies $P\big[(-\infty, x]\big] \geq \tau$ and $P\big[[x, \infty)\big] \geq 1 - \tau$. This can be rewritten

$$\tau \leq F_P(x) \leq \tau + P[\{x\}],$$

or

$$P[(-\infty, x)] \leq \tau \leq P[(-\infty, x]].$$

Since F_P is non-decreasing and right-continuous, and because the jump height at a point $x \in \mathbb{R}$ is at most equal to $P[\{x\}]$, quantiles of any order $\tau \in (0,1)$ always exist. They may, however, not be unique. Indeed, if P vanishes over some interval $[a,b]$, then F_P is constant over $[a,b]$. In particular, any point in the interval $[a,b]$ is a quantile of order $\tau = F_P(a) = F_P(b)$.

When the cdf is not invertible, it is sometimes convenient to enforce uniqueness of quantiles of arbitrary order $\tau \in (0,1)$ by letting

$$F_P^{-1}(\tau) := \inf\{z \in \mathbb{R} : F_P(z) \geq \tau\}. \tag{1.1.1}$$

The cdf and quantile maps play a vital role in statistics. They both characterize the underlying probability measure: if P_1 and P_2 are probability measures over $\mathbb{R}$, and if $F_{P_1} = F_{P_2}$ over $\mathbb{R}$, or $Q_{P_1} = Q_{P_2}$ if F_{P_1} and F_{P_2} are invertible, then $P_1 = P_2$. The cdf and quantile maps have two major uses. On the one hand, the quantile map allows one to establish confidence intervals adapted to probability measures, an important fact used in hypothesis testing. On the other hand, the cdf orders data points. These features have found many applications in hypothesis testing, outliers detection, and extreme value theory to cite only a few examples.

1.1.1 Ordering

Consider a random sample $X_1, \ldots, X_n$ drawn from P, and the associated empirical cdf F_n, defined by letting

$$F_n(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}[X_i \leq x]$$

for any $x \in \mathbb{R}$, where $\mathbb{I}[A]$ stands for the indicator function of the condition A. Given $j \in \{1, \ldots, n\}$, the number $\sigma(j) := nF_n(X_j)$ belongs to $\{1, \ldots, n\}$, and corresponds to the position of X_j among the $X_i's$ if they were ordered from smallest to largest. If P admits a density with respect to the Lebesgue measure, the event

$$\Big\{\exists\ i \neq j \text{ such that } X_i = X_j\Big\}$$

has P-probability 0. In this case, the $\sigma(i)$'s are distinct and, with P-probability 1, the map σ we constructed is a permutation of $\{1, \ldots, n\}$ such that

$$X_{\sigma^{-1}(1)} \leq \ldots \leq X_{\sigma^{-1}(n)}.$$

Therefore, the map F_n yields a bijection between the observations $\{X_1, \ldots, X_n\}$ and $\{1, \ldots, n\}$, hence an ordering of the observations $X_1, \ldots, X_n$. This makes it possible to assess the relative position of any given location $x \in \mathbb{R}$ with respect to the sample $\{X_1, \ldots, X_n\}$. In particular, the cdf yields a measure of outlyingness from $\{X_1, \ldots, X_n\}$. This construction is only made possible by the natural ordering of $\mathbb{R}$.

1.1.2 Tolerance regions

One can give tolerance regions for P in terms of F_P and Q_P. For instance, the interval

$$I_\tau = \Big[x, Q_P(F_P(x) + \tau)\Big]$$

has probability τ, for any $x \in \mathbb{R}$ and any $\tau \in (0, 1)$ such that $F_P(x) < 1 - \tau$. One can form nested tolerance intervals indexed by probability content by letting $x = x(\tau)$ be a non-increasing function of τ such that $x(\tau) < Q_P(1 - \tau)$ for any $\tau \in (0, 1)$, and

$$x(0) := \sup_{\tau \in (0,1)} x(\tau) < +\infty.$$

In this way, the intervals (I_τ) are nested and such that $P[I_\tau] = \tau$ for any $\tau \in (0, 1)$. Since the intervals (I_τ) decrease to $x(0)$ as $\tau \searrow 0$, the innermost point with respect to (I_τ) is $x(0)$. In terms of probability content, the innermost point should be the median, i.e. $x(0) = Q_P(\frac{1}{2})$. For instance, $x(\tau) = Q_P(\frac{1-\tau}{2})$ is an admissible choice. This leads to the familiy of nested intervals $(I_\tau)_{\tau \in (0,1)}$ given by

$$I_\tau = \Big[Q_P\Big(\frac{1 - \tau}{2}\Big), Q_P\Big(\frac{1 + \tau}{2}\Big)\Big] \tag{1.1.2}$$

for any $\tau \in (0, 1)$. These tolerance intervals display a natural symmetry: for any $\tau \in (0, 1)$, the median $x(0)$ belongs to I_τ, and $x(0)$ splits I_τ in two subintervals of equal probability. Of course, there are many other possible choices for $x(\tau)$.

The intervals I_τ constructed above give a better suited way of assessing "distances" with respect to a probability distribution. In this regard, the points $Q_P(\frac{1+\tau}{2})$ and $Q_P(\frac{1-\tau}{2})$ are considered "equidistant" from the median $x(0)$ with respect to P, while they may have considerably different Euclidean distances from $x(0)$. For instance, consider a χ^2 distribution with one degree of freedom. The median $Q_P(\frac{1}{2})$ is approximately 0.45494. For $\tau = 0.05$, we have

$$I_{0.05} \approx [0.40406, 0.51032],$$

with $|0.51032 - Q_P(\frac{1}{2})| \approx 0.05538$, and $|0.40406 - Q_P(\frac{1}{2})| \approx 0.05088$. For $\tau = 0.95$, we have

$$I_{0.95} \approx [0.00098, 5.02389],$$

with $|5.02389 - Q_P(\frac{1}{2})| \approx 4.56895$, and $|0.00098 - Q_P(\frac{1}{2})| \approx 0.45396$.

1.2 Multivariate extensions

Both for ordering observations and establishing tolerance intervals in $\mathbb{R}$, the underlying idea remains the same: assessing "farness" from a dataset or, more generally, from a probability distribution. In both cases, the ingredient that allows for the above constructions is the possibility of ordering points in $\mathbb{R}$. The question naturally arises of whether this is possible in spaces of larger dimensions. Indeed, $\mathbb{R}^d$ does not possess a natural ordering when $d > 1$.

It is possible to establish tolerance regions of probability measures in $\mathbb{R}^d$ by considering balls of increasing radii, centered at a fixed location. Even though these balls provide "confidence" regions in the sense of probability content, they lack any geometric meaning regarding the shape of the probability distribution. Indeed, balls do not necessarily reflect the symmetries or asymetries of the distribution. For this reason, statisticians have been trying to define notions of centrality that capture the geometry of datasets or, in general, probability measures. We call such concepts *depth functions*.

Let $d \in \mathbb{N}$ be fixed. For any probability measure P over $\mathbb{R}^d$, a depth function over $\mathbb{R}^d$ for P is a map $D_P : \mathbb{R}^d \to [0, 1]$. Small values of $D_P(x)$ are interpreted as points laying "far away" from the distribution, and large values correspond to points "close to" the distribution (they are *deep*). Obviously, not every map from $\mathbb{R}^d$ to $[0, 1]$ is admissible as a depth function, it should satisfy some minimal requirements:

1. D_P admits a global maximum, i.e. there exists $x_0 \in \mathbb{R}^d$ such that $D_P(x_0) \geq D_P(x)$ for any $x \in \mathbb{R}^d$;

2. $D_{P+v}(x + v) = D_P(x)$ for any $x \in \mathbb{R}^d$ and $v \in \mathbb{R}^d$, where $P + v$ stands for the probability measure defined by letting $(P + v)[A] = P[A - v]$ for any Borel subset $A \subset \mathbb{R}^d$;

3. $D_{UP}(Ux) = D_P(x)$ for any $x \in \mathbb{R}^d$ and $d \times d$ orthogonal matrix U, where UP stands for the probability measure defined by letting $(UP)[A] = P[U^{-1}A]$ for any Borel subset $A \subset \mathbb{R}^d$.

Assumptions 2 and 3 correspond to the fact that D_P is a geometric property, hence should not change the description of the distribution under linear isometries. One can require other desirable properties such as monotonicity along rays from the deepest point, but we do not mention them here. For an axiomatic approach of depth functions and other properties, see, e.g., [90] and [91].

Let us mention a few popular depth functions in $\mathbb{R}^d$:

- *Halfspace depth*, introduced in [98], is defined as

$$D_P(x) = \inf_{u \in S^{d-1}} P[H_{x,u}],$$

  where $H_{x,u} := \{z \in \mathbb{R}^d : u'(z - x) \geq 0\}$ is the halfspace with normal unit vector u the boundary of which passes through x; we let $u'(z-x)$ denote the usual Euclidean inner product between u and $z-x$ in $\mathbb{R}^d$. Among other features, halfspace depth is equivariant under affine transformations, vanishes at infinity, and decreases along rays from the deepest point. For further details, we refer the reader to [29], [40], [74].

- *Simplicial depth* was introduced in [60] and [61]. Letting $X_1, \ldots, X_{d+1}$ be i.i.d. random vectors with common distribution P, it is defined for any $x \in \mathbb{R}^d$ as

$$D_P(x) = P[S(X_1, \ldots, X_{d+1}) \ni x],$$

where, for any $z_1, \ldots, z_{d+1} \in \mathbb{R}^d$, we let $S(z_1, \ldots, z_{d+1})$ denote the simplex of $\mathbb{R}^d$ with vertices $z_1, \ldots, z_{d+1}$. Simplicial depth is equivariant under affine transformations and vanishes outside the convex hull of the support of the distribution. Its computation, however, is very costly; see, e.g., [10], [19], and [30].

- *Center-outward quantiles and ranks*, introduced in [21] and [44] are a multivariate extension of the usual univariate quantile map and cdf, in addition to defining a notion of depth. The center-outward rank of P is defined as the unique gradient $\nabla\varphi_P$ of a convex function $\varphi_P : \mathbb{R}^d \to \mathbb{R}$ that pushes P onto the uniform over the unit d-dimensional ball B, i.e.

$$P[(\nabla\varphi_P)^{-1}(A)] = \int_A \frac{1}{\|x\|^{d-1}}\, dx$$

for any Borel subset $A \subset B$, where $\|x\|$ stands for the Euclidean norm of x. Under mild assumptions, the map $\nabla\varphi_P$ is a homeomorphism, hence admits an inverse called the center-outward quantile map of P. The associated depth is defined, for any $x \in \mathbb{R}^d$, as

$$D_P(x) = 1 - \|\nabla\varphi_P(x)\|.$$

These quantiles, and the associated depth, are not equivariant under affine transformations. They are, however, equivariant under orthogonal transformations. In addition, they allow for controlling probability content. For further details, we refer the reader to [28], [38], and [45].

There are many other concepts of depth, each having its advantages. Some other examples include *majority depth* [64], *projection depth* ([62], [107]), *Mahalanobis depth* ([62], [64], [66]), *Oja depth* [80], *zonoid depth* [51] and L^p*-depth* [108]. Depending on the context, different properties may be required: some equivariance, (non-)vanishing depth, characterizing the underlying distribution, etc. There are also instances in which a property holds in the most useful cases, but not in general. See e.g. [75], [76], and [96] for interesting discussions about the characterization property for halfspace depth. With so many different behaviors being displayed, it becomes difficult to determine *the right* depth concept. In practice, the statistician will use the one that best matches the properties needed in the problem at hand.

This thesis is concerned with another depth concept, called *geometric quantiles, ranks and depth*. Spatial quantiles are not affine equivariant. However, they are equivariant under translations and orthogonal transformations. Furthermore, geometric ranks and depth are given in closed form for any probability measure, a feature shared by few other concepts. This leads to trivial evaluation in empirical cases, and allows for explicit asymptotic normality results while most competitors offer consistency results at best.

In the rest of this chapter, we introduce this notion in two different ways before describing some of its essential properties.

1.2.1 Two ways of introducing geometric quantiles and ranks

In this section, we present two approaches that lead to multivariate geometric quantiles and ranks.

First approach

Recall that the cdf of a probability measure P over $\mathbb{R}$ is given by $F_P(x) = P[(-\infty, x)]$ for any $x \in \mathbb{R}$. In geometrical terms, it would be more natural to consider

$$F_P^{\pm}(x) := 2F_P(x) - 1$$

instead of F_P. Indeed, while $F_P(x) \in [0,1]$, $F_P^{\pm}$ takes its values in $[-1,1]$, which is compatible with the left/right symmetry of $\mathbb{R}$. This leads to a (potentially generalized) quantile function $Q_P^{\pm} : [-1,1] \to \mathbb{R}$, defined as the (generalized) inverse of $F_P^{\pm}$. In other words, the median (the innermost quantile) is $Q_P^{\pm}(0)$, and the most extreme quantiles on the right are obtained via $Q_P^{\pm}(x)$ as $x \to 1$, while the most extreme quantiles on the left are obtained via $Q_P^{\pm}(x)$ as $x \to -1$. The quantile regions I_τ, defined in (1.1.2), can alternatively be obtained as

$$I_\tau = \{z \in \mathbb{R} : |F_P^{\pm}(z)| \leq \tau\} \tag{1.2.3}$$

for any $\tau \in (0,1)$.

Observe that

$$2\mathbb{I}[(-\infty, x]](z) - 1 = \operatorname{sign}(x - z)$$

for any $x, z \in \mathbb{R}$. It follows that

$$F_P^{\pm}(x) = \int_{\mathbb{R}} \operatorname{sign}(x - z)\, dP(z)$$

for any $x \in \mathbb{R}$. Now, let us note that

$$\operatorname{sign}(x - z) = \frac{x - z}{|x - z|}$$

for any $x, z \in \mathbb{R}$ with $x \neq z$. Therefore, we define the *multivariate geometric rank* R_P of a probability measure P over $\mathbb{R}^d$, a multivariate analog of $F_P^{\pm}$ when $d = 1$, by letting

$$R_P(\mu) = \int_{\mathbb{R}^d} \frac{\mu - z}{\|\mu - z\|} \mathbb{I}[z \neq \mu]\, dP(z)$$

for any $\mu \in \mathbb{R}^d$. The geometric rank R_P takes its values in the closed unit ball

$$\overline{B} := \{z \in \mathbb{R}^d : \|z\| \leq 1\},$$

which reduces to $[-1,1]$ when $d=1$. Similarly to (1.2.3), we let the geometric quantile regions be defined as

$$\{\mu \in \mathbb{R}^d : \|R_P(\mu)\| \leq \alpha\}$$

for any $\alpha \in [0,1]$. Since R_P takes its values in $\overline{B}$, we let the quantile of order $\alpha \in [0,1]$ in direction $u \in S^{d-1}$ be the set of solutions $x \in \mathbb{R}^d$ to the equation

$$R_P(\mu) = \alpha u.$$

The corresponding depth, introduced in [102], is the map $D_P : \mathbb{R}^d \to [0,1]$ defined as

$$D_P(\mu) = 1 - \|R_P(\mu)\|$$

for any $\mu \in \mathbb{R}^d$.

Second approach

Consider a probability measure P over $\mathbb{R}$, and let $\tau \in (0,1)$ be fixed. Let us also recall that a point $x \in \mathbb{R}$ is a quantile of order τ for P if

$$P\Big[(-\infty, x)\Big] \leq \tau \leq P\Big[(-\infty, x]\Big]. \tag{1.2.4}$$

We now define the map M_τ from $\mathbb{R}$ to $[0,\infty)$ by letting

$$M_\tau(q) := \int_{\mathbb{R}^d} \Big\{|z-q| + (2\tau-1)(z-q) - |z| - (2\tau-1)z\Big\}\, dP(z)$$

for any $q \in \mathbb{R}$. The terms $-|z| - (2\tau-1)z$ have no relevance here. We only write them so that the integral defining M_τ always exists, even when

$$\int_{\mathbb{R}^d} |z|\, dP(z) = \infty.$$

The map M_τ is convex over $\mathbb{R}$, hence continuous. Convexity also ensures that it admits left- and right-derivatives, $\partial^+ M_\tau$ and $\partial^- M_\tau$ respectively, at all points. Therefore, any minimum q of this function (if it exists) is characterized by the "first-order conditions" $\partial^- M_\tau(q) \leq 0$ and $\partial^+ M_\tau(q) \geq 0$. A straightforward computation gives, for any $q \in \mathbb{R}$,

$$\partial^- M_\tau(q) = 2\int_{\mathbb{R}} \Big\{\mathbb{I}[z<q] - \tau\Big\}\, dP(z) = 2\Big(P\Big[(-\infty,x)\Big] - \tau\Big),$$

and

$$\partial^+ M_\tau(q) = 2\int_{\mathbb{R}} \Big\{\mathbb{I}[z \leq q] - \tau\Big\}\, dP(z) = 2\Big(P\Big[(-\infty,x]\Big] - \tau\Big).$$

We conclude (i) that minimizers of M_τ always exist, and (ii) that these minimizers are precisely the quantiles of order τ for P.

We obtain a straightforward multivariate extension of univariate quantiles by replacing the absolute values $|\cdot|$ by Euclidean norms $\|\cdot\|$, and $2\tau - 1 \in (-1,1)$ by a unit vector αu, with $\alpha \in (0,1)$ and $u \in S^{d-1}$. This leads to the following definition of geometric quantiles, introduced in [17].

Definition 1.2.1. *Let P be a probability measure over $\mathbb{R}^d$. Fix $\alpha \in [0,1)$ and $u \in \mathcal{S}^{d-1}$, where $\mathcal{S}^{d-1} := \{z \in \mathbb{R}^d : \|z\|^2 = z'z = 1\}$ is the unit sphere in $\mathbb{R}^d$. We say that $\mu_{\alpha,u} = \mu_{\alpha,u}(P)$ is a* geometric quantile of order α in direction u for P *if and only if it minimizes the objective function*

$$\mu \mapsto M_{\alpha,u}(\mu) := \int_{\mathbb{R}^d} \Big\{ H_{\alpha,u}(z-\mu) - H_{\alpha,u}(z) \Big\} \, dP(z) \tag{1.2.5}$$

over $\mathbb{R}^d$, with

$$H_{\alpha,u}(z) := \|z\| + \alpha u'z. \tag{1.2.6}$$

We know explain how the two approaches we presented are related. Since $M_{\alpha,u}$ is a convex map, under differentiability assumptions the quantiles obtained in the second approach are solutions $\mu \in \mathbb{R}^d$ to the first-order condition $\nabla M_{\alpha,u}(\mu) = 0$. Computing $\nabla M_{\alpha,u}(\mu)$ by formally taking the gradient under the integral in $M_{\alpha,u}$ leads to

$$\int_{\mathbb{R}^d} \frac{z-\mu}{\|z-\mu\|} \mathbb{I}[z \neq \mu] \, dP(z) - \alpha u = 0.$$

This last equation rewrites as $R_P(\mu) = \alpha u$, which is the definition given in the first approach.

1.2.2 Essential properties of geometric quantiles and ranks

Let us now give some important properties of geometric quantiles and ranks.

Recall that the map $M_{\alpha,u}$ is convex, hence continuous for any $\alpha \in [0,1)$ and $u \in S^{d-1}$. Furthermore, it is not difficult to see that it is coercive, in the sense that

$$\lim_{\|\mu\| \to \infty} M_{\alpha,u}(\mu) = \infty,$$

provided $\alpha \in [0,1)$. Coercivity and continuity entail that $M_{\alpha,u}$ admits at least one minimizer. Therefore, geometric quantiles of arbitrary order $\alpha \in [0,1)$ in direction $u \in S^{d-1}$ for P always exist.

As a direct consequence of the convexity of $M_{\alpha,u}$, geometric quantiles are unique when P is not supported on a single line of $\mathbb{R}^d$. When P is supported on a line, the same issues we encounter in $\mathbb{R}$ can occur, for instance if P vanishes over some interval. Theorem 1 in [83] provides a refinement of this result, which we recall in the next proposition.

Proposition 1.2.2 (Paindaveine, Virta)**.** *Let P be a probability measure over $\mathbb{R}^d$. Fix $\alpha \in [0,1)$ and $u \in S^{d-1}$. Then, (i) P admits a geometric quantile $\mu_{\alpha,u}$. (ii) If P is not supported on a line, then $\mu_{\alpha,u}$ is unique. (iii) If P is not supported on a line with direction u, then $\mu_{\alpha,u}$ is unique for any $\alpha > 0$. (iv) If P is supported on the line $\mathscr{L} = \{z_0 + \lambda u : \lambda \in \mathbb{R}\}$, then any geometric quantile $\mu_{\alpha,u}$ belongs to $\mathscr{L}$; in this case, any such quantile is of the form $\mu_{\alpha,u} = z_0 + \ell_\alpha u$, where ℓ_α is a geometric quantile of order α in direction 1 for $P_{z_0,u}$, with $P_{z_0,u}$ the distribution of $u'(Z - z_0)$ when Z is a random d-vector with distribution P.*

A routine application of Lebesgue's dominated convergence theorem entails that $M_{\alpha,u}$ is differentiable at $\mu \in \mathbb{R}^d$ if $P[\{\mu\}] = 0$, with

$$\nabla M_{\alpha,u}(\mu) = R_P(\mu) - \alpha u,$$

where we recall that

$$R_P(x) = \int_{\mathbb{R}^d} \frac{x - z}{\|x - z\|} \mathbb{I}[z \neq x] \, dP(z)$$

for any $x \in \mathbb{R}^d$.

Assume that P is not supported on a line of $\mathbb{R}^d$. In this case, geometric quantiles exist and are unique. Furthermore, the convexity of $M_{\alpha,u}$ entails that the geometric quantile $\mu_{\alpha,u}$ of order $\alpha \in [0, 1)$ in direction $u \in S^{d-1}$ for P is uniquely determined by the equation $R_P(\mu_{\alpha,u}) = \alpha u$. In particular, the map $R_P : \mathbb{R}^d \to B^d$ is a bijection between $\mathbb{R}^d$ and the open unit ball B^d of $\mathbb{R}^d$. Existence and uniqueness of quantiles also allow us to define the geometric quantile map $Q_P : B^d \to \mathbb{R}^n$, defined by letting $Q_P(\alpha u) = \mu_{\alpha,u}$ for any $\alpha \in [0, 1)$ and $u \in S^{d-1}$. The previous observations imply that R_P and Q_P are inverse to one another. If we further assume that P is non-atomic, Theorem 2.6 (iii) in [50] entails that R_P is a homeomorphism, with $R_P^{-1} = Q_P$. The fact that R_P and Q_P are bijections between $\mathbb{R}^d$ and B^d is surprising: geometric quantiles span the whole space, even if P is compactly supported ! This phenomenon is typical of dimensions $d > 1$, and does not occur in $\mathbb{R}$. Indeed, probability measures over $\mathbb{R}$ are always supported on a line, by construction.

As required by the beginning of our discussion about depth functions, geometric quantiles are equivariant under translations and orthogonal transformations. If P is invariant under any orthogonal transformation, Proposition 2.2 in [42] entails that $Q_P(\alpha u) = r_\alpha u$ for any $\alpha \in [0, 1)$ and $u \in S^{d-1}$, where $r_\alpha > 0$ only depends on α. Even though geometric quantiles in direction $u \in S^{d-1}$ do not necessarily have direction u in general (see, e.g., [74]), Theorem 2.1 in [42] implies that they do so asymptotically.

Detecting outliers in datasets is one of the main uses of depth functions. Properties of extreme geometric quantiles, i.e. quantiles $\mu_{\alpha,u}$ indexed by an order $\alpha \to 1$, are studied in [41], [42], and [83]. For arbitrary sequences $(\alpha_n) \subset [0, 1)$ and $(u_n) \subset S^{d-1}$ such that $\alpha_n \to 1$ as $n \to \infty$, they show under suited assumptions i) that $\|\mu_{\alpha_n,u_n}\| \to \infty$, and (ii) that $\mu_{\alpha_n,u_n}/\|\mu_{\alpha_n,u_n}\| \to u$, as $n \to \infty$. Under additional moment assumptions, [42] provides a quantitative version of this result.

Let us consider a random sequence $X_1, \ldots, X_n, \ldots$ drawn from P, and assume that P admits a bounded density with respect to the Lebesgue measure. For any n, let P_n be the empirical measure associated to $\{X_1, \ldots, X_n\}$, i.e.

$$P_n = \sum_{i=1}^{n} \delta_{X_i},$$

where δ_z is the Dirac measure at z. With P-probability 1, the measures P_n are not supported on a single line of $\mathbb{R}^d$. In particular, *empirical geometric quantiles*, i.e. geometric quantiles of P_n, are unique with P-probability 1. For any $\alpha \in [0, 1)$ and $u \in S^{d-1}$, let $\hat{\mu}^{(n)}_{\alpha,u}$ denote the almost surely unique quantile of order α in direction u for P_n, and $\mu_{\alpha,u}$ the corresponding geometric quantile for P. Then, Theorem 3.1.1 in [17] yields the following asymptotic normality result:

$$\sqrt{n}\left(\hat{\mu}^{(n)}_{\alpha,u} - \mu_{\alpha,u}\right) = A^{-1}_{\alpha,u} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{X_i - \mu_{\alpha,u}}{\|X_i - \mu_{\alpha,u}\|} + \alpha u \right\} + r_n,$$

where, letting I_d stand for the $d \times d$ identity matrix, we define

$$A_{\alpha,u} = \int_{\mathbb{R}^d} \frac{1}{\|z - \mu_{\alpha,u}\|} \left\{ I_d - \frac{(z - \mu_{\alpha,u})(z - \mu_{\alpha,u})'}{\|z - \mu_{\alpha,u}\|^2} \right\} dP(z),$$

and r_n is almost surely $O(\log(n)/n)$ if $d \geq 3$, or $o(n^{-\beta})$ for any $\beta \in (0, 1)$ when $d = 2$.

Similarly to the cdf in dimension $d = 1$, Theorem 2.5 in [50] shows that geometric ranks characterize probability measures. Namely, if P_1 and P_2 are probability measures over $\mathbb{R}^d$ such that $R_{P_1}(\mu) = R_{P_2}(\mu)$ for any $\mu \in \mathbb{R}^d$, then $P_1 = P_2$.

For further details on geometric quantiles, ranks, and depth we refer the reader to [17], [41], [42], [71], [73], [81], [102], and [106].

1.3 Outline of the thesis

In the first chapter, we address the problem of recovering a probability measure P over $\mathbb{R}^n$ from the associated multivariate geometric rank R_P. Multivariate geometric ranks characterize probability measures (see Theorem 2.5 in [50]). If P has a density f_P, we strengthen this result and show that $f_P = \mathscr{L}_n(R_P)$, where $\mathscr{L}_n$ is a (potentially fractional) partial differential operator given in closed form, and depends on n. When P admits no density, we further show that the equality $P = \mathscr{L}_n(R_P)$ still holds in the sense of distributions (i.e. generalized functions). We thoroughly investigate the regularity properties of geometric ranks and use the PDE we established to give qualitative results on depths contours and regions. We study the local properties of the operator $\mathscr{L}_n$, and show that it is non-local when n is even. Then, we give a partial counterpart to the non-localizability in even dimensions. We use the PDE mentioned above to show that geometric ranks characterize weak convergence in the same way the univariate cdf does. Extending a result from [72], we conclude this chapter by showing that a Glivenko-Cantelli result holds for empirical ranks under weaker assumptions.

By substituting an L_p loss function for the L_1 loss function in the optimization problem defining quantiles, one obtains L_p-quantiles that dominate their classical L_1-counterparts in financial risk assessment. In the second chapter, we introduce a concept of multivariate L_p-quantiles generalizing the geometric L_1-quantiles from [17]. Rather than restricting to power loss functions, we actually allow for a large class of convex loss functions ρ. We carefully study existence and uniqueness of the resulting ρ-quantiles, both for a general probability measure over $\mathbb{R}^d$ and for a spherically symmetric one. Interestingly, the results crucially depend on ρ and on the nature of the underlying probability measure. Building on an investigation of the differentiability properties of the objective function defining ρ-quantiles, we introduce a companion concept of geometric ρ-depth, that generalizes the geometric depth from [102]. We study extreme ρ-quantiles and show in particular that extreme L_p-quantiles behave in fundamentally different ways for $p \leq 2$ and $p > 2$. Finally, we establish Bahadur representation results for sample ρ-quantiles and derive their asymptotic distributions. Throughout, we impose only very mild assumptions on the underlying probability measure, and in particular we never assume absolute continuity with respect to the Lebesgue measure.

In the third chapter, we propose a concept of quantiles for probability measures on the unit hypersphere $\mathcal{S}^{d-1}$ of $\mathbb{R}^d$. The innermost quantiles are Fréchet medians, that is, the L_1-analog of Fréchet means. Since these medians may be non-unique, we define a quantile field around each such median m. The proposed quantiles $\mu^m_{\alpha,u}$ are directional in nature: they are indexed by a scalar order $\alpha \in [0, 1)$ and a unit vector u in the tangent space $T_m\mathcal{S}^{d-1}$ to $\mathcal{S}^{d-1}$ at m. To ensure computability in any dimension d, our quantiles are essentially obtained by considering the Euclidean [17] geometric quantiles in a suitable stereographic projection of $\mathcal{S}^{d-1}$ onto $T_m\mathcal{S}^{d-1}$. Despite this link with Euclidean geometric quantiles, studying the proposed spherical quantiles requires understanding the nature of the [17] quantile in a version of the projective space where all points at infinity are identified. We thoroughly investigate the structural properties of our quantiles and we further study the asymptotic behaviour of their sample versions, which requires controlling the impact of estimating m. Our spherical quantile concept also allows for companion concepts of ranks and depth on the hypersphere. We illustrate the relevance of our construction by considering an inferential application related to testing for rotational symmetry.

1.4 A comment on the chapters

Although the three main chapters are presented as independent, one might want to estbalish links between them.

For instance, it is natural to raise the question of whether it is possible to apply the results of Chapter 2 to Chapter 3 to recover a probability measure from its multivariate geometric ρ-rank. Unfortunately, the approach used in Chapter 2 does not apply to the general framework of ρ-ranks. On the one hand, the geometric rank writes as a convolution, which makes it possible to take its Fourier transform and apply results from the theory of distributions. On the other hand, ρ-ranks write as the product between a matrix and a vector, and only the vector possesses a convolutional form, to which the ρ-rank reduces in the case of the usual geometric rank.

Because the spherical quantiles and rank of a probability measure P over S^{d-1} given in Chapter 4 are defined as the geometric quantiles and rank of a stereographic projection of P, the results of Chapters 2 and 3 readily apply. This is why we did not feel the need to further explicit this connection in the thesis.

Chapter 4 might sometimes feel like a straightforward application of standard results about geometric quantiles. Although this is partly true, the fact that the anti-median (the antipode of the Fréchet median) simultaneously plays the role of "infinity" in the Euclidean case, and belongs to the space itself makes it a setting distinct from a purely Euclidean one. Indeed, the very definition of our spherical quantiles and rank must take into account a "projective" effect due to the stereographic projection through the anti-median. For this reason, both the concepts we define and the proofs we give in Chapter 4 are more subtle than direct applications of well-known results.

The three chapters forming this thesis have been submitted to different journals, which explains why they sometimes do not share the exact same structure and mathematical conventions.

Chapter 2

Recovering a probability measure from its multivariate geometric rank

2.1 Introduction

The cumulative distribution function (cdf) and quantiles of a probability measure over $\mathbb{R}$ play a central role in probability and statistics. They characterize the underlying probability measure, and allow one to control the probability content of a given region. One of the most notable uses of quantiles is arguably the design of hypotheses tests in which one is able to establish confidence regions for the estimation of a quantity of interest. Furthermore, the cdf pushes the probability measure from which it derives onto the uniform distribution over the interval $[0, 1]$. This gives a very simple way to generate random observations that follow a given probability law. For these reasons, a lot of effort has been made over the past decades to extend the notions of cdf and quantiles of a probability measure to a multivariate setting. It has long been known that considering quantiles of a probability measure P over $\mathbb{R}^n$ componentwise leads to quantile regions that are not equivariant with respect to orthogonal transformations, which automatically discarded this approach. The first and probably most famous attempt to establish a multivariate analog of cdf's and quantiles is the concept of *halfspace depth*, introduced in [98] which, to any point of $\mathbb{R}^n$, associate a non-negative number (its *depth*). The regions of points the depth of which does not exceed a given treshold value are interpreted as analogs of quantile regions in $\mathbb{R}$. Many other concepts have followed, such as simplicial depth [60] and projection depth [107] to cite only a few. Other approaches have been adopted, attempting to define a proper notion of multivariate quantiles and cumulative distribution functions (called *ranks*), some of the most notable being based on regression quantiles [46], or optimal transport [44]. We also refer the reader to [88] for a review on the topic.

Among the concepts extending quantiles and cdf's to a multivariate setting, one very popular approach is that of *geometric* multivariate ranks and quantiles, introduced in [17]. Spatial ranks and quantiles enjoy important advantages over other competing approaches. Among them, let us stress that geometric ranks are available in closed form, which leads to trivial evaluation in the empirical case, unlike most competing concepts. As a consequence, explicit Bahadur-type representations and asymptotic normality results are provided in [17] and [50], when competing approaches offer at best consistency results only. Spatial ranks and quantiles also allow for direct extensions in infinite-dimensional Hilbert spaces ; see, e.g., [15] and [22]. We refer the reader to [81] for an overview of the scope of applications geometric ranks offer.

Let P be a probability measure P over $\mathbb{R}^n$. A *geometric quantile* of order $\alpha \in [0,1)$ in direction $u \in S^{n-1}$ for P is defined as an arbitrary minimizer of the objective function

$$x \mapsto O^P_{\alpha,u}(x) := \int_{\mathbb{R}^n} \Big\{ |z-x| - |z| - (\alpha u, x) \Big\} \, dP(z),$$

where $|y| := \sqrt{(y,y)}$ is the Euclidean norm of $y \in \mathbb{R}^n$ and $(u,v) = \sum_{i=1}^n u_i v_i$ is the Euclidean inner product between $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$ and $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$. For any $x \in \mathbb{R}^n$, let also

$$R_P(x) = \int_{\mathbb{R}^n \setminus \{z\}} \frac{x-z}{|x-z|} \, dP(z)$$

denote the *geometric rank* map of P. In the univariate case $n = 1$, it is showed in [32] that geometric quantiles of order $\alpha \in [0,1)$ in direction $u \in \{-1,+1\}$ reduce to the usual quantiles of order $(\alpha u + 1)/2 \in [0,1)$. Still when $n = 1$, we have

$$R_P(x) = \int_{\mathbb{R}} \mathrm{sign}(x-z) \, dP(z),$$

and

$$R_P(x) = 2F_P(x) - 1$$

for any $x \in \mathbb{R}$, where $F_P(x) := P[(-\infty, x]]$ stands for the usual univariate cdf of P. This motivates calling Q_P and R_P multivariate analogs of the univariate quantile map and univariate cdf, respectively.

In a general multivariate framework $n \geq 1$, we prove in Chapter 3 that $O^P_{\alpha,u}$ is continuously differentiable over $\mathbb{R}^n$ and that

$$\nabla O^P_{\alpha,u}(x) = R_P(x) - \alpha u$$

for any $x \in \mathbb{R}^n$, provided P has no atoms. Further requiring that P is not supported on a single line of $\mathbb{R}^n$, it is proved in [83] that $O^P_{\alpha,u}$ is strictly convex over $\mathbb{R}^n$ and, therefore, that geometric quantiles of order α in direction u for P are unique for any $\alpha \in [0,1)$ and $u \in S^{n-1}$; let us write this quantile as $Q_P(\alpha u)$. This implies that $Q_P(\alpha u)$, i.e. the unique minimizer of $O^P_{\alpha,u}$, is the unique solution $x \in \mathbb{R}^n$ to the equation

$$R_P(x) = \alpha u.$$

Under the above assumptions, we show in Chapter 3 that the quantile map $\alpha u \mapsto Q_P(\alpha u)$ is invertible with inverse $Q_P^{-1} = R_P$. This provides another motivation to regard R_P as a natural multivariate analog of the univariate cdf.

The so-called *depth contours* and *depth regions* of P are of particular interest. Assume that P is not supported on a single line of $\mathbb{R}^n$ so that geometric quantiles for P are unique. For any $\beta \in [0, 1)$, we let

$$\mathcal{D}_P^\beta = \left\{ Q_P(\alpha u) : \alpha \in [0, \beta],\ u \in S^{n-1} \right\},$$

and

$$\mathcal{C}_P^\beta = \left\{ Q_P(\beta u) : u \in S^{n-1} \right\}$$

be the *depth region* and *depth contour* of order β for P, respectively. Depth regions provide a family of smooth compact arc-connected and nested centrality regions, while depth contours are disjoint compact and arc-connected $(n-1)$-dimensional smooth manifolds (see Section 2.7).

As we already mentioned, the conceptual and computational simplicity of geometric ranks and quantiles allow for explicit qualitative and quantitative results. Therefore, geometric ranks and quantiles are well-understood; see, e.g. [41] and [42] for interesting features of geometric quantiles. Similarly to their univariate counterpart, it is well-know that geometric ranks characterize probability measures in arbitrary dimension n : if P and Q are Borel probability measures over $\mathbb{R}^n$, and if $R_P(x) = R_Q(x)$ for any $x \in \mathbb{R}^n$, then $P = Q$ (see Theorem 2.5 in [50]). Note that this very desirable property is also shared by ranks based on optimal transport and, when P admits a sufficiently smooth density, the density can be recovered from the rank via a (highly non-linear) partial differential equation, see [44]. The characterization property is not shared by the concept of halfspace depth (see [75]). However, halfspace depth possesses the characterization property within an important class of probability measures ; see [96], who gave the first positive result for empirical probability measures by algorithmically reconstructing the measure. We refer the reader to [76] for a review on the question of characterization for halfspace depth. Therefore, it is most natural to explore the possibility of recovering a probability measure from its geometric rank. In the present chapter, we show that any Borel probability measure P over $\mathbb{R}^n$ can be reconstructed from its geometric rank R_P through a (potentially fractional) linear partial differential equation involving R_P only. We further show that this result holds even when P admits no density; this extends the characterization, given by Theorem 2.5 in [50], with a degree of generality that outperforms similar results known for halfspace-depth and quantiles based on optimal transport.

The structure of this chapter is as follows. In section 2.2, the main definitions are stated. We discuss the strategy used to recover a probability measure P over $\mathbb{R}^n$ knowing its multivariate geometric rank R_P only. Some notations and usual spaces are introduced in Section 2.3. Section 2.4 is devoted to a brief review on distribution theory and Sobolev spaces. We introduce fractional Laplacians, which are key ingredients all along the chapter, in Section 2.5. In Section 2.6, we establish the PDE relating an arbitrary probability measure P over $\mathbb{R}^n$ to its multivariate geometric rank R_P in the sense of distributions. We thoroughly investigate the regularity properties of geometric ranks, and give sufficient conditions for the above PDE to hold pointwise. By exploiting the results of Section 2.6, we establish some regularity properties of geometric quantile contours in Section 2.7. In Section 2.8, we give a refinement of the characterization property of geometric ranks, for odd dimensions only, before stating a partial counterpart to the non-local nature of the PDE in even dimensions. In Section 2.9, we prove that geometric ranks characterize convergence in distribution. We conclude this chapter by proving a Glivenko-Cantelli result for geometric ranks in Section 2.10.

2.2 Main results

Throughout, $\mathbb{I}[A]$ will denote the indicator function of the condition A.

In the following definition, we introduce a quantity strongly related to the map $O^P_{\alpha,u}$ introduced in Section 2.1.

Definition 2.2.1. *Let $n \geq 1$, and P be a Borel probability measure over $\mathbb{R}^n$. We define the map $g_P : \mathbb{R}^n \to \mathbb{R}$ by letting*

$$g_P(x) = \mathbb{E}\big[|x - Z| - |Z|\big]$$

for any $x \in \mathbb{R}^n$, where Z is a random n-vector with law P.

We obivously have $O^P_{\alpha,u}(x) = g_P(x) - (\alpha u, x)$ for any $x \in \mathbb{R}^n$, so that

$$\nabla O^P_{\alpha,u} = \nabla g_P - \alpha u$$

whenever $O^P_{\alpha,u}$ is differentiable. In view of the discussion of Section 2.1, it is clear that g_P is strongly related to R_P (see also Definition 2.2.2).

The triangle inequality entails that g_P is well-defined, irrespective of the probability measure P : no moment assumption is made. It is further easy to see that g_P is continuous over $\mathbb{R}^n$. Theorem 3.5.2 in Chapter 3 implies that g_P is continuously differentiable over an open subset $U \subset \mathbb{R}^n$ if and only if P has no atoms over U. In that case, we have $R_P(x) = \nabla g_P(x)$ for any $x \in U$, where R_P is given in the next definition.

Definition 2.2.2. *Let $n \geq 1$, and P be a Borel probability measure over $\mathbb{R}^n$. The geometric rank R_P of P is the map $R_P : \mathbb{R}^n \to \mathbb{R}^n$ defined by letting*

$$R_P(x) = \mathbb{E}\left[\frac{x - Z}{|x - Z|}\mathbb{I}[Z \neq x]\right]$$

for any $x \in \mathbb{R}^n$, where Z is a random n-vector with law P.

To recover P from R_P, it is crucial to notice that R_P is a convolution between P and a fixed kernel, which we will denote by K in the sequel. Formally taking the Fourier transform $\mathscr{F}(R_P)$ of R_P (in the sense of distributions, i.e. generalized functions) gives

$$\mathscr{F}(R_P) = \mathscr{F}(K)\mathscr{F}(P). \tag{2.2.1}$$

This fact was already noticed in [50], and we used this idea as a building block to prove the results we present. Recovering P now essentially amounts to isolating $\mathscr{F}(P)$ in

(2.2.1), and taking the inverse Fourier transform of $\mathscr{F}(P)$. To obtain an explicit formula for $\mathscr{F}(P)$, we will need an explicit formula for $\mathscr{F}(K)$. Therefore, let us introduce the kernel $K : \mathbb{R}^n \to \mathbb{R}^n$, defined as

$$K(x) = \begin{cases} \frac{x}{|x|} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

At least formally, we have $R_P(x) = (K * P)(x)$. When confusion about the dimension is possible, we will write K_n instead of K to emphasize that we consider K over $\mathbb{R}^n$.

We will show that the Fourier transform $\mathscr{F}(K_n)$ of K_n is given, up to a multiplicative constant C_n that only depends on n, by

$$\text{P.V.}\Big(\frac{\xi}{|\xi|^{n+1}}\Big),$$

where P.V. stands for *principal value.* In other words, we have

$$\int_{\mathbb{R}^n} K_n(x)\widehat{\psi}(x)\,dx = C_n \lim_{\eta\to 0}\int_{\mathbb{R}^n\setminus B_\eta} \frac{\xi}{|\xi|^{n+1}}\psi(\xi)\,d\xi$$

for any $\psi \in \mathscr{S}(\mathbb{R}^n)$, where B_η stands for the ball of radius $\eta > 0$ centered at the origin, and $\mathscr{S}(\mathbb{R}^n)$ (or, equivalently, $\mathscr{S}(\mathbb{R}^n, \mathbb{C})$) stands for the complex-valued Schwarz class over $\mathbb{R}^n$. Before stating our main result, let us introduce the operator $\mathscr{L}_n$, that will play a key role. It involves a constant γ_n, defined as

$$\frac{1}{\gamma_n} = 2^n \pi^{\frac{n-1}{2}} \Gamma\Big(\frac{n+1}{2}\Big).$$

The domain $D(\mathscr{L}_n)$ of the operator $\mathscr{L}_n$ is a space of distributions that depends on n, namely we let $D(\mathscr{L}_n) = \mathscr{S}^n(\mathbb{R}^n)'$, the space of $\mathbb{C}^n$-valued tempered distributions (see Section 2.4.1 for further details) if n is odd, and $D(\mathscr{L}_n) = \mathscr{S}^n_{1/2}(\mathbb{R}^n)'$ (see Section 2.5) if n is even.

Definition 2.2.3 (The operator $\mathscr{L}_n$ and its adjoint $\mathscr{L}_n^*$)**.** *Let $n \in \mathbb{N}$ with $n \geq 1$. Define the (potentially fractional) differential operator $\mathscr{L}_n : D(\mathscr{L}_n) \subset \mathscr{S}^n(\mathbb{R}^n)' \to \mathscr{S}(\mathbb{R}^n)'$ by*

$$\mathscr{L}_n := \gamma_n \begin{cases} (-\Delta)^{\frac{n-1}{2}} \nabla\,\cdot & \text{if } n \text{ is odd,} \\ (-\Delta)^{\frac{1}{2}} (-\Delta)^{\frac{n-2}{2}} \nabla\,\cdot & \text{if } n \text{ is even,} \end{cases}$$

where $\nabla\,\cdot$ is the divergence operator, $(-\Delta)^k$ stands for the Laplacian operator $-\Delta$ taken k times successively when $k \in \mathbb{N}$, and $(-\Delta)^{\frac{1}{2}}$ denotes the fractional Laplacian introduced in Section 2.5.

Define the formal adjoint $\mathscr{L}_n^* : D(\mathscr{L}_n^*) \subset \mathscr{S}(\mathbb{R}^n)' \to \mathscr{S}^n(\mathbb{R}^n)'$ of $\mathscr{L}_n$ by

$$\mathscr{L}_n^* := \gamma_n \begin{cases} \nabla(-\Delta)^{\frac{n-1}{2}} & \text{if } n \text{ is odd,} \\ \nabla(-\Delta)^{\frac{1}{2}} (-\Delta)^{\frac{n-2}{2}} & \text{if } n \text{ is even,} \end{cases}$$

where ∇ stands for the gradient operator, and

$$D(\mathscr{L}_n^*) = \begin{cases} \mathscr{S}(\mathbb{R}^n)' & \text{if } n \text{ is odd,} \\ \mathscr{S}_{1/2}(\mathbb{R}^n)' & \text{if } n \text{ is even.} \end{cases}$$

We call $\mathscr{L}_n^*$ the formal adjoint of $\mathscr{L}_n$ because, letting $\langle \cdot, \cdot \rangle$ denote the distributional bracket (see Definition 2.4.1), we have

$$\langle \mathscr{L}_n \Lambda, \varphi \rangle = \langle \Lambda, \mathscr{L}_n^* \varphi \rangle$$

for any $\Lambda \in D(\mathscr{L}_n)$ and $\varphi \in \mathscr{S}(\mathbb{R}^n)$, and

$$\langle \mathscr{L}_n^* T, \Psi \rangle = \langle T, \mathscr{L}_n \Psi \rangle$$

for any $T \in D(\mathscr{L}_n^*)$ and $\Psi \in \mathscr{S}(\mathbb{R}^n, \mathbb{C}^n)$. In particular, we have

$$\int_{\mathbb{R}^n} (\mathscr{L}_n \Psi)(x) \varphi(x) \, dx = \int_{\mathbb{R}^n} (\Psi(x), \mathscr{L}_n^*(\varphi)(x)) \, dx$$

for any $\Psi \in \mathscr{S}(\mathbb{R}^n, \mathbb{C}^n)$ and $\varphi \in \mathscr{S}(\mathbb{R}^n)$.

Taking Fourier transforms shows that all differential operators involved in the definition of $\mathscr{L}_n$ and $\mathscr{L}_n^*$ commute with each other over $D(\mathscr{L}_n)$ and $D(\mathscr{L}_n^*)$, respectively. This legitimates writing $\mathscr{L}_n$ and $\mathscr{L}_n^*$ in the more compact forms

$$\mathscr{L}_n = \gamma_n \, (-\Delta)^{\frac{n-1}{2}} \nabla \, \cdot,$$

and

$$\mathscr{L}_n^* = \gamma_n \, \nabla (-\Delta)^{\frac{n-1}{2}},$$

irrespective of $n \geq 1$. In $\mathbb{R}$, notice that $\mathscr{L}_1$ and $\mathscr{L}_1^*$ simply reduce to

$$\mathscr{L}_1 = \frac{1}{2} \frac{d}{dx} = \mathscr{L}_1^*.$$

We now state our main result.

Theorem 2.2.4. *Let $n \geq 1$, and P be a Borel probability measure over $\mathbb{R}^n$. Then, R_P belongs to the domain $D(\mathscr{L}_n)$ of $\mathscr{L}_n$, and the equality*

$$P = \mathscr{L}_n(R_P)$$

holds in $\mathscr{S}(\mathbb{R}^n)'$, i.e.

$$\int_{\mathbb{R}^n} \psi(x) \, dP(x) = \int_{\mathbb{R}^n} \Big(R_P(x), (\mathscr{L}_n^* \psi)(x) \Big) \, dx$$

for any $\psi \in \mathscr{S}(\mathbb{R}^n)$.

The previous theorem naturally raises the question of whether it is possible to recover a pointwise equality $f_P = \mathscr{L}_n(R_P)$ when P admits a density f_P. When n is odd, R_P should be at least n times differentiable. When n is even, R_P should be at least $n-1$ times differentiable, and such that $(-\Delta)^{\frac{1}{2}}(-\Delta)^{\frac{n-2}{2}}(\nabla \cdot R_P)$ exists pointwise. Differentiability up to order $n-1$ is obtained in Proposition 2.6.8, via a standard argument based on dominated convergence. Since the $(n-1)$th derivatives of R_P behave like

$$x \mapsto \int_{\mathbb{R}^n} \frac{1}{|x - z|^{n-1}} f_P(z) \, dz,$$

and since $\frac{1}{|x|^n}$ is not integrable near the origin nor at infinity in $\mathbb{R}^n$, the differentiability of order n cannot be addressed by the dominated convergence argument. Our strategy consists in using the PDE we established, i.e.

$$f_P = \gamma_n \, (-\Delta)^{\frac{n-1}{2}} (\nabla \cdot R_P),$$

and use elliptic regularity to conclude that $\nabla \cdot R_P \in H^{n-1}_{\text{loc}}$; see Section 2.4.2 for a definition of the spaces H^k. This fact alone, however, is not enough to deduce that $R_P \in H^n_{\text{loc}}$ in general. Nevertheless, we established that $R_P = \nabla g_P$ (see Definition 2.2.1 and the comments below). Consequently, we will have $-\Delta g_P \in H^{n-1}_{\text{loc}}$ which, by elliptic regularity, will lead to $g_P \in H^{n+1}_{\text{loc}}$, hence $R_P \in H^n_{\text{loc}}$.

Theorem 2.2.5 (Odd dimensions)**.** *Let $n \geq 3$ be odd, and P be a Borel probability measure over $\mathbb{R}^n$. Let $\Omega \subset \mathbb{R}^n$ be an open subset, and assume that P admits a density $f_\Omega \in L^1(\Omega)$ over Ω with respect to the Lebesgue measure. Then, we have the following :*

1. *If $f_\Omega \in L^p_{\text{loc}}(\Omega)$ for some $p > n$, then $R_P \in \mathscr{C}^{n-1}(\Omega) \cap H^n_{\text{loc}}(\Omega)$. In particular, R_P admits weak derivatives of order n in Ω, and we have*

$$f_\Omega(x) = (\mathscr{L}_n R_P)(x)$$

for almost any $x \in \Omega$.

2. *If $f_\Omega \in \mathscr{C}^{k,\alpha}_{\text{loc}}(\Omega)$ for some $k \in \mathbb{N}$ and some $\alpha \in (0,1)$, then $R_P \in \mathscr{C}^{k+n}(\Omega)$, and we have*

$$f_\Omega(x) = (\mathscr{L}_n R_P)(x)$$

for any $x \in \Omega$.

Theorem 2.2.6 (Even dimensions)**.** *Let $n \geq 2$ be even, and P a Borel probability measure on $\mathbb{R}^n$. Assume that P admits a density $f_P \in L^1(\mathbb{R}^n)$ with respect to the Lebesgue measure. Let us define*

$$R_P^{(n-1)} := \gamma_n \, (-\Delta)^{\frac{n-2}{2}} (\nabla \cdot R_P).$$

Then, we have the following :

1. *If $n > 2$ and $f_P \in L^p_{\text{loc}}(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ for some $p > n$, then $R_P \in \mathscr{C}^{n-1}(\mathbb{R}^n) \cap H^n_{\text{loc}}(\mathbb{R}^n)$. In particular, R_P admits weak derivatives of order n in $\mathbb{R}^n$, and we have*

$$f_P(x) = (\mathscr{L}_n R_P)(x) = ((-\Delta)^{\frac{1}{2}} R_P^{(n-1)})(x)$$

for almost any $x \in \mathbb{R}^n$. In addition, $R_P^{(n-1)} \in H^1(\mathbb{R}^n)$, and we have

$$((-\Delta)^{\frac{1}{2}} R_P^{(n-1)})(x) = 2\pi \mathscr{F}^{-1}\Big(|\xi| \mathscr{F} R_P^{(n-1)}\Big)(x)$$

for almost any $x \in \mathbb{R}^n$.

2. If $f_P \in \mathscr{C}^{k,\alpha}(\mathbb{R}^n)$ for some $k \in \mathbb{N}$ and $\alpha \in (0,1)$, then $R_P \in \mathscr{C}^{k+n}(\mathbb{R}^n)$ and we have

$$f_P(x) = (\mathscr{L}_n R_P)(x)$$

for any $x \in \mathbb{R}^n$. In addition, we have

$$(\mathscr{L}_n R_P)(x) = c_{n,1/2} \lim_{\eta \to 0} \int_{\mathbb{R}^n \setminus B_\eta(x)} \frac{R_P^{(n-1)}(x) - R_P^{(n-1)}(z)}{|x-z|^{n+1}}\, dz$$

for any $x \in \mathbb{R}^n$ (see Section 2.5 for the definition of $c_{n,1/2}$).

The fundamental difference between odd and even dimensions lies in the local nature of the statement in odd dimensions and the global nature of the statement in even dimensions. This is a direct consequence of the fact that $\mathscr{L}_n$ is a purely local differential operator when n is odd, while $\mathscr{L}_n$ is non-local when n is even due to the presence of $(-\Delta)^{\frac{1}{2}}$ in $\mathscr{L}_n$.

In even dimensions, we require $f_P \in L^2(\mathbb{R}^n)$ although we do not ask it in odd dimensions. This is again due to the local nature of the statement in odd dimensions, which actually requires $f_P \in L^2_{\text{loc}}$. But this condition is automatically verified since we already asked $f_P \in L^p_{\text{loc}}$ with $p > n \geq 2$. Now, the fact that we require $f_P \in L^2(\mathbb{R}^n)$ in even dimensions implies that $\frac{1}{|x|^{n-1}}$ does not belong to $L^2(\mathbb{R}^n)$ when $n = 2$. This is why $n = 2$ is excluded in even dimensions if no Hölder regularity holds.

2.3 Notations

Let $n \geq 1$, and $U \subset \mathbb{R}^n$ be an open subset.

- $\mathbb{N} = \{0, 1, 2, \ldots\}$ is the collection of natural numbers.
- We denote the inner product over $\mathbb{R}^n$ by $(\cdot, \cdot)$.
- For any $x \in \mathbb{R}^n$ and $r > 0$, we let $B_r(x)$ and B_r denote the open ball centered at x with radius r and the ball centered at the origin with radius r, respectively.
- For any subset $A \subset \mathbb{R}^n$, we write $\overline{A}$ for the closure of A with respect to the usual topology of $\mathbb{R}^n$.
- For any function $u : \mathbb{R}^n \to \mathbb{C}$ that is k-times differentiable, we let

  $$\partial^\alpha u(x) := \frac{\partial^{|\alpha|} u}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}}(x)$$

  for any $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ such that $|\alpha| := \sum_{j=1}^n \alpha_j \leq k$. By convention, we let $\partial^\alpha u := u$ if $\alpha = (0, \ldots, 0)$.
- For any $k \geq 0$, $\mathscr{C}^k(U)$ stands for the collection of functions $u : U \to \mathbb{C}$ that are k-times differentiable, and such that $\partial^\alpha u$ is continuous over U for any $\alpha \in \mathbb{N}^n$ with $|\alpha| \leq k$.
- For any $k \geq 0$, $\mathscr{C}^k_b(U)$ is the collection of functions $u \in \mathscr{C}^k(U)$ such that $\partial^\alpha u$ is bounded over U for any $\alpha \in \mathbb{N}^n$ with $|\alpha| \leq k$.

- The set $\mathscr{C}_c^k(U)$ stands for the subset of $\mathscr{C}_b^k$ made of functions whose support is compact and contained in U. The set $\mathscr{C}_c^\infty(U)$ of infinitely differentiable maps with compact support in U is also denoted $\mathscr{D}(U)$.
- We let $\mathscr{C}_0(\mathbb{R}^n)$ denote the collection of (complex-valued) continuous functions that converge to 0 at infinity.
- The Hölder space $\mathscr{C}^{k,\alpha}(U)$, $k \in \mathbb{N}$ and $\alpha \in (0,1]$, stands for the collection of functions $u \in \mathscr{C}^k(U)$ such that $\partial^\beta u$ is bounded over U for any $\beta \in \mathbb{N}^n$ with $|\beta| \leq k$, and such that $\partial^\beta u$ is α-Hölder continuous over U when $|\beta| = k$. Similarly, we let $\mathscr{C}^{k,\alpha}_{\text{loc}}(U)$ be the collection of functions $u \in \mathscr{C}^k(U)$ such that $u \in \mathscr{C}^{k,\alpha}(V)$ for any open bounded subset $V \subset U$ such that $\overline{V} \subset U$.
- When V is a collection of functions $u : \mathcal{T} \to \mathbb{C}^n$ defined over a topological space $\mathcal{T}$, we let V_{loc} denote the collection of functions $u : \mathcal{T} \to \mathbb{C}^n$ such that the restriction $u_{|K}$ of u to any compact set $K \subset \mathcal{T}$ belongs to V.
- For any $u \in L^1(\mathbb{R}^n)$, we define the Fourier transform $\widehat{u}$ of u by letting

$$\widehat{u}(\xi) = \int_{\mathbb{R}^n} u(x) e^{-2i\pi(x,\xi)}\, dx$$

  for any $\xi \in \mathbb{R}^n$. We let $\mathscr{F} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ denote the Fourier transform over $L^2(\mathbb{R}^n)$, defined as the unique continuous extension to $L^2(\mathbb{R}^n)$ of the restriction of the Fourier transform on $\mathscr{S}(\mathbb{R}^n)$. We will also denote by $\mathscr{F}$ the Fourier transform acting on the space $\mathscr{S}(\mathbb{R}^n)'$ of tempered distributions on $\mathbb{R}^n$ (see Section 2.4.1), and acting componentwise on $\mathscr{S}^k(\mathbb{R}^n)'$, for any $k \geq 1$.

In the sequel, we will refer to the following statement as Green's formula, that can be found in Appendix C.2 of [31].

Theorem 2.3.1 (Green's formula)**.** *Let $\Omega \subset \mathbb{R}^n$ be a regular and bounded open subset, and $\partial\Omega$ denote its boundary. Let $u, v \in \mathscr{C}^1(\overline{\Omega})$. For any $i \in \{1, \ldots, n\}$, we have*

$$\int_\Omega (\partial_i u(x)) v(x)\, dx = \int_{\partial\Omega} u(x) v(x) \nu_i(x)\, d\sigma(x) - \int_\Omega u(x) \partial_i v(x)\, dx,$$

where $\nu_i(x)$ is the ith component of the outer unit normal vector to Ω at x, and σ the surface area measure on $\partial\Omega$.

2.4 Brief review of distribution theory and Sobolev spaces

In this section, we review some of the basic analytical tools we will need : distribution theory and Sobolev spaces.

2.4.1 Distributions

The main reference we used for this section is [87].

Let $U \subset \mathbb{R}^n$ be an open subset. Let us recall that the set of infinitely differentiable functions whose support is compact and included in U is denoted in the chapter by $\mathscr{C}_c^\infty(U)$. We endow $\mathscr{C}_c^\infty(U)$ with the following notion of convergence: a sequence $(\varphi_k) \subset \mathscr{C}_c^\infty(U)$ converges to $\varphi \in \mathscr{C}_c^\infty(U)$ in the space $\mathscr{C}_c^\infty(U)$ if there exists a compact subset $K \subset U$ such that $\operatorname{supp}(\varphi_k) \subset K$ for any k and such that

$$\sup_{x\in K} |\partial^\alpha(\varphi_k - \varphi)(x)| \to 0$$

as $k \to \infty$ for any $\alpha \in \mathbb{N}^n$.

Definition 2.4.1 (Distribution)**.** *A distribution on U is a linear map*

$$T : \mathscr{C}_c^\infty(U) \to \mathbb{C}, \;\; \varphi \mapsto \langle T, \varphi\rangle$$

which is continuous with respect to the convergence on $\mathscr{C}_c^\infty(U)$. The set of all distributions on U is denoted $\mathscr{D}(U)'$.

Examples 2.4.2. We list here typical examples of distributions and a few usual ways to obtain distributions from other distributions.

- Any function $f \in L^1_{\mathrm{loc}}(U)$ gives rise to a distribution T_f on U which, by an obivous abuse of notation we also write f, by letting

$$\langle f, \varphi\rangle := \int_U f(x)\varphi(x)\, dx$$

for any $\varphi \in \mathscr{C}_c^\infty(U)$.

- Similarly, any Borel measure μ on U that is finite over compact subsets of U leads to a distribution on U by letting

$$\langle \mu, \varphi\rangle := \int_U \varphi(x)\, d\mu(x)$$

for any $\varphi \in \mathscr{C}_c^\infty(U)$. In particular, any Borel probability measure is a distribution on any open subset of $\mathbb{R}^n$.

- If $T \in \mathscr{D}(U)'$ is a distribution on U, we define its distributional derivatives $\partial^\alpha T$, $\alpha \in \mathbb{N}^n$, by letting

$$\langle \partial^\alpha T, \varphi\rangle := (-1)^{|\alpha|} \langle T, \partial^\alpha \varphi\rangle$$

for any $\varphi \in \mathscr{C}_c^\infty(U)$.

- For any smooth function $f \in \mathscr{C}^\infty(U)$ and distribution $T \in \mathscr{D}(U)'$ on U, we define the distribution fT by letting

$$\langle fT, \varphi\rangle := \langle T, f\varphi\rangle$$

for any $\varphi \in \mathscr{C}_c^\infty(U)$.

Distributions are stable with respect to multiplication by smooth functions, and taking derivatives. Other common operations, such as convolution and Fourier transform, do not leave the space $\mathscr{C}_c^\infty(U)$ invariant and cannot be directly defined on $\mathscr{D}(U)'$. Consequently, we need to use another class of test functions, and the corresponding new distributions, on which we can apply these operations. This is the role of tempered distributions, that rely on the Schwarz class $\mathscr{S}(\mathbb{R}^n)$ defined as

$$\mathscr{S}(\mathbb{R}^n) = \Big\{ f \in \mathscr{C}^\infty(\mathbb{R}^n) : \sup_{x\in\mathbb{R}^n} (1+|x|)^m |\partial^\alpha f(x)| < \infty,\ \forall m \in \mathbb{N},\ \forall \alpha \in \mathbb{N}^n \Big\}.$$

Following our definition of $\mathscr{C}^\infty(\mathbb{R}^n)$, functions from $\mathscr{S}(\mathbb{R}^n)$ are complex-valued (see Section 2.3).

The set $\mathscr{S}(\mathbb{R}^n)$ is a vector space. It is also stable by multiplication, multiplication by smooth functions all derivatives of which have at most polynomial growth at infinity, convolution, differentiation and Fourier transform. We further have the inclusion

$$\mathscr{C}_c^\infty(U) \subset \mathscr{S}(\mathbb{R}^n)$$

for any open subset $U \subset \mathbb{R}^n$. The set $\mathscr{S}(\mathbb{R}^n, \mathbb{C}^k)$, with $k \geq 1$, will stand for the collection of vector fields $\Psi = (\psi_1, \dots, \psi_k)$ for which every component ψ_i belongs to $\mathscr{S}(\mathbb{R}^n)$. Similarly to the space $\mathscr{C}_c^\infty(\mathbb{R}^n)$, we endow $\mathscr{S}(\mathbb{R}^n)$ with an adequate notion of convergence. A sequence $(\psi_k) \subset \mathscr{S}(\mathbb{R}^n)$ converges to $\psi \in \mathscr{S}(\mathbb{R}^n)$ in the space $\mathscr{S}(\mathbb{R}^n)$ if

$$\sup_{x\in\mathbb{R}^n} |(1+|x|)^m \partial^\alpha(\psi_k - \psi)(x)| \to 0$$

as $k \to \infty$ for any $m \in \mathbb{N}$ and $\alpha \in \mathbb{N}^n$.

Definition 2.4.3 (Tempered distribution)**.** *A tempered distribution is a linear map*

$$T : \mathscr{S}(\mathbb{R}^n) \to \mathbb{C},\ \ \psi \mapsto \langle T, \psi \rangle$$

which is continuous with respect to the convergence on $\mathscr{S}(\mathbb{R}^n)$. The set of tempered distributions is denoted $\mathscr{S}(\mathbb{R}^n)'$.

For the sake of simplicity we let

$$\mathscr{S}^k(\mathbb{R}^n)' := (\mathscr{S}(\mathbb{R}^n)')^k$$

for any $k \geq 1$ be the set of linear maps

$$T = (T_1, \dots, T_k) : \mathscr{S}(\mathbb{R}^n, \mathbb{C}^k) \to \mathbb{C}^k,\ \ \Psi = (\psi_1, \dots, \psi_k) \mapsto (\langle T_1, \psi_1 \rangle, \dots, \langle T_k, \psi_k \rangle)$$

such that $T_i \in \mathscr{S}(\mathbb{R}^n)'$ for any $i = 1, \dots, k$. We let all operations described above act on $\mathscr{S}^k(\mathbb{R}^n)'$ componentwise. Therefore, the identities we stated remain valid on $\mathscr{S}^k(\mathbb{R}^n)'$.

It is easy to see that if a sequence $(\varphi_k) \subset \mathscr{C}_c^\infty(\mathbb{R}^n)$ converges to some $\varphi \in \mathscr{C}_c^\infty(\mathbb{R}^n)$ in the space $\mathscr{C}_c^\infty(\mathbb{R}^n)$, then convergence also holds in the space $\mathscr{S}(\mathbb{R}^n)'$. In particular, any tempered distribution is a distribution over $\mathbb{R}^n$.

Examples 2.4.4. Below, we list typical examples of tempered distributions, and a few usual ways to obtain tempered distributions from other tempered distributions.

- If $f : \mathbb{R}^n \to \mathbb{C}$ is a measurable map such that

$$\frac{f(x)}{(1+|x|)^m} \in L^p(\mathbb{R}^n)$$

for some $m \in \mathbb{N}$ and $p \in [1, \infty)$, then $f \in \mathscr{S}(\mathbb{R}^n)'$ by letting

$$\langle f, \psi \rangle := \int_{\mathbb{R}^n} f(x)\psi(x)\, dx$$

for any $\psi \in \mathscr{S}(\mathbb{R}^n)$.

- Similarly, if μ is a Borel measure on $\mathbb{R}^n$ such that

$$\int_{\mathbb{R}^n} \frac{1}{(1+|x|)^m}\, d\mu(x) < \infty$$

for some $m \in \mathbb{N}$, then $\mu \in \mathscr{S}(\mathbb{R}^n)'$.

- Let $T \in \mathscr{S}(\mathbb{R}^n)'$ and $f \in \mathscr{C}^\infty(\mathbb{R}^n)$. We have already mentioned that the product fT is a distribution over $\mathbb{R}^n$. For fT to be tempered, we need $f\psi$ to be a Schwarz function for any $\psi \in \mathscr{S}(\mathbb{R}^n)$. This will be the case, for instance, if f and all its derivatives have at most polynomial growth at infinity. If no restriction is imposed on the growth of f and its derivatives, the product fT might not be tempered; consider, e.g., the tempered distribution $T \equiv 1$, and the smooth function $f(x) = e^x$.

- If $T \in \mathscr{S}(\mathbb{R}^n)'$ is a tempered distribution, then the derivative $\partial^\alpha T$ is also a tempered distribution for any $\alpha \in \mathbb{N}^n$.

- If $T \in \mathscr{S}(\mathbb{R}^n)'$, we define its Fourier transform $\mathscr{F}T$ by letting

$$\langle \mathscr{F}T, \psi \rangle = \langle\ T,\ \widehat{\psi}\ \rangle$$

for any $\psi \in \mathscr{S}(\mathbb{R}^n)$. The map $\mathscr{F}T$ is a tempered distribution. Just as for smooth functions over $\mathbb{R}^n$, the equalities

$$\mathscr{F}(\partial^\alpha T) = (2i\pi\xi)^\alpha \mathscr{F}(T), \quad \text{and} \quad \partial^\alpha \mathscr{F}(T) = \mathscr{F}((-2i\pi x)^\alpha T)$$

hold in $\mathscr{S}(\mathbb{R}^n)'$ for any $\alpha \in \mathbb{N}^n$.

Proposition 2.4.5 (Convolution)**.** *Let $T \in \mathscr{S}(\mathbb{R}^n)'$ and $\psi \in \mathscr{S}(\mathbb{R}^n)$. We define the convolution $T * \psi$ by letting*

$$(T * \psi)(x) = \langle T, \psi(x - \cdot)\rangle$$

for any $x \in \mathbb{R}^n$. The map $T * \psi$ belongs to $\mathscr{C}^\infty(\mathbb{R}^n)$, and we have

$$\partial^\alpha (T * \psi) = (\partial^\alpha T) * \psi = T * (\partial^\alpha \psi)$$

over $\mathbb{R}^n$ for any $\alpha \in \mathbb{N}^n$. Furthermore, $T * \psi$ has polynomial growth. In particular, $T * \psi$ is a tempered distribution over $\mathbb{R}^n$, and the equality

$$\mathscr{F}(T * \psi) = \mathscr{F}(T)\ \widehat{\psi}$$

holds in $\mathscr{S}(\mathbb{R}^n)'$.

2.4.2 Sobolev spaces

The main reference we used for this section is [31].

Let $U \subset \mathbb{R}^n$ be an open subset. A function $u \in L^1_{\text{loc}}(U)$ has weak derivatives of order k in U if, for any $\alpha \in \mathbb{N}^n$ with $|\alpha| \leq k$, the distributional derivative $\partial^\alpha u$ is actually a function, and belongs to $L^1_{\text{loc}}(U)$, i.e. there exists $v_\alpha \in L^1_{\text{loc}}(U)$ such that

$$\int_U u(x)\partial^\alpha \varphi(x)\, dx = (-1)^{|\alpha|} \int_U v_\alpha(x)\varphi(x)\, dx$$

for any $\varphi \in \mathscr{C}^\infty_c(U)$. We write $v_\alpha = \partial^\alpha u$. When $u \in \mathscr{C}^k(U)$, then $\partial^\alpha u$ coincides with the usual partial derivative of u. In the sequel, when no regularity of $u \in L^1_{\text{loc}}(U)$ is assumed, then $\partial^\alpha u$ will always stand for a distributional derivative.

For any integer $k \geq 1$, we define the Sobolev space

$$H^k(U) = \Big\{u \in L^2(U) : \partial^\alpha u \in L^2(U),\ \forall \alpha \in \mathbb{N}^n, |\alpha| \leq k\Big\}.$$

Since $\mathscr{F}(\partial^\alpha u) = (2i\pi x)^\alpha \mathscr{F}u$ in $\mathscr{S}(\mathbb{R}^n)'$, and since a function belongs to $L^2(\mathbb{R}^n)$ if and only if its distributional Fourier transform does, the condition "$u \in L^2(\mathbb{R}^n)$ and $\partial^\alpha u \in L^2(\mathbb{R}^n)$" is equivalent to "$u \in L^2(\mathbb{R}^n)$ and $x^\alpha \mathscr{F}u \in L^2(\mathbb{R}^n)$". It follows that we can, equivalently, define $H^k(\mathbb{R}^n)$ as

$$H^k(\mathbb{R}^n) = \Big\{u \in L^2(\mathbb{R}^n) : \int_{\mathbb{R}^n} (1 + |\xi|^{2k})|\mathscr{F}u(\xi)|^2\, d\xi < \infty\Big\}.$$

For any real $s > 0$, we finally let

$$\begin{aligned} H^s(\mathbb{R}^n) &= \Big\{u \in L^2(\mathbb{R}^n) : \int_{\mathbb{R}^n} (1 + |\xi|^{2s})|\mathscr{F}u(\xi)|^2\, d\xi < \infty\Big\} \\ &= \Big\{u \in L^2(\mathbb{R}^n) : (1 + |\xi|^2)^{s/2}\mathscr{F}u(\xi) \in L^2(\mathbb{R}^n)\Big\}. \end{aligned}$$

For any $s > 0$, the set $H^s(\mathbb{R}^n)$ is a Hilbert space, equipped with the inner product

$$(u, v)_{H^s(\mathbb{R}^n)} = \int_{\mathbb{R}^n} (1 + |\xi|^2)^s \mathscr{F}u(\xi)\overline{\mathscr{F}v(\xi)}\, d\xi.$$

Sobolev spaces are particularly approriate to study the regularity of distributional solutions u to the Laplace equation $-\Delta u = f$ when u and f satisfy mild assumptions, described in the next definition.

Definition 2.4.6 (Weak Laplacian)**.** *Let $n \geq 1$, and $\Omega \subset \mathbb{R}^n$ be an open and bounded set. Let $g \in L^2(\Omega)$, and $u \in H^1(\Omega)$. We say that u satisfies $-\Delta u = f$ in the weak sense in Ω if*

$$\int_\Omega \langle \nabla u(x), \nabla \varphi(x)\rangle\, dx = \int_\Omega f(x)\varphi(x)\, dx$$

for any $\varphi \in \mathscr{C}^\infty_c(\Omega)$.

The following proposition is Theorem 2 of §6.3.1 in [31], and will play a crucial role in our proofs.

Proposition 2.4.7 (Elliptic regularity, I)**.** *Let $n \geq 1$, and $\Omega \subset \mathbb{R}^n$ be an open and bounded set. Let $g \in H^k(\Omega)$ for some $k \in \mathbb{N}$, and $u \in H^1(\Omega)$ be such that $-\Delta u = g$ in the weak sense in Ω. Then $u \in H^{k+2}_{\mathrm{loc}}(\Omega)$, i.e $u \in H^{k+2}(V)$ for any open subset $V \subset \Omega$ such that $\overline{V} \subset \Omega$. In particular, u admits weak derivatives of order $k+2$ in Ω, and we have $-\Delta u = f$ almost everywhere in Ω, where $\Delta u = \sum_{i=1}^n \partial_i^2 u$ and $\partial_1^2 u, \ldots, \partial_n^2 u$ are weak derivatives of u.*

The following proposition is Corollary 2.17 in [33]. It will also play an important role in our proofs.

Proposition 2.4.8 (Elliptic regularity, II)**.** *Let $n \geq 1$, and B_1 be the open unit ball of $\mathbb{R}^n$. Let $f \in \mathscr{C}^{k,\alpha}(B_1)$ for some $\alpha \in (0,1)$ and $k \in \mathbb{N}$, and let $u \in H^1(B_1) \cap L^\infty(B_1)$ be such that $-\Delta u = f$ in the weak sense in B_1. Then $u \in \mathscr{C}^{k+2,\alpha}(B_1)$.*

We will use a direct generalization of this proposition, stated in the next corollary. We prove it in Appendix 2.11

Corollary 2.4.9. *Let $n \geq 1$, and $\Omega \subset \mathbb{R}^n$ be an open subset. Let $f \in \mathscr{C}^{k,\alpha}_{\mathrm{loc}}(\Omega)$ for some $\alpha \in (0,1)$ and $k \in \mathbb{N}$, and let $u \in H^1_{\mathrm{loc}}(\Omega) \cap L^\infty_{\mathrm{loc}}(\Omega)$ be such that $-\Delta u = f$ in the weak sense in Ω. Then $u \in \mathscr{C}^{k+2,\alpha}_{\mathrm{loc}}(\Omega)$.*

2.5 Introduction to fractional Laplacians

Different definitions of fractional Laplacians exist. Some rely on Fourier transform, others on singular integrals, or Sobolev spaces. They all coincide for functions with enough regularity, such as the Schwarz class, but may differ in general, or at least be defined over different domains. In this section, we provide a self-contained introduction to fractional Laplacians. The approach we present is based on Fourier transforms, because it appears under this form in our proofs. The main references we used for this section are [59], [78], [94], and [95].

Let us fix $u \in \mathscr{S}(\mathbb{R}^n)$. Recalling that

$$\mathscr{F}((-\Delta)^\ell u) = (2\pi|\xi|)^{2\ell} \mathscr{F} u$$

for any integer $\ell \geq 0$, we naturally let

$$((-\Delta)^s u)(x) := (2\pi)^{2s} \mathscr{F}^{-1}(|\xi|^{2s} \mathscr{F} u(\xi))(x)$$

for any real $s > 0$ and $x \in \mathbb{R}^n$. We now make a comment on the factor $(2\pi)^{2s}$ in the definition of $(-\Delta)^s u$. Obivously, it is a consequence of our choice of normalization in the definition of the Fourier transform. For another normalization in the Fourier transform, $\mathscr{F}_{a,b}$ say, defined by

$$(\mathscr{F}_{a,b} u)(\xi) := \frac{1}{b} \int_{\mathbb{R}^n} u(x) e^{-ia(x,\xi)}\, dx$$

for some $a > 0$ and $b > 0$, we have

$$(\mathscr{F}_{a,b}^{-1} u)(\xi) = b\Big(\frac{a}{2\pi}\Big)^n \int_{\mathbb{R}^n} u(x) e^{ia(x,\xi)}\, dx.$$

It is easy to show that

$$a^{2s}\mathscr{F}_{a,b}^{-1}(|\xi|^{2s}\mathscr{F}_{a,b}u(\xi))(x) = \mathscr{F}_{1,1}^{-1}(|\xi|^{2s}\mathscr{F}_{1,1}u(\xi))(x)$$

for any $x \in \mathbb{R}^n$. It follows that any choice of normalization in the Fourier transform leads to the same value of $(-\Delta)^s u$ if we let

$$((-\Delta)^s u)(x) = a^{2s}\mathscr{F}_{a,b}^{-1}(|\xi|^{2s}\mathscr{F}_{a,b}u(\xi))(x).$$

In the sequel, we will be working with $a = 2\pi$ and $b = 1$.

When $s = n + \sigma$, with $n \in \mathbb{N}$ and $\sigma \in (0,1)$, taking the Fourier transform readily implies that

$$(-\Delta)^s u = (-\Delta)^\sigma((-\Delta)^n u),$$

where $(-\Delta)^n$ is the usual differential operator $-\Delta$ taken n times. Let us therefore restrict to $s \in (0,1)$. It is proved in [78] (see Proposition 3.3) that, in this case, we have

$$((-\Delta)^s u)(x) = c_{n,s} \lim_{\eta\to 0}\int_{\mathbb{R}^n\setminus B_\eta(x)} \frac{u(x)-u(z)}{|x-z|^{n+2s}}\,dz,$$

for any $x \in \mathbb{R}^n$, and some constant $c_{n,s}$ that only depends on n and s. Note that in the reference mentioned above, the normalization in the Fourier transform corresponds to $a = 1$ and $b = (2\pi)^{\frac{n}{2}}$ in our previous discussion. The value of the constant $c_{n,s}$ can be found in [95] (see Theorem 1), and is given by

$$c_{n,s} = \frac{s(1-s)4^s\Gamma(n/2+s)}{|\Gamma(2-s)|\pi^{n/2}}.$$

We will now explain how one can extend the domain of $(-\Delta)^s$. It is easy to see that

$$\int_{\mathbb{R}^n}((-\Delta)^s u)(x)v(x)\,dx = \int_{\mathbb{R}^n} u(x)((-\Delta)^s v)(x)\,dx$$

for any $u, v \in \mathscr{S}(\mathbb{R}^n)$. Consequently, it is tempting to define the fractional Laplacian $(-\Delta)^s T$ of an arbitrary tempered distribution $T \in \mathscr{S}(\mathbb{R}^n)'$ (recall that Fourier transforms are involved in the definition of $(-\Delta)^s$) by letting

$$\langle(-\Delta)^s T, \psi\rangle := \langle T, (-\Delta)^s\psi\rangle \tag{2.5.2}$$

for any $\psi \in \mathscr{S}(\mathbb{R}^n)$. However, this approach must be discarded because $(-\Delta)^s\psi$ does not belong to $\mathscr{S}(\mathbb{R}^n)$ in general. The regularity of $(-\Delta)^s\psi$ is established in the next proposition, which is stated in [94] but not proved. For the sake of completeness, we provide a proof of this proposition in Appendix 2.11.

Proposition 2.5.1. *Let $n \geq 1$, $s \in (0,1)$, and $u \in \mathscr{S}(\mathbb{R}^n)$. Then, $(-\Delta)^s u \in \mathscr{C}^\infty(\mathbb{R}^n)$, and we have*

$$\sup_{x\in\mathbb{R}^n} |(1+|x|^{n+2s})\partial^\alpha((-\Delta)^s u)(x)| < \infty \tag{2.5.3}$$

for any $\alpha \in \mathbb{N}^n$. In addition, for any $\alpha \in \mathbb{N}^n$ we have

$$\begin{aligned}\sup_{x\in\mathbb{R}^n} &|(1+|x|^{n+2s})\partial^\alpha((-\Delta)^s u)(x)| \\ &\lesssim |\partial^\alpha u|_{L^1(\mathbb{R}^n)} + \sup_{z\in\mathbb{R}^n}\Big((1+|z|)^{n+2}|\nabla^2(\partial^\alpha u)(z)|\Big),\end{aligned} \tag{2.5.4}$$

where, for any smooth function ψ, we let $|\nabla^2\psi(z)|$ stand for the operator norm of the Hessian matrix $\nabla^2\psi(z)$ of ψ at z.

According to the previous result, the space

$$\mathscr{S}_s(\mathbb{R}^n) := \{\psi \in \mathscr{C}^\infty(\mathbb{R}^n) : \sup_{x\in\mathbb{R}^n} |(1+|x|^{n+2s})\partial^\alpha\psi(x)| < \infty,\ \forall\alpha \in \mathbb{N}^n\}$$

is more adequate for defining fractional Laplacians by duality. For any $k \geq 1$, we similarly define $\mathscr{S}_s(\mathbb{R}^n, \mathbb{C}^k)$ as the collection of vector fields $\Psi = (\psi_1, \ldots, \psi_k)$ for which $\psi_i \in \mathscr{S}_s(\mathbb{R}^n)$ for any $i = 1, \ldots, k$. We endow $\mathscr{S}_s(\mathbb{R}^n)$ with the following convergence: a sequence $(\psi_k) \subset \mathscr{S}_s(\mathbb{R}^n)$ converges to $\psi \in \mathscr{S}_s(\mathbb{R}^n)$ in the space $\mathscr{S}_s(\mathbb{R}^n)$ if

$$\sup_{x\in\mathbb{R}^n} |(1+|x|^{n+2s})\partial^\alpha(\psi_k - \psi)(x)| \to 0$$

as $k \to \infty$ for any $\alpha \in \mathbb{N}^n$.

Definition 2.5.2 (Distributions for the fractional Laplacian)**.** *Let $s \in (0,1)$. We let $\mathscr{S}_s(\mathbb{R}^n)'$ be the set of linear maps*

$$T : \mathscr{S}_s(\mathbb{R}^n) \to \mathbb{C},\ \ \psi \mapsto \langle T, \psi\rangle$$

which are continuous with respect to the convergence on $\mathscr{S}_s(\mathbb{R}^n)$.

Similarly to tempered distributions, we let $\mathscr{S}_s^k(\mathbb{R}^n)'$ denote the space $(\mathscr{S}_s(\mathbb{R}^n)')^k$ for any $k \in \mathbb{N}$ with $k \geq 1$; see Definition 2.4.3 and the comments below.

Since $\mathscr{S}(\mathbb{R}^n) \subset \mathscr{S}_s(\mathbb{R}^n)$, we have $\mathscr{S}_s(\mathbb{R}^n)' \subset \mathscr{S}(\mathbb{R}^n)'$. In particular, distributions in $\mathscr{S}_s(\mathbb{R}^n)'$ are more regular than those from $\mathscr{S}(\mathbb{R}^n)'$.

Proposition 2.5.1 entails that if a sequence $(\psi_k) \subset \mathscr{S}(\mathbb{R}^n)$ converges to 0 in the space $\mathscr{S}(\mathbb{R}^n)$, then the sequence $((-\Delta)^s\psi_k)$ converges to 0 in the space $\mathscr{S}_s(\mathbb{R}^n)$ as $k \to \infty$. Therefore, if $(-\Delta)^sT$ is defined according to (2.5.2), then $(-\Delta)^sT$ is a tempered distribution for any $T \in \mathscr{S}_s(\mathbb{R}^n)'$.

Definition 2.5.3 (Fractional Laplacian)**.** *Let $s \in (0,1)$. For any $T \in \mathscr{S}_s(\mathbb{R}^n)'$, we let $(-\Delta)^sT : \mathscr{S}(\mathbb{R}^n) \to \mathbb{C}$ be the linear map defined by*

$$\langle(-\Delta)^sT, \psi\rangle := \langle T, (-\Delta)^s\psi\rangle$$

for any $\psi \in \mathscr{S}(\mathbb{R}^n)$. We have $(-\Delta)^sT \in \mathscr{S}(\mathbb{R}^n)'$.

The class $\mathscr{S}_s(\mathbb{R}^n)$ is obviously closed under differentiation. However, it is not closed under Fourier transform. Consequently, the space $\mathscr{S}_s(\mathbb{R}^n)'$ is closed under differentiation, but not under Fourier transform. In addition, it is easy to see that

$$\partial^\alpha(-\Delta)^s u = (-\Delta)^s\partial^\alpha u$$

for any $u \in \mathscr{S}(\mathbb{R}^n)$ and $\alpha \in \mathbb{N}^n$. This implies that

$$\partial^\alpha(-\Delta)^sT = (-\Delta)^s\partial^\alpha T$$

for any $T \in \mathscr{S}_s(\mathbb{R}^n)'$ and $\alpha \in \mathbb{N}^n$.

Let us now give examples of tempered distributions that also belong to $\mathscr{S}_s(\mathbb{R}^n)'$. It is trivial to see that any (signed) measure μ such that

$$\int_{\mathbb{R}^n} \frac{1}{1+|x|^{n+2s}}\, d|\mu|(x) < \infty$$

belongs to $\mathscr{S}_s(\mathbb{R}^n)'$. In particular, any Borel probability measure over $\mathbb{R}^n$ belongs to $\mathscr{S}_s(\mathbb{R}^n)'$. This also entails that any function $u : \mathbb{R}^n \to \mathbb{C}$ belongs to $\mathscr{S}_s(\mathbb{R}^n)'$, provided that

$$\int_{\mathbb{R}^n} \frac{|u(x)|}{1+|x|^{n+2s}}\, dx < \infty.$$

In particular, we have $L^p(\mathbb{R}^n) \subset \mathscr{S}_s(\mathbb{R}^n)'$ for any $p \in [1, +\infty]$.

We provide a result that allows one to compute $(-\Delta)^s u$ explicitly. This result is proved in Appendix 2.11.

Proposition 2.5.4. *Let $s \in (0,1)$, and $u \in \mathscr{S}_s(\mathbb{R}^n)'$. Then, we have the following :*

1. *If $u \in H^{2s}(\mathbb{R}^n)$, then $(-\Delta)^s u \in L^2(\mathbb{R}^n)$, and we have*

$$(-\Delta)^s u = (2\pi)^{2s} \mathscr{F}^{-1}(|\xi|^{2s} \mathscr{F} u)$$

   in $L^2(\mathbb{R}^n)$;

2. *If $\mathscr{F} u \in L^1_{\text{loc}}(\mathbb{R}^n)$ and $|\xi|^{2s} \mathscr{F} u(\xi) \in L^1(\mathbb{R}^n)$, then $(-\Delta)^s u \in \mathscr{C}_0(\mathbb{R}^n)$; in this case, we have*

$$((-\Delta)^s u)(x) = (2\pi)^{2s} \mathscr{F}^{-1}(|\xi|^{2s} \mathscr{F} u)(x)$$

   for any $x \in \mathbb{R}^n$;

3. *If, for some open subset $\Omega \subset \mathbb{R}^n$ and $\alpha \in (0, 2-2s)$, we have $u \in \mathscr{C}^{k,\beta}(\Omega)$ with $k = \lfloor 2s + \alpha \rfloor$ and $\beta = 2s + \alpha - k$, then $(-\Delta)^s u \in \mathscr{C}^0(\Omega)$; in this case, we have*

$$(-\Delta)^s(x) = c_{n,s} \lim_{\eta \to 0} \int_{\mathbb{R}^n \setminus B_\eta(x)} \frac{u(x) - u(z)}{|x-z|^{n+2s}}\, dz$$

   for any $x \in \Omega$.

2.6 Recovering a probability measure from its geometric rank

Theorem 2.6.1. *Let $n \geq 1$, P be a Borel probability measure over $\mathbb{R}^n$, and assume that P admits a density $f_P \in \mathscr{S}(\mathbb{R}^n)$ with respect to the Lebesgue measure. Then $R_P \in \mathscr{C}^\infty(\mathbb{R}^n) \cap D(\mathscr{L}_n)$, and we have*

$$f_P(x) = (\mathscr{L}_n R_P)(x)$$

for any $x \in \mathbb{R}^n$ (see Definition 2.2.3 for the definition of $\mathscr{L}_n$ and $D(\mathscr{L}_n)$).

Since $\mathscr{L}_1 = \frac{1}{2}\frac{d}{dx}$ and $R_P = 2F_P - 1$ over $\mathbb{R}$ when $n = 1$, where F_P is the usual cumulative distribution function of P, we recover the well-known fact

$$F_P'(x) = f_P(x).$$

The proof of Theorem 2.6.1 requires the next two lemmas, which we prove in Appendix 2.11.

Lemma 2.6.2. *Let $n \geq 2$ be an integer, and $\alpha \in (0, n)$ be a real number. Then the Fourier transform of the tempered distribution $1/|x|^\alpha$ is given by*

$$\mathscr{F}\Big(\frac{1}{|x|^\alpha}\Big)(\xi) = \frac{\Gamma(\frac{n-\alpha}{2})}{\pi^{\frac{n}{2}-\alpha}\Gamma(\frac{\alpha}{2})}\frac{1}{|\xi|^{n-\alpha}}.$$

Lemma 2.6.3. *The derivative of the tempered distribution $1/|x|^{n-1}$ is given by*

$$\nabla\Big(\frac{1}{|x|^{n-1}}\Big) = -(n-1)\ \text{P.V.}\Big(\frac{x}{|x|^{n+1}}\Big)$$

in $\mathscr{S}^n(\mathbb{R}^n)'$.

We are now able to give the proof of Theorem 2.6.1.

Proof of Theorem 2.6.1. Let us first recall that

$$R_P(x) = (K * f_P)(x)$$

for any $x \in \mathbb{R}^n$, where K is the kernel introduced in Section 2.2. Since $R_P \in L^\infty(\mathbb{R}^n, \mathbb{R}^n)$, we have $R_P \in \mathscr{S}^n_{1/2}(\mathbb{R}^n)' \subset \mathscr{S}^n(\mathbb{R}^n)'$ (see Section 2.5) so that $R_P \in D(\mathscr{L}_n)$, the domain of $\mathscr{L}_n$, irrespective of n. Since $K \in \mathscr{S}^n(\mathbb{R}^n)'$ and $f_P \in \mathscr{S}(\mathbb{R}^n)$, Proposition 2.4.5 entails that $R_P \in \mathscr{C}^\infty(\mathbb{R}^n)$.

For $n = 1$, recall that $R_P = 2F_P - 1$, where

$$F_P(x) = \int_{-\infty}^{x} f_P(t)\, dt$$

is the cumulative distribution function of P. By the fundamental theorem of calculus, we then have

$$(\mathscr{L}_n R_P)(x) = \gamma_n\ (-\Delta)^{\frac{n-1}{2}}(\nabla \cdot R_P)(x) = \frac{1}{2}R_P'(x) = \frac{1}{2}(2F_P - 1)' = f_P(x).$$

Therefore, the claim is proved when $n = 1$.

Now assume that $n \geq 2$. Since $K \in L^\infty(\mathbb{R}^n, \mathbb{R}^n)$, we have $K \in \mathscr{S}^n(\mathbb{R}^n)'$. It follows from Proposition 2.4.5 that

$$\mathscr{F}(R_P) = \mathscr{F}(K)\ \widehat{f_P} \tag{2.6.5}$$

in $\mathscr{S}^n(\mathbb{R}^n)'$, since $f_P \in \mathscr{S}(\mathbb{R}^n)$. Lemma 2.6.2 and the fact that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ yield

$$\mathscr{F}\Big(\frac{1}{|x|}\Big)(\xi) = \frac{\Gamma(\frac{n-1}{2})}{\pi^{\frac{n-1}{2}}}\frac{1}{|\xi|^{n-1}}$$

in $\mathscr{S}(\mathbb{R}^n)'$. From the identities stated before Proposition 2.4.5, we deduce that

$$\mathscr{F}(K) = \mathscr{F}\Big(\frac{x}{|x|}\Big) = -\frac{1}{2i\pi}\nabla\mathscr{F}\Big(\frac{1}{|x|}\Big) = -\frac{1}{2i\pi}\frac{\Gamma(\frac{n-1}{2})}{\pi^{\frac{n-1}{2}}}\nabla\Big(\frac{1}{|\xi|^{n-1}}\Big)$$

in $\mathscr{S}^n(\mathbb{R}^n)'$. Recalling that $x\Gamma(x) = \Gamma(x+1)$ for every $x > 0$, Lemma 2.6.3 yields

$$(\mathscr{F}K)(\xi) = \frac{\Gamma(\frac{n+1}{2})}{i\pi^{\frac{n+1}{2}}}\ \text{P.V.}\Big(\frac{\xi}{|\xi|^{n+1}}\Big) \tag{2.6.6}$$

in $\mathscr{S}^n(\mathbb{R}^n)'$. Equation (2.6.5) then rewrites

$$(\mathscr{F}R_P)(\xi) = \frac{\Gamma(\frac{n+1}{2})}{i\pi^{\frac{n+1}{2}}}\ \text{P.V.}\Big(\frac{\xi}{|\xi|^{n+1}}\Big)\widehat{f_P}(\xi)$$

in $\mathscr{S}^n(\mathbb{R}^n)'$. It is easy to see that

$$\Big(\xi,\ \text{P.V.}(\frac{\xi}{|\xi|^{n+1}})\Big) := \sum_{i=1}^{n}\xi_i\ \text{P.V.}(\frac{\xi_i}{|\xi|^{n+1}}) = \frac{1}{|\xi|^{n-1}}$$

in $\mathscr{S}(\mathbb{R}^n)'$. Further note that

$$\big(\xi,\ (\mathscr{F}R_P)(\xi)\big) = \sum_{i=1}^{n}\xi_i\mathscr{F}((R_P)_i)(\xi) = \frac{1}{2i\pi}\sum_{i=1}^{n}\mathscr{F}(\partial_i(R_P)_i)(\xi) = \frac{1}{2i\pi}\mathscr{F}(\nabla\cdot R_P)(\xi)$$

in $\mathscr{S}(\mathbb{R}^n)'$. It follows that

$$\frac{1}{2i\pi}\mathscr{F}(\nabla\cdot R_P)(\xi) = \frac{\Gamma(\frac{n+1}{2})}{i\pi^{\frac{n+1}{2}}}\frac{1}{|\xi|^{n-1}}\widehat{f_P}(\xi) \tag{2.6.7}$$

in $\mathscr{S}(\mathbb{R}^n)'$. Let us now consider two cases. (A) Assume that $n \geq 3$ is odd. Therefore, $\frac{n-1}{2} \in \mathbb{N}$ and we have

$$\widehat{f_P}(\xi) = \frac{1}{2}\frac{\pi^{\frac{n-1}{2}}}{\Gamma(\frac{n+1}{2})}|\xi|^{n-1}\mathscr{F}(\nabla\cdot R_P)(\xi) = \gamma_n\ \mathscr{F}((-\Delta)^{\frac{n-1}{2}}\nabla\cdot R_P)(\xi)$$

in $\mathscr{S}(\mathbb{R}^n)'$, where

$$\gamma_n = \frac{1}{2^n\pi^{\frac{n-1}{2}}\Gamma(\frac{n+1}{2})}.$$

Therefore, the equality

$$f_P = \gamma_n\ (-\Delta)^{\frac{n-1}{2}}(\nabla\cdot R_P) = \mathscr{L}_n(R_P)$$

holds in $\mathscr{S}(\mathbb{R}^n)'$. But $R_P \in \mathscr{C}^\infty(\mathbb{R}^n)$, which ensures that the r.h.s. of the last equality is a proper continuous function. Since f_P is also continuous and equality holds in the sense of distributions, equality also holds pointwise. (B) Assume that $n \geq 2$ is even. Since $n-2$ is even, we deduce from (2.6.7) that

$$\frac{\widehat{f_P}(\xi)}{|\xi|} = \frac{1}{2}\frac{\pi^{\frac{n-1}{2}}}{\Gamma(\frac{n+1}{2})}|\xi|^{n-2}\mathscr{F}(\nabla\cdot R_P)(\xi) = \frac{1}{2}\frac{\pi^{\frac{n-1}{2}}}{\Gamma(\frac{n+1}{2})}\frac{1}{(2\pi)^{n-2}}\mathscr{F}((-\Delta)^{\frac{n-2}{2}}(\nabla\cdot R_P))(\xi)$$

holds in $\mathscr{S}(\mathbb{R}^n)'$. Let us recall that $R_P \in \mathscr{S}^n_{1/2}(\mathbb{R}^n)'$. Since $\mathscr{S}^n_{1/2}(\mathbb{R}^n)'$ is closed with respect to differentiation, we have $u \in \mathscr{S}_{1/2}(\mathbb{R}^n)'$, with $u := (-\Delta)^{\frac{n-2}{2}}(\nabla\cdot R_P)$. It is clear that $\mathscr{F}u \in L^1_{\text{loc}}(\mathbb{R}^n)$ since $\widehat{f_P}(\xi)/|\xi| \in L^1_{\text{loc}}(\mathbb{R}^n)$ (recall that $n \geq 2$), and that $|\xi|\mathscr{F}u(\xi) \in L^1(\mathbb{R}^n)$ since $f_P \in \mathscr{S}(\mathbb{R}^n)$. It follows from Proposition 2.5.4 that $(-\Delta)^{\frac{1}{2}}(-\Delta)^{\frac{n-2}{2}}(\nabla \cdot R_P) \in \mathscr{C}_0(\mathbb{R}^n)$ and that

$$\widehat{f_P} = \gamma_n\ \mathscr{F}((-\Delta)^{\frac{1}{2}}(-\Delta)^{\frac{n-2}{2}}(\nabla \cdot R_P))$$

holds in $\mathscr{S}(\mathbb{R}^n)'$, where γ_n is the same constant as in (A). We deduce that

$$f_P = \gamma_n\ (-\Delta)^{\frac{1}{2}}(-\Delta)^{\frac{n-2}{2}}(\nabla \cdot R_P) = \mathscr{L}_n(R_P)$$

in $\mathscr{S}(\mathbb{R}^n)'$. Since both sides of this last equality are continuous, equality also holds pointwise over $\mathbb{R}^n$, which concludes the proof. ■

Definition 2.6.4 (Characteristic function)**.** *Let $n \geq 1$, and P be a Borel probability measure over $\mathbb{R}^n$. For any $\xi \in \mathbb{R}^n$, let*

$$\phi_P(\xi) := \mathbb{E}[e^{-2i\pi(\xi,X)}]$$

be the characteristic function of P, where X is a random n-vector with law P. Then, ϕ_P is the distributional Fourier transform of the tempered distribution P, i.e. $\phi_P = \mathscr{F}(P)$ in $\mathscr{S}(\mathbb{R}^n)'$ (see Section 2.4.1 for the definition and some properties of the space $\mathscr{S}(\mathbb{R}^n)'$ of tempered distributions). Letting $\varphi_P(\xi) = \mathbb{E}[e^{i\pi(\xi,X)}]$ denote the usual characteristic function of P, we have

$$\phi_P(\xi) = \varphi_P(-2\pi\xi)$$

for any $\xi \in \mathbb{R}^n$.

Theorem 2.6.5. *Let $n \geq 1$, and P be a Borel probability measure over $\mathbb{R}^n$. Then R_P belongs to the domain $D(\mathscr{L}_n)$ of $\mathscr{L}_n$. Furthermore, the distribution $\mathscr{F}(\mathscr{L}_n R_P)$ is a continous function over $\mathbb{R}^n$, and we have*

$$\phi_P(\xi) = \mathscr{F}(\mathscr{L}_n R_P)(\xi)$$

for any $\xi \in \mathbb{R}^n$. In other words, the equality

$$P = \mathscr{L}_n(R_P)$$

holds in $\mathscr{S}(\mathbb{R}^n)'$.

The following lemma, which we need to prove Theorem 2.6.5, is stated in [6], Corollary 2.2.10.

Lemma 2.6.6. *Let Q and $(Q_k)_{k\geq 1}$ be Borel probability measures over $\mathbb{R}^n$ such that Q_k converges to Q in distribution as $k \to \infty$. Let $g : \mathbb{R}^n \to \mathbb{C}$ be a bounded and measurable map such that $Q(D_g) = 0$, where we let*

$$D_g := \{x \in \mathbb{R}^n : g \text{ is not continuous at } x\}.$$

Then $\int_{\mathbb{R}^n} g\, dQ_k \to \int_{\mathbb{R}^n} g\, dQ$ as $k \to \infty$.

Proof of Theorem 2.6.5. Assume first that there exists a sequence of probability measures (Q_k) over $\mathbb{R}^n$ such that (Q_k) converges in distribution (i.e. in law) to P as $k \to \infty$ and such that, for any k, Q_k admits a density $f_k \in \mathscr{S}(\mathbb{R}^n)$ with respect to the Lebesgue measure. For any k, let us denote R_{Q_k} the geometric rank associated to the probability measure Q_k. Since, for any k, Q_k admits the density $f_k \in \mathscr{S}(\mathbb{R}^n)$, Theorem 2.6.1 entails that

$$f_k(x) = (\mathscr{L}_n R_{Q_k})(x)$$

for any $x \in \mathbb{R}^n$ and k. Let $\psi \in \mathscr{S}(\mathbb{R}^n)$. For any k, we then have

$$\int_{\mathbb{R}^n} \psi(x) f_k(x)\, dx = \int_{\mathbb{R}^n} (R_{Q_k}, \mathscr{L}_n^*(\psi))\, dx. \tag{2.6.8}$$

We are going to show that the l.h.s. of (2.6.8) converges to $\int_{\mathbb{R}^n} \psi(x)\, dP(x)$ and that the r.h.s. of (2.6.8) converges to $\int_{\mathbb{R}^n} (R_P, \mathscr{L}_n^*(\psi))\, dx$ as $k \to \infty$.

In order to show that the l.h.s. of (2.6.8) converges, it is enough to observe that, since (Q_k) converges in distribution to P and ψ is continuous and bounded over $\mathbb{R}^n$, we have

$$\int_{\mathbb{R}^n} \psi(x) f_k(x)\, dx = \int_{\mathbb{R}^n} \psi(x)\, dQ_k(x) \to \int_{\mathbb{R}^n} \psi(x)\, dP(x) \tag{2.6.9}$$

as $k \to \infty$. Let us now show that the r.h.s. of (2.6.8) converges. Let us start by showing that R_{Q_k} converges almost everywhere to R_P. For any $x \in \mathbb{R}^n$, let $g_x(z) := \frac{x-z}{\|x-z\|}\mathbb{I}[z \neq x]$ for any $z \in \mathbb{R}^n$. With the notations of Lemma 2.6.6, we have $D_{g_x} = \{x\}$. Let

$$A := \{x \in \mathbb{R}^n : P[\{x\}] > 0\}.$$

Then A is at most countable and we have $P[D_{g_x}] = 0$ for all $x \in \mathbb{R}^n \setminus A$. Since g_x is bounded and measurable for all $x \in \mathbb{R}^n$, Lemma 2.6.6 entails that

$$R_{Q_k}(x) = \int_{\mathbb{R}^n} g_x(z)\, dQ_k(z) \to \int_{\mathbb{R}^n} g_x(z)\, dP(z) = R_P(x)$$

for any $x \in \mathbb{R}^n \setminus A$ as $k \to \infty$. Since A is at most countable, we have $R_{Q_k} \to R_P$ almost everywhere. In order to apply the dominated convergence theorem to the r.h.s. of (2.6.8), observe that $\mathscr{L}_n^*(\psi) \in L^1(\mathbb{R}^n)$. Indeed, if n is even, we have $(-\Delta)^{\frac{n-2}{2}}\psi \in \mathscr{S}(\mathbb{R}^n)$ so that $(-\Delta)^{\frac{1}{2}}((-\Delta)^{\frac{n-2}{2}}\psi) \in \mathscr{S}_{1/2}(\mathbb{R}^n)$, whence

$$\mathscr{L}_n^*(\psi) = \nabla((-\Delta)^{\frac{1}{2}}(-\Delta)^{\frac{n-2}{2}}\psi) \in \mathscr{S}_{1/2}(\mathbb{R}^n, \mathbb{C}^n) \subset L^1(\mathbb{R}^n)$$

since $\psi \in \mathscr{S}(\mathbb{R}^n)$. If n is odd, then $\nabla\big((-\Delta)^{\frac{n-1}{2}}\psi\big)$ trivially belongs to $\mathscr{S}(\mathbb{R}^n, \mathbb{C}^n)$ which is a subset of $L^1(\mathbb{R}^n)$. Since the sequence of functions $(R_{Q_k})_k$ is uniformly norm-bounded by 1 and since $R_{Q_k} \to R_P$ almost everywhere as $k \to \infty$, Lebesgue's dominated convergence theorem entails that

$$\int_{\mathbb{R}^n} (R_{Q_k}, \mathscr{L}_n^*(\psi))\, dx \to \int_{\mathbb{R}^n} (R_P, \mathscr{L}_n^*(\psi))\, dx \tag{2.6.10}$$

as $k \to \infty$. Putting (2.6.8), (2.6.9) and (2.6.10) together yields

$$\int_{\mathbb{R}^n} \psi(x)\, dP(x) = \int_{\mathbb{R}^n} (R_P, \mathscr{L}_n^*(\psi))\, dx$$

for any $\psi \in \mathscr{S}(\mathbb{R}^n)$, which yields

$$P = \mathscr{L}_n(R_P)$$

in $\mathscr{S}(\mathbb{R}^n)'$.

It remains to show that there indeed exists a sequence of probability measures (Q_k) over $\mathbb{R}^n$ that converges in distribution to P and such that Q_k admits a density $f_k \in \mathscr{S}(\mathbb{R}^n)$ with respect to the Lebesgue measure. Let X be a random vector with law P. Let $\rho \in \mathscr{C}_c^\infty(\mathbb{R}^n)$ be such that $0 \le \rho \le 1$ and $\int_{\mathbb{R}^n} \rho(x)\, dx = 1$. In particular, ρ is a probability density over $\mathbb{R}^n$. Let then Y be a random vector with density ρ. For any k, let $X_k := X + \frac{1}{k} Y$. Since (X_k) converges to X in probability, then X_k converges to X in distribution. Observe that X_k admits the density $p_k := \rho_k * P$ with respect to the Lebesgue measure, where $\rho_k(x) := k^n \rho(kx)$ for any k and $x \in \mathbb{R}^n$. In particular, $p_k \in \mathscr{C}^\infty(\mathbb{R}^n)$ since $\rho_k \in \mathscr{C}_c^\infty(\mathbb{R}^n)$. Since $X_k \to X$ in distribution, we have

$$\int_{\mathbb{R}^n} g(x) p_k(x)\, dx \to \int_{\mathbb{R}^n} g(x)\, dP(x)$$

for any $g \in \mathscr{C}_b^0(\mathbb{R}^n)$ as $k \to \infty$. For any k, let $r_k > 0$ be such that

$$\int_{\mathbb{R}^n \setminus B_{r_k}} p_k(x)\, dx < \frac{1}{k}.$$

For any k, let $\chi_k \in \mathscr{C}_c^\infty(\mathbb{R}^n)$ be such that $0 \le \chi_k \le 1$ over $\mathbb{R}^n$, $\chi_k = 1$ over B_{r_k} and $\chi_k = 0$ over $\mathbb{R}^n \setminus B_{1+r_k}$. Let then $f_k(x) := \chi_k(x) p_k(x)$ for any k and $x \in \mathbb{R}^n$. Since $(p_k) \subset \mathscr{C}^\infty(\mathbb{R}^n)$, we have $(f_k) \subset \mathscr{C}_c^\infty(\mathbb{R}^n)$. In particular, $(f_k) \subset \mathscr{S}(\mathbb{R}^n)$. Let $g \in \mathscr{C}_b^0(\mathbb{R}^n)$. We have that

$$\begin{aligned}
&\Big| \int_{\mathbb{R}^n} g(x) f_k(x)\, dx - \int_{\mathbb{R}^n} g(x) p_k(x)\, dx \Big| \\
&\qquad \le \int_{\mathbb{R}^n} |g(x)| (\chi_k(x) - 1) p_k(x)\, dx \\
&\qquad \le \int_{\mathbb{R}^n \setminus B_{r_k}} g(x) p_k(x)\, dx \\
&\qquad \le \|g\|_{L^\infty(\mathbb{R}^n)} \int_{\mathbb{R}^n \setminus B_{r_k}} p_k(x)\, dx \\
&\qquad \le \frac{1}{k} \|g\|_{L^\infty(\mathbb{R}^n)}.
\end{aligned}$$

Since $\int_{\mathbb{R}^n} g(x) p_k(x)\, dx \to \int_{\mathbb{R}^n} g(x)\, dP(x)$, it follows that

$$\int_{\mathbb{R}^n} g(x) f_k(x)\, dx \to \int_{\mathbb{R}^n} g(x)\, dP(x)$$

for any $g \in \mathscr{C}_b^0(\mathbb{R}^n)$. Letting Q_k be the probability measure with density $f_k \in \mathscr{S}(\mathbb{R}^n)$ for any k yields the conclusion. ■

If P admits a density f_P with respect to the Lebesgue measure over $\mathbb{R}^n$, Theorem 2.6.5 entails that $f_P = \mathscr{L}_n R_P$ holds in the sense of tempered distributions. Consequently, it is natural to look for simple conditions that would ensure that the previous equality holds pointwise. If that is the case, one can $\mathscr{L}_n R_P$ by successively applying the differential operators involved in the definition of $\mathscr{L}_n$ to R_P, without restricting to smooth densities with rapid decay at infinity as in Theorem 2.6.1. We start with the univariate case, which is already well-known.

Theorem 2.6.7 (Univariate case)**.** *Let P be a Borel probability measure over $\mathbb{R}$. Let $\Omega \subset \mathbb{R}$ be an open subset, and assume that P is non-atomic over Ω. Then $R_P \in \mathscr{C}^0(\Omega)$. If, in addition, we assume that P admits a density $f_\Omega \in L^1(\Omega)$ over Ω with respect to the Lebesgue measure, then R_P admits a weak derivative R'_P, and we have*

$$f_\Omega(x) = \gamma_1 R'_P(x) \tag{2.6.11}$$

for almost any $x \in \Omega$. If we further assume that f_Ω is continuous over Ω, then R_P is continuously differentiable over Ω, and (2.6.11) holds pointwise over Ω, where R'_P is now the usual derivative.

In dimension $n \geq 2$, the operator $\mathscr{L}_n$ requires the computation of n derivatives when n is odd, and $n-1$ derivatives when n is even. In any case, we wish to reach a regularity of order at least $n-1$. We achieve this in the following proposition.

Proposition 2.6.8 (Intermediate regularity)**.** *Let $n \geq 2$, and P be a Borel probability measure over $\mathbb{R}^n$. Let $\Omega \subset \mathbb{R}^n$ be an open subset. We have the following :*

1. *If P is non-atomic over Ω, then $R_P \in \mathscr{C}^0(\Omega)$;*
2. *For some integer $\ell \in [1, n-1]$ and some real $p > \frac{n}{n-\ell}$, assume that P admits a density $f_\Omega \in L^1(\Omega)$ over Ω with respect to the Lebesgue measure, and that $f_\Omega \in L^p_{\text{loc}}(\Omega)$. Let Z be a random n-vector with law P. Then $R_P(x) = \mathbb{E}[K(x-Z)] \in \mathscr{C}^\ell(\Omega)$, and we have*
$$\partial^\alpha R_P(x) = \mathbb{E}[(\partial^\alpha K)(x-Z)]$$
for any $x \in \Omega$ and $\alpha \in \mathbb{N}^n$ with $|\alpha| \leq \ell$;
3. *Under the assumptions of point 2 of this proposition, if we further assume that the density $f_{\mathbb{R}^n} =: f_P$ belongs to $L^p(\mathbb{R}^n)$, then $\partial^\alpha R_P$ converges to 0 at infinity for any $\alpha \in \mathbb{N}^n$ such that $1 \leq |\alpha| \leq \ell$.*

In particular, if $f_\Omega \in L^{n+\varepsilon}_{\text{loc}}(\Omega)$ for some $\varepsilon > 0$, then $R_P \in \mathscr{C}^{n-1}(\Omega)$. If $f_\Omega \in L^{n+\varepsilon}(\mathbb{R}^n)$ for some $\varepsilon > 0$, then $R_P \in \mathscr{C}^{n-1}_b(\mathbb{R}^n)$.

Proof of Proposition 2.6.8. 1. The fact that $R_P \in \mathscr{C}^0(\Omega)$ is a direct consequence of Lebesgue's dominated convergence theorem, provided P is non-atomic over Ω.

2. We are going to prove the result by induction. By the first part of the proof, we have $R_P \in \mathscr{C}^0(\Omega)$. In addition, we trivially have that $|R_P(x)| \leq 1$ for any $x \in \mathbb{R}^n$, so that $R_P \in \mathscr{C}^0_b(\mathbb{R}^n)$.

Let $0 \le k \le \ell - 1$ and assume that $R_P \in \mathscr{C}^k(\Omega)$ with

$$\partial^\alpha R_P(x) = \mathbb{E}[(\partial^\alpha K)(x - X)] \tag{2.6.12}$$

for any $x \in \Omega$ and $\alpha \in \mathbb{N}^n$ with $|\alpha| \le k$. Let us show that $R_P \in \mathscr{C}^{k+1}(\Omega)$ (and $R_P \in \mathscr{C}_b^{k+1}(\mathbb{R}^n)$ if $f_P \in L^p(\mathbb{R}^n)$) and that (2.6.12) holds for any $\alpha \in \mathbb{N}^n$ with $|\alpha| \le k+1$. To that purpose, let $\alpha \in \mathbb{N}^n$ with $|\alpha| = k$, $x \in \Omega$ and $r > 0$ be such that $\overline{B_r(x)} \subset \Omega$. Let $j \in \{1, \dots, n\}$ and e_j be the jth vector of the canonical basis of $\mathbb{R}^n$. We are going to show that

$$\frac{\partial^\alpha R_P(x + he_j) - \partial^\alpha R_P(x)}{h} \to \mathbb{E}[(\partial_j \partial^\alpha K)(x - Z)]$$

$h \to 0$ and that the limit is continuous over Ω. Without loss of generality, let us assume that $|h| < \kappa$ for some $\kappa < d(x, \partial\Omega)$, so that $x + he_j \in B_\kappa(x) \subset \Omega$. For any h, let

$$S_h := \{x + she_j : s \in [0, 1]\}$$

be the line segment from x to $x + he_j$. Since $S_h \subset \Omega$ and P has a density over Ω, we have

$$\frac{\partial^\alpha R_P(x + he_j) - \partial^\alpha R_P(x)}{h} = \mathbb{E}\left[\frac{(\partial^\alpha K)(x + he_j - Z) - (\partial^\alpha K)(x - Z)}{h} \mathbb{I}[Z \in \mathbb{R}^n \setminus S_h]\right]$$

for any h. In order to take the limit as $h \to 0$ under the above expectation, we will show that the integrand is a uniformly P-integrable family indexed by h and converges P-almost surely as $h \to 0$. Since $K \in \mathscr{C}^\infty(\mathbb{R}^n \setminus \{0\})$, observe that

$$\frac{(\partial^\alpha K)(x + he_j - z) - (\partial^\alpha K)(x - z)}{h} = \int_0^1 (\partial_j \partial^\alpha K)(x + she_j - z)\, ds$$

for any $z \in \mathbb{R}^n \setminus S_h$. The latter obviously converges to $(\partial_j \partial^\alpha K)(x - z)$ as $h \to 0$, for any $z \in \mathbb{R}^n \setminus S_h$. Let us now show that the family of random vectors

$$\left(\int_0^1 (\partial_j \partial^\alpha K)(x + she_j - Z) \mathbb{I}[Z \in \mathbb{R}^n \setminus S_h]\, ds \right)_{|h| < \kappa}$$

is uniformly P-integrable. It is enough to show that there exists $\delta > 0$ such that

$$\sup_{|h| < \kappa} \mathbb{E}\left[\Big| \int_0^1 (\partial_j \partial^\alpha K)(x + she_j - Z)\, ds \Big|^{1+\delta} \mathbb{I}[Z \in \mathbb{R}^n \setminus S_h] \right] < \infty.$$

Let $\delta > 0$ be arbitrary for now and let us fix its value later on. Observe that $|\partial^\beta K(x)| \le C_\beta |x|^{-|\beta|}$ for any $x \in \mathbb{R}^n \setminus \{0\}$, any $\beta \in \mathbb{N}^n$ and some positive constant C_β. Therefore, there exists $C > 0$ such that

$$\begin{aligned} &\Big| \int_0^1 (\partial_j \partial^\alpha K)(x + she_j - z)\, ds \Big|^{1+\delta} \\ &\quad\le \int_0^1 |(\partial_j \partial^\alpha K)(x + she_j - z)|^{1+\delta}\, ds \\ &\quad\le C \int_0^1 \frac{1}{|x + she_j - z|^{(1+k)(1+\delta)}}\, ds \end{aligned}$$

for any $z \in \mathbb{R}^n \setminus S_h$ and h, by Jensen's inequality. It follows from Fubini's theorem that

$$\begin{aligned}
&\sup_{|h|<\kappa} \mathbb{E}\bigg[\Big|\int_0^1 (\partial_j \partial^\alpha K)(x + she_j - Z)\,ds\Big|^{1+\delta} \mathbb{I}[Z \in \mathbb{R}^n \setminus S_h]\bigg] \\
&\qquad \le C \sup_{|h|<\kappa} \int_0^1 \mathbb{E}\bigg[\frac{1}{|x + she_j - Z|^{(1+k)(1+\delta)}} \mathbb{I}[Z \in \mathbb{R}^n \setminus S_h]\bigg]\,ds \\
&\qquad \le C \sup_{|h|<\kappa} \sup_{s\in[0,1]} \mathbb{E}\bigg[\frac{1}{|x + she_j - Z|^{(1+k)(1+\delta)}} \mathbb{I}[Z \in \mathbb{R}^n \setminus S_h]\bigg].
\end{aligned}$$

Let us fix h such that $|h| < \kappa$ and $s \in [0, 1]$. We have that

$$\begin{aligned}
&\mathbb{E}\bigg[\frac{1}{|x + she_j - Z|^{(1+k)(1+\delta)}} \mathbb{I}[Z \in \mathbb{R}^n \setminus S_h]\bigg] \\
&\qquad \le \frac{1}{r^{(1+k)(1+\delta)}} + \mathbb{E}\bigg[\frac{1}{|x + she_j - Z|^{(1+k)(1+\delta)}} \mathbb{I}[Z \in B_r(x + she_j) \setminus S_h]\bigg].
\end{aligned}$$

Since $B_r(x + she_j) \setminus S_h \subset \Omega$ and P admits a density f_Ω over Ω, Hölder's inequality yields

$$\begin{aligned}
&\mathbb{E}\bigg[\frac{1}{|x + she_j - Z|^{(1+k)(1+\delta)}} \mathbb{I}[Z \in B_r(x + she_j) \setminus S_h]\bigg] \\
&\qquad = \int_{B_r(x+she_j)\setminus S_h} \frac{1}{|x + she_j - z|^{(1+k)(1+\delta)}} f_\Omega(z)\,dz \\
&\qquad \le \bigg(\int_{B_r} \frac{1}{|z|^{q(1+k)(1+\delta)}}\,dz\bigg)^{1/q} \bigg(\int_{B_r(x+she_j)} |f_\Omega(z)|^p\,dz\bigg)^{1/p},
\end{aligned}$$

where p is such that $f_\Omega \in L^p_{\mathrm{loc}}(\Omega)$ and $q = \frac{p}{p-1}$ is the conjugate exponent of p. The fact that $k \le \ell - 1$ and $p > \frac{n}{n-\ell}$ implies that $p > \frac{n}{n-(1+k)}$ and $q < \frac{n}{1+k}$. Let us therefore choose $\delta > 0$ small enough such that $q < \frac{n}{(1+k)(1+\delta)}$. In particular, we have

$$\int_{B_r} \frac{1}{|z|^{q(1+k)(1+\delta)}}\,dz < \infty.$$

Since $\overline{B_\kappa(x)} \subset \Omega$ and $f_\Omega \in L^p_{\mathrm{loc}}(\Omega)$, we also have that

$$\int_{B_r(x+she_j)} |f_\Omega(z)|^p\,dz \le \int_{B_\kappa(x)} |f_\Omega(z)|^p\,dz < \infty$$

uniformly in $|h| < \kappa$ and $s \in [0, 1]$. We deduce that

$$\sup_{|h|<\kappa} \mathbb{E}\bigg[\Big|\int_0^1 (\partial_j \partial^\alpha K)(x + she_j - Z)\,ds\Big|^{1+\delta} \mathbb{I}[Z \in \mathbb{R}^n \setminus S_h]\bigg] < \infty.$$

Therefore, the family of random vectors

$$\bigg(\int_0^1 (\partial_j \partial^\alpha K)(x + she_j - Z)\mathbb{I}[Z \in \mathbb{R}^n \setminus S_h]\,ds\bigg)_{|h|<\kappa}$$

is uniformly P-integrable. It follows from Lebesgue-Vitali's theorem that

$$\frac{\partial^\alpha R_P(x+he_j) - \partial^\alpha R_P(x)}{h} \to \mathbb{E}[(\partial_j \partial^\alpha K)(x-Z)]$$

for any $x \in \Omega$ as $h \to 0$. Let us show that $x \mapsto \mathbb{E}[(\partial_j \partial^\alpha K)(x-Z)]$ is continuous over Ω. Let $x \in \Omega$ and $(x_m) \subset \Omega$ be a sequence converging to x as $m \to \infty$. The familiy of random vectors $((\partial_j \partial^\alpha K)(x_m - Z))_{n\in\mathbb{N}}$ converges P-almost surely to $(\partial_j \partial^\alpha K)(x-Z)$ as $m \to \infty$ since P is non-atomic, and is uniformly P-integrable since

$$\sup_{m\in\mathbb{N}} \mathbb{E}[|(\partial_j \partial^\alpha K)(x_m - Z)|^{1+\eta}] \lesssim \sup_{m\in\mathbb{N}} \mathbb{E}\Big[\frac{1}{|x_m - Z|^{(1+k)(1+\eta)}}\Big] < \infty$$

for η small enough, by the previous computations. It follows that $\partial^\alpha R_P \in \mathscr{C}^1(\Omega)$ and that

$$\partial_j \partial^\alpha R_P(x) = \mathbb{E}[(\partial_j \partial^\alpha K)(x-Z)]$$

for any $x \in \mathbb{R}^n$. Since $\alpha \in \mathbb{N}^n$ with $|\alpha| = k$ was arbitrary, we deduce that $R_P \in \mathscr{C}^{k+1}(\Omega)$ and that (2.6.12) holds for any $\alpha \in \mathbb{N}^n$ with $|\alpha| \le k+1$. This proves the result, by induction.

3. The second part of the proof implies that $R_P \in \mathscr{C}^\ell(\mathbb{R}^n)$. Let $\alpha \in \mathbb{N}^n$ with $1 \le |\alpha| \le \ell$ and let us show that $\partial^\alpha R_P$ converges to 0 at infinity. For the sake of convenience, let us write $k := |\alpha|$. We have already noticed that

$$|\partial^\alpha R_P(x)| \le C\, \mathbb{E}\Big[\frac{1}{|x-Z|^k} \mathbb{I}[Z \ne x]\Big] =: C\, h(x)$$

for any $x \in \mathbb{R}^n$ and some positive constant C. Therefore, it is enough to show that h converges to 0 at infinity. Let $(x_m) \subset \mathbb{R}^n$ be such that $|x_m| \to \infty$ as $m \to \infty$. A standard application of Lebesgue's dominated convergence theorem entails that

$$\mathbb{E}\Big[\frac{1}{|x_m - Z|^k} \mathbb{I}[Z \in \mathbb{R}^n \setminus B_1(x)]\Big] \to 0$$

as $m \to \infty$. Next observe that

$$\begin{aligned}
&\Big|\mathbb{E}\Big[\frac{1}{|x_m - Z|^k} \mathbb{I}[B_1(x)]\Big]\Big| \\
&\qquad = \int_{B_1(x_m)} \frac{1}{|z - x_m|^k} f_P(z)\, dz \\
&\qquad \le C\Big(\int_{B_1(x_m)} |f_P(z)|^p\, dz\Big)^{1/p} \Big(\int_{B_1} \frac{1}{|z|^{qk}}\, dz\Big)^{1/q},
\end{aligned}$$

where $q = \frac{p}{p-1}$ is the conjugate exponent of p. Since $p > \frac{n}{n-\ell}$ and $k \le \ell$, we have $p > \frac{n}{n-k}$ whence $q < \frac{n}{k}$. In particular, $qk < n$. It follows that

$$\int_{B_1} \frac{1}{|z|^{qk}}\, dz < \infty.$$

It remains to show that $\int_{B_1(x_m)} |f_P(z)|^p \, dz \to 0$ as $m \to \infty$. Since $f_P \in L^p(\mathbb{R}^n)$, let ν be the non-negative finite measure defined by

$$\nu(B) := \int_B |f_P(z)|^p \, dz$$

for any Borel subset $B \subset \mathbb{R}^n$. We then have

$$\int_{B_1(x_m)} |f_P(z)|^p \, dz = \nu(B_1(x_m))$$

for any m. Furthermore, we have $\nu(\mathbb{R}^n \setminus B_{|x_m|-1}) \to 0$ as $m \to \infty$ since ν is finite and $|x_m| \to \infty$ as $m \to \infty$. It follows that

$$\mathbb{E}\Big[\frac{1}{|x_m - Z|^k} \mathbb{I}[B_1(x_m)]\Big] \to 0$$

as $m \to \infty$. We deduce that $\partial^\alpha R_P$ converges to 0 at infinity for any $\alpha \in \mathbb{N}^n$ with $1 \leq |\alpha| \leq \ell$, which concludes the proof. ■

To reach differentiability of order n, one cannot take $\ell = n$ formally in Theorem 2.6.8. If that were true, and if $f_P \in L^\infty(\mathbb{R}^n)$, we would have $R_P \in \mathscr{C}^n(\mathbb{R}^n)$; this would imply that $f_P \in \mathscr{C}^0(\mathbb{R}^n)$ when n is odd by Theorem 2.6.5, which will not be the case in general. However, when $f \in \mathscr{C}^0(\Omega)$ we show in the next two theorems that $R_P \in \mathscr{C}^n(\Omega)$ with $f_P = \mathscr{L}_n(R_P)$, under the additional very mild assumption that f_P belongs to some Hölder class.

Theorem 2.6.9 (Odd dimensions)**.** *Let $n \geq 3$ be odd, and P be a Borel probability measure over $\mathbb{R}^n$. Let $\Omega \subset \mathbb{R}^n$ be an open subset, and assume that P admits a density $f_\Omega \in L^1(\Omega)$ over Ω with respect to the Lebesgue measure. Then, we have the following :*

1. *If $f_\Omega \in L^p_{\text{loc}}(\Omega)$ for some $p > n$, then $R_P \in \mathscr{C}^{n-1}(\Omega) \cap H^n_{\text{loc}}(\Omega)$. In particular, R_P admits weak derivatives of order n in Ω, and we have*

$$f_\Omega(x) = (\mathscr{L}_n R_P)(x)$$

   for almost any $x \in \Omega$.

2. *If $f_\Omega \in \mathscr{C}^{k,\alpha}_{\text{loc}}(\Omega)$ for some $k \in \mathbb{N}$ and some $\alpha \in (0,1)$, then $R_P \in \mathscr{C}^{k+n}(\Omega)$, and we have*

$$f_\Omega(x) = (\mathscr{L}_n R_P)(x)$$

   for any $x \in \Omega$.

Proof of Theorem 2.6.9. 1. Let us recall that $\nabla g_P = R_P$ over Ω since P is non-atomic over Ω (see Definition 2.2.1 and the comments below). Let us also recall that

$$P = \gamma_n \, (-\Delta)^{\frac{n-1}{2}} R_P$$

in $\mathscr{S}(\mathbb{R}^n)'$ by Theorem 2.6.5. Since $\nabla \cdot R_P = \nabla \cdot (\nabla g_P) = \Delta g_P$ over Ω, we have

$$-f_\Omega = \gamma_n(-\Delta)^{\frac{n+1}{2}} g_P$$

in $\mathscr{D}(\Omega)'$. Since $f_\Omega \in L^p_{\text{loc}}(\Omega)$ with $p > n$, Proposition 2.6.8 entails that $R_P \in \mathscr{C}^{n-1}(\Omega)$. In particular, we have $g_P \in \mathscr{C}^n(\Omega)$.

We are now going to use elliptic regularity to improve on the regularity of g_P over Ω. Let us first fix an open and bounded subset $U \subset \Omega$ such that $\overline{U} \subset \Omega$. Let U_1 be an open and bounded subset such that $\overline{U} \subset U_1$ and $\overline{U_1} \subset \Omega$. Since $g_P \in \mathscr{C}^n(\Omega)$ and U_1 is bounded, we have $(-\Delta)^{\frac{n-1}{2}} g_P \in H^1(U_1)$ with

$$- f_\Omega = \gamma_n \, (-\Delta)\big((-\Delta)^{\frac{n-1}{2}} g_P\big) \tag{2.6.13}$$

in $\mathscr{D}(U_1)'$. Since $f_\Omega \in L^p_{\text{loc}}(\Omega)$ with $p > n \geq 3$, we have $f_\Omega \in L^2_{\text{loc}}(\Omega)$. In particular, $f_\Omega \in L^2(U_1)$ since U_1 is bounded. It follows from Proposition 2.4.7 that $(-\Delta)^{\frac{n-1}{2}} g_P \in H^2_{\text{loc}}(U_1)$ and that (2.6.13) holds almost everywhere. Let us fix another open and bounded subset U_2 such that $\overline{U} \subset U_2$ and $\overline{U_2} \subset U_1$. Since $g_P \in \mathscr{C}^{n-2}(\Omega)$ and U_2 is bounded, $\Delta^{\frac{n-3}{2}} g_P \in H^1(U_2)$ satisfies

$$(-\Delta)\big((-\Delta)^{\frac{n-3}{2}} g_P\big) = (-\Delta)^{\frac{n-1}{2}} g_P.$$

Since $(-\Delta)^{\frac{n-1}{2}} g_P \in H^2(U_2)$, elliptic regularity implies that $(-\Delta)^{\frac{n-3}{2}} g_P \in H^4_{\text{loc}}(U_2)$. Proceeding by induction, we construct open and bounded decreasing subsets

$$U_1 \supset U_2 \supset \ldots \supset U_{\frac{n+1}{2}} \supset U$$

such that $\overline{U} \subset U_k$ and $(-\Delta)^{\frac{n+1-2k}{2}} g_P \in H^{2k}_{\text{loc}}(U_k)$ for any $k = 1, \ldots \frac{n+1}{2}$. For $k = \frac{n+1}{2}$, we find that $g_P \in H^{n+1}_{\text{loc}}(U_{\frac{n+1}{2}})$. In particular, we have $g_P \in H^{n+1}(U)$. Since U was arbitrary in the first place, we conclude that $g_P \in H^{n+1}_{\text{loc}}(\Omega)$. In particular, $R_P \in H^n_{\text{loc}}(\Omega)$.

2. The second part of the statement is proved similarly to the first part by replacing Proposition 2.4.7 by Proposition 2.4.8 and Corollary 2.4.9. Applying the same bootstrap method, we only need to prove that $(-\Delta)^{\frac{n+1-2k}{2}} g_P \in L^\infty_{\text{loc}}(\Omega)$ for any $k = 1, \ldots, \frac{n+1}{2}$ in order to apply Corollary 2.4.9. But this immediately follows from the fact that $g_P \in \mathscr{C}^n(\Omega)$ since $R_P \in \mathscr{C}^{n-1}(\Omega)$ and $R_P = \nabla g_P$ (see Definition 2.2.2). ∎

Theorem 2.6.10 (Even dimensions). *Let $n \geq 2$ be even, and P a Borel probability measure on $\mathbb{R}^n$. Assume that P admits a density $f_P \in L^1(\mathbb{R}^n)$ with respect to the Lebesgue measure. Let us define*

$$R_P^{(n-1)} := \gamma_n \, (-\Delta)^{\frac{n-2}{2}} (\nabla \cdot R_P).$$

Then, we have the following :

1. *If $n > 2$ and $f_P \in L^p_{\text{loc}}(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ for some $p > n$, then $R_P \in \mathscr{C}^{n-1}(\mathbb{R}^n) \cap H^n_{\text{loc}}(\mathbb{R}^n)$. In particular, R_P admits weak derivatives of order n in $\mathbb{R}^n$, and we have*

$$f_P(x) = (\mathscr{L}_n R_P)(x) = ((-\Delta)^{\frac{1}{2}} R_P^{(n-1)})(x)$$

for almost any $x \in \mathbb{R}^n$. In addition, $R_P^{(n-1)} \in H^1(\mathbb{R}^n)$, and we have

$$((-\Delta)^{\frac{1}{2}} R_P^{(n-1)})(x) = 2\pi \mathscr{F}^{-1}\Big(|\xi| \mathscr{F} R_P^{(n-1)}\Big)(x)$$

for almost any $x \in \mathbb{R}^n$.

2. *If* $f_P \in \mathscr{C}^{k,\alpha}(\mathbb{R}^n)$ *for some* $k \in \mathbb{N}$ *and* $\alpha \in (0,1)$*, then* $R_P \in \mathscr{C}^{k+n}(\mathbb{R}^n)$ *and we have*

$$f_P(x) = (\mathscr{L}_n R_P)(x)$$

for any $x \in \mathbb{R}^n$*. In addition, we have*

$$(\mathscr{L}_n R_P)(x) = c_{n,1/2} \lim_{\eta \to 0} \int_{\mathbb{R}^n \setminus B_\eta(x)} \frac{R_P^{(n-1)}(x) - R_P^{(n-1)}(z)}{|x-z|^{n+1}}\, dz$$

for any $x \in \mathbb{R}^n$ *(see Section 2.5 for the definition of* $c_{n,1/2}$*).*

Proof Theorem 2.6.10. 1. Since $f_P \in L^p_{\text{loc}}(\mathbb{R}^n)$ with $p > n$, Proposition 2.6.8 entails that $R_P \in \mathscr{C}^{n-1}(\mathbb{R}^n)$. In particular $R_P^{(n-1)}$ is well-defined and continuous over $\mathbb{R}^n$. Let us show that $R_P^{(n-1)} \in L^2(\mathbb{R}^n)$. Proposition 2.6.8 entails that

$$R_P^{(n-1)}(x) = \gamma_n \mathbb{E}[((-\Delta)^{\frac{n-2}{2}}(\nabla \cdot K))(x-Z)],$$

where Z is a random n-vector with law P. It follows that

$$|R_P^{(n-1)}(x)| \lesssim \mathbb{E}\Big[\frac{1}{|x-Z|^{n-1}}\Big] =: h(x).$$

Therefore, it is enough to show that $h \in L^2(\mathbb{R}^n)$. Let us write

$$h(x) = \Big(\frac{1}{|z|^{n-1}} * f_P\Big)(x) = (u_1 * f_P)(x) + (u_2 * f_P)(x),$$

with $u_1(z) = \frac{1}{|z|^{n-1}}\mathbb{I}[|z| < 1]$ and $u_2(z) = \frac{1}{|z|^{n-1}}\mathbb{I}[|z| > 1]$. Since $u_1 \in L^1(\mathbb{R}^n)$ and $f_P \in L^2(\mathbb{R}^n)$, Hausdorff-Young's inequality entails that $u_1 * f_P \in L^2(\mathbb{R}^n)$. Since $n > 2$, we have $u_2 \in L^2(\mathbb{R}^n)$. Since $f_P \in L^1(\mathbb{R}^n)$, Hausdorff-Young's inequality yields $u_2 * f_P \in L^2(\mathbb{R}^n)$. It follows that $h \in L^2(\mathbb{R}^n)$, hence $R_P^{(n-1)} \in L^2(\mathbb{R}^n)$. Recall that $f_P = (-\Delta)^{\frac{1}{2}} R_P^{(n-1)}$ in $\mathscr{S}(\mathbb{R}^n)'$, with $R_P^{(n-1)} \in L^2(\mathbb{R}^n)$ and $f_P \in L^2(\mathbb{R}^n)$. Arguing as in the proof of Proposition 2.5.4, it is easy to show that this implies that $R_P^{(n-1)} \in H^1(\mathbb{R}^n)$. Proposition 2.5.4 therefore entails that

$$(-\Delta)^{\frac{1}{2}} R_P^{(n-1)} = 2\pi \mathscr{F}^{n-1}\big(|\xi| \mathscr{F} R_P^{(n-1)}\big)$$

in $L^2(\mathbb{R}^n)$ and that the equality

$$f_P = (-\Delta)^{\frac{1}{2}} R_P^{(n-1)}$$

also holds in $L^2(\mathbb{R}^n)$, i.e. almost everywhere. The same bootstrap argument than in the proof of Theorem 2.6.9 yields $g_P \in H^{n+1}_{\text{loc}}(\mathbb{R}^n)$, whence $R_P \in H^n_{\text{loc}}(\mathbb{R}^n)$.

2. The proof proceeds in two main steps. We first show that the fact that $f_P \in \mathscr{C}^{k,\alpha}(\mathbb{R}^n)$ entails that $R_P^{(n-1)} \in \mathscr{C}^{k+1,\alpha}_{\text{loc}}(\mathbb{R}^n)$. We will then show that this implies that $R_P \in \mathscr{C}^{k+n}(\mathbb{R}^n)$. Irrespective of the value of $k \in \mathbb{N}$, we will have $R_P^{(n-1)} \in \mathscr{C}^{1,\alpha}_{\text{loc}}(\mathbb{R}^n)$, which will entail that

$$(-\Delta)^{1/2} R_P^{(n-1)}(x) = c_{n,1/2} \lim_{\eta \to 0} \int_{\mathbb{R}^n \setminus B_\eta(x)} \frac{R_P^{(n-1)}(x) - R_P^{(n-1)}(z)}{|x-z|^{n+1}}\, dz$$

for any $x \in \mathbb{R}^n$, by Proposition 2.5.4.

Let us first show that $R_P^{(n-1)} \in \mathscr{C}^{k+1}(\mathbb{R}^n)$. Observe that $f_P \in L^\infty$, since $f_P \in \mathscr{C}^{0,\alpha}(\mathbb{R}^n)$. Theorem 2.6.8 then entails that $R_P \in \mathscr{C}_b^{n-1}(\mathbb{R}^n)$. In particular, $R_P^{(n-1)} \in \mathscr{C}^0(\mathbb{R}^n)$ is bounded over $\mathbb{R}^n$ so that $(-\Delta)^{1/2} R_P^{(n-1)}$ is a well-defined tempered distribution. Recall that the equality

$$f_P = (-\Delta)^{1/2} R_P^{(n-1)}$$

holds in $\mathscr{S}(\mathbb{R}^n)'$ by Theorem 2.6.5. If $k = 0$, we have $f_P \in \mathscr{C}^{0,\alpha}(\mathbb{R}^n)$ and $R_P^{(n-1)} \in L^\infty(\mathbb{R}^n)$ with $f_P = (-\Delta)^{1/2} R_P^{(n-1)}$ in $\mathscr{S}(\mathbb{R}^n)'$. Proposition 2.8 in [94] then entails that $R_P^{(n-1)} \in \mathscr{C}^{1,\alpha}(\mathbb{R}^n)$. Now assume that $k \geq 1$. Since $f_P \in L^1(\mathbb{R}^n) \subset \mathscr{S}_{1/2}(\mathbb{R}^n)'$, then $(-\Delta)^{1/2} f_P$ is well-defined. Furthermore, we have $(-\Delta)^{1/2} f_P \in \mathscr{C}^{k-1,\alpha}(\mathbb{R}^n)$ by Proposition 2.7 in [94], since $f_P \in \mathscr{C}^{k,\alpha}(\mathbb{R}^n)$. We then have

$$(-\Delta)^{1/2} f_P = -\Delta R_P^{(n-1)}$$

in $\mathscr{S}(\mathbb{R}^n)'$ with $(-\Delta)^{1/2} f_P \in \mathscr{C}^{k-1,\alpha}(\mathbb{R}^n)$ and $R_P^{(n-1)} \in H^1_{\text{loc}}(\mathbb{R}^n) \cap L^\infty_{\text{loc}}(\mathbb{R}^n)$ (recall that we just showed that $R_P^{(n-1)} \in \mathscr{C}^{1,\alpha}(\mathbb{R}^n)$ when considering the case $k = 0$). Therefore, Corollary 2.4.9 entails that $R_P^{(n-1)} \in \mathscr{C}^{k+1,\alpha}_{\text{loc}}(\mathbb{R}^n)$.

Let us now show that $R_P \in \mathscr{C}^{k+n}(\mathbb{R}^n)$. Recall that $R_P = \nabla g_P$ (see Definition 2.2.2 and the comments below) and that $R_P \in \mathscr{C}^{n-1}(\mathbb{R}^n)$, so that $g_P \in \mathscr{C}^n(\mathbb{R}^n)$. Consequently, we have

$$-R_P^{(n-1)}(x) = \gamma_n (-\Delta)^n g_P(x)$$

for any $x \in \mathbb{R}^n$. Since $R_P^{(n-1)} \in \mathscr{C}^{k+1,\alpha}_{\text{loc}}(\mathbb{R}^n)$ and $(-\Delta)^{\frac{n-2}{2}} g_P \in H^1_{\text{loc}}(\mathbb{R}^n) \cap L^\infty_{\text{loc}}(\mathbb{R}^n)$, we have $(-\Delta)^{\frac{n-2}{2}} g_P \in \mathscr{C}^{k+3,\alpha}_{\text{loc}}(\mathbb{R}^n)$ by Corollary 2.4.9. Repeating the argument recursively, we find that $g_P \in \mathscr{C}^{k+n+1,\alpha}_{\text{loc}}(\mathbb{R}^n)$. In particular, we have $R_P \in \mathscr{C}^{k+n,\alpha}_{\text{loc}}(\mathbb{R}^n)$, which entails that $R_P \in \mathscr{C}^{k+n}(\mathbb{R}^n)$. This concludes the proof. ■

2.7 Depth regions and probability content

Consider a probability measure P over $\mathbb{R}^n$, with $n \geq 2$. For any $\beta \in [0,1)$ and $u \in S^{n-1}$, recall that a geometric quantile of order β in direction u for P is an arbitrary minimizer of the objective function $O^P_{\beta,u}$, introduced in Section 2.1. When P is not supported on a single line of $\mathbb{R}^n$, Theorem 1 in [83] implies that the geometric quantile of order β in direction u for P is unique for any $\beta \in [0,1)$ and $u \in S^{n-1}$; we denote it by $Q_P(\beta u)$. Under these assumptions, we define the geometric quantile regions $\mathcal{D}^\beta_P$ and contours $\mathcal{C}^\beta_P$ of arbitrary order $\beta \in [0,1)$ in the next definition.

Definition 2.7.1 (Depth contours and regions)**.** *Let $n \geq 2$ and P be a probability measure over $\mathbb{R}^n$. Assume that P is not supported on a single line of $\mathbb{R}^n$. For any $\beta \in [0,1)$, we define the depth region $\mathcal{D}^\beta_P$ and depth contour $\mathcal{C}^\beta_P$ of order β for P as*

$$\mathcal{D}^\beta_P = \Big\{ Q_P(\alpha u) : \alpha \in [0,\beta],\ u \in S^{n-1} \Big\}$$

and

$$\mathcal{C}^\beta_P = \Big\{ Q_P(\beta u) : u \in S^{n-1} \Big\}.$$

When P is non-atomic and not supported on a line of $\mathbb{R}^n$, Proposition 3.6.2 in Chapter 3 entails that the map Q_P is continuous over the open unit ball B_1. It directly follows that

$$\mathcal{D}_P^\beta = Q_P(\beta\, \overline{B_1})$$

is compact and arc-connected, and that

$$\mathcal{C}_P^\beta = Q_P(\beta\, S^{n-1})$$

is compact and arc-connected as well. Furthermore, the depth regions $(D_P^\beta)_{\beta\in[0,1)}$ are obivously nested, while depth contours $(\mathcal{C}_P^\beta)_{\beta\in[0,1)}$ are disjoint. Although depth regions are convex in most cases, they may fail to be convex in general ; see [74] for a detailed and quantified discussion of the shape of depth regions.

To state regularity properties of depth contours, let us first rewrite depth contours in terms of the rank map R_P. Theorem 3.6.4 in Chapter 3 entails that $x = Q_P(\alpha u)$ if and only if $R_P(x) = \alpha u$. This allows one to rewrite

$$\mathcal{D}_P^\beta = \Big\{x \in \mathbb{R}^n : |R_P(x)| \leq \beta\Big\}$$

and

$$\mathcal{C}_P^\beta = \Big\{x \in \mathbb{R}^n : |R_P(x)| = \beta\Big\}.$$

The results of Section 2.6 may now easily be used to derive regularity properties of depth contours, as we show in the next proposition.

Proposition 2.7.2 (Regularity of depth contours)**.** *Let $n \geq 2$, and P be a probability measure over $\mathbb{R}^n$. Assume that P admits a density $f_P \in L^1(\mathbb{R}^n)$ with respect to the Lebesgue measure. We have the following :*

1. *If $f_P \in L^p_{\text{loc}}(\mathbb{R}^n)$ for some real $p > \frac{n}{n-\ell}$ and some integer $\ell \in [1, n-1]$, then the depth contour $\mathcal{C}_P^\beta$ is an $(n-1)$-dimensional manifold of class $\mathscr{C}^\ell$, for any $\beta \in [0,1)$;*
2. *If $f_P \in \mathscr{C}^{k,\alpha}(\mathbb{R}^n)$ for some $k \in \mathbb{N}$ and $\alpha \in (0,1)$, then the depth contour $\mathcal{C}_P^\beta$ is an $(n-1)$-dimensional manifold of class $\mathscr{C}^{k+n}$, for any $\beta \in [0,1)$.*

Proof of Proposition 2.7.2. Proposition 2.6.8, Theorem 2.6.9 and Theorem 2.6.10 yield that R_P has the stated regularity, $R_P \in \mathscr{C}^j(\mathbb{R}^n)$ say, and that

$$\partial^\alpha R_P(x) = \mathbb{E}[(\partial^\alpha K)(x - Z)]$$

for any $x \in \mathbb{R}^n$ and $\alpha \in \mathbb{N}^n$ with $|\alpha| \leq j$, where Z is a random n-vector with law P. Let us fix $\beta \in [0,1)$. Let $g_\beta(x) := |R_P(x)|^2 - \beta^2$. Then $g_\beta \in \mathscr{C}^j(\mathbb{R}^n)$ since the map $z \mapsto |z|^2$ is smooth over $\mathbb{R}^n$. We obviously have that

$$\mathcal{C}_P^\beta = \{x \in \mathbb{R}^n : g_\beta(x) = 0\}.$$

Let $z = (\tilde{z}, z_n) \in \mathbb{R}^{n-1} \times \mathbb{R}$ be such that $z \in \mathcal{C}_P^\beta$ and let us assume that $\nabla g_\beta(z) \neq 0$. The implicit function theorem entails that there exists an open neighbourhood $U \subset \mathbb{R}^{n-1}$ of

$\tilde{z}$, an open neighbourhood $V \subset \mathbb{R}^n$ of z, and a map $\varphi \in \mathscr{C}^j(U, \mathbb{R})$ such that $\varphi(\tilde{z}) = z_n$ and

$$V \cap \mathcal{C}_P^\beta = \{(\tilde{x}, \varphi(\tilde{x})) : \tilde{x} \in U\}.$$

In other words, in a neighbourhood of z, $\mathcal{C}_P^\beta$ is the graph of a function of class $\mathscr{C}^j$, which proves the claim.

It remains to show that $\nabla g_\beta(z) \neq 0$. Since $R_P \in \mathscr{C}^1(\mathbb{R}^n)$, we have

$$\nabla g_\beta(z) = 2 J_{R_P}(z)^T R_P(z),$$

where $J_{R_P}(z)^T$ stands for the transpose of the Jacobian matrix of R_P at z. Recall that $\partial_j R_P(z) = \mathbb{E}[(\partial_j K)(z - Z)]$ and that

$$J_K(x) = \frac{1}{|x|}\Big(I_n - \frac{xx^T}{|x|^2}\Big)$$

for any $x \in \mathbb{R}^n \setminus \{0\}$, where I_n stands for the $n \times n$ identity matrix. Therefore, we have

$$J_{R_P}(z) = \mathbb{E}\Big[\frac{1}{|z-Z|}\Big(I_n - \frac{(z-Z)(z-Z)^T}{|z-Z|^2}\Big)\mathbb{I}[Z \neq z]\Big].$$

The matrix J_{R_P} is obviously symmetric and non-negative definite. Let us show that it is positive definite. Assume, ad absurdum, that there exists $v \in \mathbb{R}^n$ such that $|v| = 1$ and $v^T J_{R_P}(z) v = 0$, i.e.

$$\mathbb{E}\Big[\frac{1}{|z-Z|}\Big(1 - \Big\langle v, \frac{z-Z}{|z-Z|}\Big\rangle^2\Big)\mathbb{I}[Z \neq z]\Big] = 0.$$

We then have

$$\frac{1}{|z-Z|}\Big(1 - \Big\langle v, \frac{z-Z}{|z-Z|}\Big\rangle^2\Big)\mathbb{I}[Z \neq z] = 0$$

P-almost surely. Since P admits a density, we have $\frac{1}{|z-Z|}\mathbb{I}[Z \neq z] \neq 0$ with P-probability 1. Consequently, we have

$$\Big|\Big\langle v, \frac{z-Z}{|z-Z|}\Big\rangle\Big| = 1$$

with P-probability 1. This implies that P is supported on the line through z with direction v, a contradiction. We deduce that $J_{R_P}(z)$ is positive definite, hence invertible. It follows that $J_{R_P}(z)^T R_P(z) \neq 0$, whence $\nabla g_\beta(z) \neq 0$. This concludes the proof. ■

Unlike center-outward quantiles based on optimal transport [44], geometric quantile regions are not indexed by their probability content, i.e. we do not have $P[\mathcal{D}_P^\beta] = \beta$ in general. However, one can in principle re-index quantile regions so that they match their probability content. Assume that P admits a density f_P over $\mathbb{R}^n$ such that $f_P(x) > 0$ for any $x \in \mathbb{R}^n$. Let $\theta_P(\beta) = P[\mathcal{D}_P^\beta]$ for any $\beta \in [0,1)$. Since quantile regions are nested, the map θ_P is monotone non-decreasing. The assumptions on P further ensure that $\theta_P : [0,1) \to [0,1)$ is continuous, increasing, and surjective. In particular, θ_P is bijective. It follows that the re-indexed quantile regions

$$\widetilde{\mathcal{D}}_P^\beta := \mathcal{D}_P^{\theta_P^{-1}(\beta)}$$

match their probability content, i.e. we have $P[\tilde{\mathcal{D}}_P^\beta] = \beta$ for any $\beta \in [0,1)$. We similarly define the re-indexed quantile contours

$$\widetilde{\mathcal{C}}_P^\beta := \mathcal{C}_P^{\theta_P^{-1}(\beta)}$$

for any $\beta \in [0,1)$. This suggests defining an alternative rank function $\widetilde{R}_P(x)$. To do so, observe that $x \in \widetilde{\mathcal{C}}_P^\beta$ if and only if

$$\Big|\frac{\beta}{\theta_P^{-1}(\beta)} R_P(x)\Big| = \beta.$$

When the previous equality holds, we have $\beta = \theta(|R_P(x)|)$. This suggests letting

$$\widetilde{R}_P(x) = \theta_P(|R_P(x)|)\frac{R_P(x)}{|R_P(x)|}$$

for any $x \in \mathbb{R}^n$. It follows that

$$\widetilde{\mathcal{D}}_P^\beta = \{x \in \mathbb{R}^n : |\widetilde{R}_P(x)| \leq \beta\}$$

and

$$\widetilde{\mathcal{C}}_P^\beta = \{x \in \mathbb{R}^n : |\widetilde{R}_P(x)| = \beta\}$$

for any $\beta \in [0,1)$. Letting Z denote a random n-vector with law P, it is clear that $\theta_P(|R_P(Z)|)$ is uniformly distributed over $[0,1)$. Indeed, we have

$$P\Big[\theta_P(|R_P(Z)|) \leq \beta\Big] = P\Big[|R_P(Z)| \leq \theta_P^{-1}(\beta)\Big] = P\big[\widetilde{\mathcal{D}}_P^\beta\big] = \beta$$

for any $\beta \in [0,1)$. In fact, θ_P is the cdf of $|R_P(Z)|$.

Recall that

$$f_P(x) = \gamma_n(-\Delta)^{\frac{n-1}{2}}(\nabla \cdot R_P)(x)$$

for any $x \in \mathbb{R}^n$, by Theorem 2.6.9 and Theorem 2.6.10. Interchanging the order of differential operators yields

$$f_P = \gamma_n \nabla \cdot \big((-\Delta)^{\frac{n-1}{2}} R_P\big)$$

over $\mathbb{R}^n$, where $(-\Delta)^{\frac{n-1}{2}} R_P$ is the operator $(-\Delta)^{\frac{n-1}{2}}$ applied componentwise to R_P. For any (regular) open and bounded subset $\Omega \subset \mathbb{R}^n$, we have by divergence's theorem

$$P[\Omega] = \int_\Omega f_P(x)\,dx = \gamma_n \int_{\partial\Omega} \Big((-\Delta)^{\frac{n-1}{2}} R_P(x), \nu(x)\Big)\,dH_{n-1}(x), \tag{2.7.14}$$

for any $\beta \in [0,1)$, where H_{n-1} denotes the $(n-1)$-dimensional Hausdorff measure, and $\nu(x)$ is the outer unit normal vector to Ω at x. The probability content of an arbitrary open subset is then controlled by $(-\Delta)^{\frac{n-1}{2}} R_P$. Notice that $(-\Delta)^{\frac{n-1}{2}} R_P$ and R_P actually coincide when $n = 1$. Therefore, Equation (2.7.14) is a multivariate analog of the well-known equality

$$P\Big[Z \in (a,b)\Big] = F_P(b) - F_P(a)$$

when $n = 1$, and where F_P stands for the usual univariate cdf.

2.8 Localization issues

In this section we investigate the local properties of the operator $\mathscr{L}_n$. As we have already mentionned, the operator $\mathscr{L}_n = (-\Delta)^{\frac{n-1}{2}}\nabla\cdot$ display substantially different behaviours in odd and even dimensions. This is due the nature of $(-\Delta)^{\frac{n-1}{2}}$, which depends on whether $\frac{n-1}{2}$ is an integer or not. When $\frac{n-1}{2} \in \mathbb{N}$, then $(-\Delta)^{\frac{n-1}{2}}$ is the classical differential operator that consists in applying the Laplacian $-\Delta$ successively $\frac{n-1}{2}$ times. This operator is local in nature : if f_1 and f_2 are smooth functions that coincide over an open subset $U \subset \mathbb{R}^n$, then $(-\Delta)^{\frac{n-1}{2}} f_1$ and $(-\Delta)^{\frac{n-1}{2}} f_2$ also coincide over U. When n is even, then $\frac{n-1}{2} \in \mathbb{R} \setminus \mathbb{N}$; in this case, we write $(-\Delta)^{\frac{n-1}{2}} = (-\Delta)^{1/2}(-\Delta)^{\frac{n-2}{2}}$. Although $(-\Delta)^{1/2}$ acts like a derivative in terms of regularity (see Proposition 2.6 in [94]), it is also known to be a non-local operator.

Multivariate geometric ranks characterize probability measures in arbitrary dimension n : if P and Q are Borel probability measures over $\mathbb{R}^n$ and if $R_P(x) = R_Q(x)$ for any $x \in \mathbb{R}^n$, then $P = Q$ (see Theorem 2.5 in [50]). When n is odd, we provide a refinement of this result in the next proposition, thanks to the local nature of $\mathscr{L}_n$.

Proposition 2.8.1. *Let $n \geq 1$ be odd. Let also P and Q be a Borel probability measures over $\mathbb{R}^n$. Let $\Omega \subset \mathbb{R}^n$ be an open subset, and assume that $R_P(x) = R_Q(x)$ for any $x \in \Omega$. Then, P and Q coincide over Ω, i.e. $P(E) = Q(E)$ for any Borel subset $E \subset \Omega$.*

Proof of Proposition 2.8.1. By Theorem 2.6.5, we have

$$\int_{\mathbb{R}^n} \psi(x)\, dP(x) = \int_{\mathbb{R}^n} (R_P(x), (\mathscr{L}_n^* \psi)(x))\, dx$$

and

$$\int_{\mathbb{R}^n} \psi(x)\, dQ(x) = \int_{\mathbb{R}^n} (R_Q(x), (\mathscr{L}_n^* \psi)(x))\, dx$$

for any $\psi \in \mathscr{S}(\mathbb{R}^n)$. In particular, the above equalities hold for any $\psi \in \mathscr{C}_c^\infty(\Omega)$. Let us observe that $\mathscr{L}_n^* = \gamma_n \nabla(-\Delta^{\frac{n-1}{2}})$ is a (non-fractional) differential operator, since $\frac{n-1}{2}$ is an integer. In particular, $\mathscr{L}_n^* \psi$ is also supported in Ω. Since $R_P = R_Q$ over Ω, we then have

$$\int_\Omega \psi(x)\, dP(x) = \int_\Omega \psi(x)\, dQ(x)$$

for any $\psi \in \mathscr{C}_c^\infty(\Omega)$. It follows that $P(E) = Q(E)$ for any Borel subset $E \subset \Omega$. ∎

When n is even, the operator $\mathscr{L}_n$ is non-local. In particular, the proof of Proposition 2.8.1 does not apply. We present two approaches attempting to recover a localization result similar to Proposition 2.8.1.

Consider a probability measure P over $\mathbb{R}^n$ with n even. The first idea that naturally comes to mind is to embed P into $\mathbb{R}^{n+1}$ (with $n+1$ odd); this gives rise to a probability measure P^* supported on the hyperplane $x_{n+1} = 0$ of $\mathbb{R}^{n+1}$. Proposition 2.8.1 now applies to P^*. The other approach consists in localizing the operator $(-\Delta)^{1/2}$. For a smooth function u over $\mathbb{R}^n$, computing $(-\Delta)^{1/2}u$ can be achieved by first solving $-\Delta U = 0$ over $\mathbb{R}_+^{n+1} := \mathbb{R}^n \times (0, \infty)$ subject to the boundary condition $U(\tilde{x}, 0) = u(\tilde{x})$ for any $\tilde{x} \in \mathbb{R}^n$. One then has

$$((-\Delta)^{1/2}u)(\tilde{x}) = -\lim_{x_{n+1}\to 0} (\partial_{n+1}U)(\tilde{x}, x_{n+1})$$

for any $\tilde{x} \in \mathbb{R}^n$. This formulation is now local with respect to U since the values of $\partial_{n+1}U$ in some open subset $\Omega \subset \mathbb{R}^{n+1}_+$ depend on the values of U on Ω only. For further details on this method, we refer the reader to [11] and [94].

It turns out that both approaches are equivalent. This is the content of the next proposition, in which we will show that the density f_P of a probability measure P over $\mathbb{R}^n$ (n even) can be recovered through $\lim_{x_{n+1}\to 0} \partial_{n+1}U(\tilde{x}, x_{n+1})$, where $U(\tilde{x}, x_{n+1})$ is issentially equal to

$$(-\Delta)^{\frac{n-2}{2}}(\nabla \cdot R_{P^*})(\tilde{x}, x_{n+1}),$$

and solves $-\Delta U = 0$ over $\mathbb{R}^{n+1}_+$.

Proposition 2.8.2. *Let $n \geq 2$ be even, and P be a Borel probability measure over $\mathbb{R}^n$. Assume that P admits a density $f_P \in L^1(\mathbb{R}^n)$ with respect to the Lebesgue measure, and that $f_P \in \mathscr{C}^{0,\alpha}(\mathbb{R}^n)$ for some $\alpha \in (0, 1)$. Let P^* denote the probability measure over $\mathbb{R}^{n+1}$ supported on the hyperplane $x_{n+1} = 0$ with density f_P with respect to the n-dimensonal Hausdorff measure $\mathcal{H}_n$. Let Z be a random n-vector with law P, and Z^* be a random $(n+1)$-vector with law P^*. Define*

$$U(x) = 2\gamma_{n+1}\mathbb{E}\Big[((-\Delta)^{\frac{n-2}{2}}(\nabla \cdot K_{n+1}))(x - Z^*)\Big]$$

for any $x \in \mathbb{R}^{n+1}_+ := \mathbb{R}^n \times (0, \infty)$, and

$$u(\tilde{x}) = \gamma_n(-\Delta)^{\frac{n-2}{2}}(\nabla \cdot R_P)(\tilde{x})$$

for any $\tilde{x} \in \mathbb{R}^n$. We have $U \in \mathscr{C}^\infty(\mathbb{R}^{n+1}_+)$ and $u \in \mathscr{C}^1(\mathbb{R}^n)$. In addition, the following holds :

1. *$U(x) = 2\gamma_{n+1}(-\Delta)^{\frac{n-2}{2}}(\nabla \cdot R_{P^*})(x)$ and $-\Delta U(x) = 0$, for any $x \in \mathbb{R}^{n+1}_+$;*

2. *for any $\tilde{x} \in \mathbb{R}^n$, $U(\tilde{x}, 0) = u(\tilde{x})$ and*

$$f_P(\tilde{x}) = ((-\Delta)^{1/2}u)(\tilde{x}) = \lim_{x_{n+1} \overset{>}{\to} 0} -(\partial_{n+1}U)(\tilde{x}, x_{n+1}).$$

In practice, Proposition 2.8.2 entails that one can recover f_P by applying purely (local) differential operators to the geometric rank associated to P^* instead of P. We summarize this in the following corollary.

Corollary 2.8.3. *Let $n \geq 2$ be even, and P be a Borel probability measure over $\mathbb{R}^n$. Assume that P admits a density $f_P \in L^1(\mathbb{R}^n)$ with respect to the Lebesgue measure and that $f_P \in \mathscr{C}^{0,\alpha}(\mathbb{R}^n)$ for some $\alpha \in (0, 1)$. Let P^* denote the probability measure over $\mathbb{R}^{n+1}$ supported on the hyperplane $x_{n+1} = 0$ with density f_P with respect to the n-dimensonal Hausdorff measure $\mathcal{H}_n$. Then,*

$$f_P(\tilde{x}) = -2\gamma_{n+1} \lim_{x_{n+1} \overset{>}{\to} 0} \partial_{n+1}(-\Delta)^{\frac{n-2}{2}}(\nabla \cdot R_{P^*})(\tilde{x}, x_{n+1})$$

for any $\tilde{x} \in \mathbb{R}^n$.

Proof of Proposition 2.8.2. Since P^* admits the null density over the open subset $\mathbb{R}^{n+1}_+ := \mathbb{R}^n \times (0, \infty)$ of $\mathbb{R}^{n+1}$, Proposition 2.6.8 entails that $R_{P^*} \in \mathscr{C}^n(\mathbb{R}^{n+1}_+)$ and that

$$\partial^\alpha R_{P^*}(x) = \mathbb{E}[(\partial^\alpha K_{n+1})(x - Z^*)]$$

for any $x \in \mathbb{R}^{n+1}_+$ and $\alpha \in \mathbb{N}^{n+1}$ with $|\alpha| \leq n$, where Z^* is a random $(n+1)$-vector with law P^*. Letting

$$U(x) := 2\gamma_{n+1}\mathbb{E}\Big[((-\Delta)^{\frac{n-2}{2}}(\nabla \cdot K_{n+1}))(x - Z^*)\Big]$$

for any $x \in \mathbb{R}^{n+1}_+$, we then have

$$U(x) = 2\gamma_{n+1}(-\Delta)^{\frac{n-2}{2}}(\nabla \cdot R_{P^*})(x)$$

for any $x \in \mathbb{R}^{n+1}_+$. Theorem 2.6.9 further implies that $-\Delta U(x) = 0$ for any $x \in \mathbb{R}^{n+1}_+$.

Let us show that $U(\tilde{x}, 0) = u(\tilde{x})$ for any $\tilde{x} \in \mathbb{R}^n$, where

$$u(\tilde{x}) = \gamma_n(-\Delta)^{\frac{n-2}{2}}(\nabla \cdot R_P)(\tilde{x}).$$

First observe that $R_P \in \mathscr{C}^{n-1}(\mathbb{R}^n)$, and

$$u(\tilde{x}) = \gamma_n\mathbb{E}\Big[((-\Delta)^{\frac{n-2}{2}}(\nabla \cdot K_n))(\tilde{x} - Z)\Big]$$

for any $\tilde{x} \in \mathbb{R}^n$, by Proposition 2.6.8. Let us compute explicitly $(-\Delta)^{\frac{n-2}{2}}(\nabla \cdot K_n)$ and $(-\Delta)^{\frac{n-2}{2}}(\nabla \cdot K_{n+1})$. It is easy to see that

$$(\nabla \cdot K_n)(\tilde{x}) = (n-1)\frac{1}{|\tilde{x}|}$$

for any $\tilde{x} \in \mathbb{R}^n \setminus \{0\}$. Easy computations further show that

$$(-\Delta)^\ell \frac{1}{|\tilde{x}|} = \Lambda_{n,\ell}\frac{1}{|\tilde{x}|^{2\ell+1}}, \tag{2.8.15}$$

for any $\tilde{x} \in \mathbb{R}^n \setminus \{0\}$, where

$$\Lambda_{n,\ell} = \prod_{j=1}^{\ell}(2j-1)(n-2j-1) \tag{2.8.16}$$

for any $1 \leq \ell \leq \frac{n-2}{2}$. It follows that

$$(-\Delta)^{\frac{n-2}{2}}(\nabla \cdot K_n)(\tilde{x}) = (n-1)\Lambda_{n,\frac{n-2}{2}}\frac{1}{|\tilde{x}|^{n-1}}$$

for any $\tilde{x} \in \mathbb{R}^n \setminus \{0\}$. The same computations yield

$$(-\Delta)^{\frac{n-2}{2}}(\nabla \cdot K_{n+1})(x) = n\Lambda_{n+1,\frac{n-2}{2}}\frac{1}{|x|^{n-1}}$$

for any $x \in \mathbb{R}^{n+1} \setminus \{0\}$. Using the fact that

$$\prod_{j=1}^{k}(2j-1) = \frac{(2k)!}{2^k k!}$$

for any integer $k \geq 1$, it is easy to see that

$$\Lambda_{n,\frac{n-2}{2}} = \left(\frac{\Gamma(n-1)}{2^{\frac{n-2}{2}}\Gamma(\frac{n}{2})}\right)^2$$

and $\Lambda_{n+1,\frac{n-2}{2}} = \Gamma(n-1)$. It follows that

$$U(x) = 2\gamma_{n+1} n\Gamma(n-1)\mathbb{E}\left[\frac{1}{|x-Z^*|^{n-1}}\mathbb{I}[Z \neq x]\right]$$

for any $x \in \mathbb{R}^{n+1}_+$ and that

$$u(\tilde{x}) = \gamma_n(n-1)\left(\frac{\Gamma(n-1)}{2^{\frac{n-2}{2}}\Gamma(\frac{n}{2})}\right)^2 \mathbb{E}\left[\frac{1}{|\tilde{x}-Z|^{n-1}}\mathbb{I}[Z \neq \tilde{x}]\right]$$

for any $\tilde{x} \in \mathbb{R}^n$. In particular, we have

$$\begin{aligned} U(\tilde{x},0) &= 2n\gamma_{n+1}\Gamma(n-1)\mathbb{E}\left[\frac{1}{|\tilde{x}-Z|^{n-1}}\mathbb{I}[Z \neq \tilde{x}]\right] \\ &= 2n\gamma_{n+1}\Gamma(n-1) \times \frac{2^{n-2}\Gamma(\frac{n}{2})^2}{\gamma_n(n-1)\Gamma(n-1)^2}u(\tilde{x}) \\ &= \frac{\gamma_{n+1}}{\gamma_n} \times \frac{2^{n-1}n\Gamma(\frac{n}{2})^2}{\Gamma(n)}u(\tilde{x}) \end{aligned}$$

for any $\tilde{x} \in \mathbb{R}^n$. Using the fact that

$$\Gamma\left(k+\frac{1}{2}\right) = \frac{\sqrt{\pi}\,\Gamma(2k+1)}{2^{2k}\Gamma(k+1)} \tag{2.8.17}$$

for any $k \in \mathbb{N}$ leads to

$$\frac{\gamma_{n+1}}{\gamma_n} = \frac{\Gamma(n+1)}{2^{n+1}\Gamma(\frac{n}{2}+1)^2} = \frac{n\Gamma(n)}{2^{n+1}(\frac{n}{2}\Gamma(\frac{n}{2}))^2} = \frac{\Gamma(n)}{2^{n-1}n\Gamma(\frac{n}{2})^2}.$$

It follows that $U(\tilde{x},0) = u(\tilde{x})$ for any $\tilde{x} \in \mathbb{R}^n$.

Let us now compute $-\partial_{n+1}U(\tilde{x},x_{n+1})$ for any $(\tilde{x},x_{n+1}) \in \mathbb{R}^n \times (0,\infty)$. We have already noticed that

$$\partial_{n+1}U(x) = 2\gamma_{n+1}\mathbb{E}\left[\partial_{n+1}((-\Delta)^{\frac{n-2}{2}}(\nabla \cdot K_{n+1}))(x-Z^*)\right]$$

for any $x \in \mathbb{R}^{n+1}_+$. Writing $Z^* = (Z^*_1, \ldots, Z^*_{n+1})$, we then have

$$-(\partial_{n+1}U)(\tilde{x},x_{n+1}) = 2\gamma_{n+1}\Gamma(n+1)\mathbb{E}\left[\frac{x_{n+1}-Z^*_{n+1}}{|x-Z^*|^{n+1}}\mathbb{I}[Z \neq x]\right]$$

for any $(\tilde{x},x_{n+1}) \in \mathbb{R}^n \times (0,\infty)$. Let us now show that

$$\mathbb{E}\left[\frac{x_{n+1}-Z^*_{n+1}}{|x-Z^*|^{n+1}}\mathbb{I}[Z \neq x]\right] \to \frac{\pi^{\frac{n+1}{2}}}{\Gamma(\frac{n+1}{2})}f_P(\tilde{x})$$

as $x_{n+1} \to 0$. Letting $\mathcal{H}_n$ denote the n-dimensional Haussdorf measure in $\mathbb{R}^{n+1}$, we have for any $x \in \mathbb{R}^{n+1}_+$

$$\begin{aligned}
&\mathbb{E}\Big[\frac{x_{n+1} - Z^*_{n+1}}{|x - Z^*|^{n+1}}\mathbb{I}[Z \neq x]\Big] \\
&\quad = \int_{\{z_{n+1}=0\}} \frac{x_{n+1} - z_{n+1}}{\Big(|\tilde{x} - \tilde{z}|^2 + (x_{n+1} - z_{n+1})^2\Big)^{\frac{n+1}{2}}} f_P(\tilde{z}) d\mathcal{H}_n(\tilde{z}, z_{n+1}) \\
&\quad = \int_{\mathbb{R}^n} \frac{x_{n+1}}{(|\tilde{x} - \tilde{z}|^2 + x_{n+1}^2)^{\frac{n+1}{2}}} f_P(\tilde{z}) d\tilde{z} \\
&\quad = \int_{\mathbb{R}^n} \frac{1}{x_{n+1}^n} \frac{1}{\Big(1 + |\frac{\tilde{x}-\tilde{z}}{x_{n+1}}|^2\Big)^{\frac{n+1}{2}}} f_P(\tilde{z}) d\tilde{z} \\
&\quad = \int_{\mathbb{R}^n} \frac{1}{(1 + |\tilde{z}|^2)^{\frac{n+1}{2}}} f_P(\tilde{x} - x_{n+1}\tilde{z}) d\tilde{z}.
\end{aligned}$$

It follows that

$$\mathbb{E}\Big[\frac{x_{n+1} - Z^*_{n+1}}{|x - Z^*|^{n+1}}\mathbb{I}[Z \neq x]\Big] \to f_P(\tilde{x}) \int_{\mathbb{R}^n} \frac{1}{(1 + |\tilde{z}|^2)^{\frac{n+1}{2}}} \, d\tilde{z}$$

as $x_{n+1} \to 0$. Indeed, we have

$$\begin{aligned}
&\Big| \int_{\mathbb{R}^n} \frac{1}{(1 + |\tilde{z}|^2)^{\frac{n+1}{2}}} f_P(\tilde{x} - x_{n+1}\tilde{z}) d\tilde{z} - f_P(\tilde{x}) \int_{\mathbb{R}^n} \frac{1}{(1 + |\tilde{z}|^2)^{\frac{n+1}{2}}} \, d\tilde{z} \Big| \\
&\qquad \leq [f_P]_{\mathscr{C}0,\alpha} |x_{n+1}|^\alpha \int_{\mathbb{R}^n} \frac{|\tilde{z}|^\alpha}{(1 + |\tilde{z}|^2)^{\frac{n+1}{2}}} \, d\tilde{z},
\end{aligned}$$

where $[f_P]_{\mathscr{C}0,\alpha} := \sup_{x \neq y} \frac{|f_P(x) - f_P(y)|}{|x-y|^\alpha}$. The latter converges to 0 as $x_{n+1} \to 0$ because

$$\int_{\mathbb{R}^n} \frac{|\tilde{z}|^\alpha}{(1 + |\tilde{z}|^2)^{\frac{n+1}{2}}} \, d\tilde{z} < \infty$$

since $0 < \alpha < 1$. Furthermore, one can show that

$$\int_{\mathbb{R}^n} \frac{1}{(1 + |\tilde{z}|^2)^{\frac{n+1}{2}}} \, d\tilde{z} = S_{n-1} \frac{\sqrt{\pi}\, \Gamma(n/2)}{2\Gamma(\frac{n+1}{2})} = \frac{\pi^{\frac{n+1}{2}}}{\Gamma(\frac{n+1}{2})},$$

where $S_{n-1} = 2\pi^{\frac{n}{2}}/\Gamma(\frac{n}{2})$ is the surface area of the $(n-1)$-dimensional sphere of $\mathbb{R}^n$. We deduce that

$$-(\partial_{n+1} U)(\tilde{x}, x_{n+1}) \to 2\gamma_{n+1}\Gamma(n+1) \frac{\pi^{\frac{n+1}{2}}}{\Gamma(\frac{n+1}{2})} f_p(\tilde{x})$$

as $x_{n+1} \to 0$. Using again (2.8.17), we see that

$$2\gamma_{n+1}\Gamma(n+1) \frac{\pi^{\frac{n+1}{2}}}{\Gamma(\frac{n+1}{2})} = 1.$$

It follows that

$$-(\partial_{n+1}U)(\tilde{x}, x_{n+1}) \to f_P(\tilde{x})$$

as $x_{n+1} \to 0$. Since $f \in \mathscr{C}^{0,\alpha}(\mathbb{R}^n)$, Theorem 2.6.10 entails that

$$f_P(\tilde{x}) = \gamma_n(-\Delta)^{1/2}(-\Delta)^{\frac{n-2}{2}}(\nabla \cdot R_P)(\tilde{x}) = ((-\Delta)^{1/2}u)(\tilde{x})$$

for any $\tilde{x} \in \mathbb{R}^n$. This concludes the proof. ■

2.9 Weak convergence via geometric ranks

In this section, we explore the link between the convergence in distribution of probability measures over $\mathbb{R}^n$ and pointwise convergence of their geometric ranks. Similarly to the univariate case, we show that convergence in distribution is equivalent to pointwise convergence of the corresponding geometric ranks away from atomicity points of the limiting distribution.

If a sequence of probability measures (P_k) over $\mathbb{R}^n$ converges in distribution to a probability measure P, it is easy to show that $R_{P_k}(x) \to R_P(x)$ for any $x \in \mathbb{R}^n$ such that $P[\{x\}] = 0$. When $n = 1$, the converse also holds : convergence in distribution is equivalent to pointwise convergence of the cdf's away from atomicity points of the limiting distribution. When $n > 1$, proving the converse requires understanding how P is related to R_P. Theorem 2.5 in [50] entails that R_P characterizes P : if P and Q are probability measures such that $R_P(x) = R_Q(x)$ for any $x \in \mathbb{R}^n$, then $P = Q$. This result is abstract and does not provide an effective way to recover P from R_P. By using the PDE we established in Section 2.6, we are able to prove the following result.

Theorem 2.9.1. *Let P and $(P_k)_{k\geq 1}$ be Borel probability measures over $\mathbb{R}^n$. Let*

$$A = \{y \in \mathbb{R}^n : P[\{y\}] > 0\}$$

be the set of atoms of P. Then (P_k) converges to P in distribution as $k \to \infty$ if and only if $R_{P_k}(x) \to R_P(x)$ for any $x \in \mathbb{R}^n \setminus A$ as $k \to \infty$.

Proof of Theorem 2.9.1. First assume that (P_k) converges in distribution to Q. For any $x \in \mathbb{R}^n$, let $g_x(z) := \frac{x-z}{\|x-z\|}\mathbb{I}[z \neq x]$ for any $z \in \mathbb{R}^n$. With the notations of Lemma 2.6.6, we have $D_{g_x} = \{x\}$. Therefore, we have $P[D_{g_x}] = 0$ for any $x \in \mathbb{R}^n \setminus A$. Since g_x is bounded and measurable for any $x \in \mathbb{R}^n$, Lemma 2.6.6 entails that

$$R_{P_k}(x) = \int_{\mathbb{R}^n} g_x(z)\, dP_k(z) \to \int_{\mathbb{R}^n} g_x(z)\, dP(z) = R_P(x)$$

for any $x \in \mathbb{R}^n \setminus A$ as $k \to \infty$.

Now assume that $R_{P_k}(x) \to R_P(x)$ for any $x \in \mathbb{R}^n \setminus A$. In particular, $R_{P_k} \to R_P$ almost everywhere since A is at most countable. We are going to show that

$$\int_{\mathbb{R}^n} \psi(z)\, dP_k(z) \to \int_{\mathbb{R}^n} \psi(z)\, dP(z)$$

for any $\psi \in \mathscr{S}(\mathbb{R}^n)$ as $k \to \infty$. For this purpose, let $\psi \in \mathscr{S}(\mathbb{R}^n)$. Theorem 2.6.5 entails that

$$\int_{\mathbb{R}^n} \psi(z)\, dP_k(z) = \int_{\mathbb{R}^n} \langle R_{P_k}(z), (\mathscr{L}_n^* \psi)(z) \rangle \, dz,$$

and that

$$\int_{\mathbb{R}^n} \psi(z)\, dP(z) = \int_{\mathbb{R}^n} \langle R_P(z), (\mathscr{L}_n^* \psi)(z) \rangle \, dz.$$

Assume that $\mathscr{L}_n^*(\psi) \in L^1(\mathbb{R}^n)$. Since $R_{P_k} \to R_P$ almost everywhere and (R_{P_k}) is norm-bounded by 1, Lebesgue's dominated convergence theorem entails that

$$\int_{\mathbb{R}^n} \langle R_{P_k}(z), (\mathscr{L}_n^* \psi)(z) \rangle \, dz \to \int_{\mathbb{R}^n} \langle R_P(x), (\mathscr{L}_n^* \psi)(z) \rangle \, dz$$

as $k \to \infty$. This yields

$$\int_{\mathbb{R}^n} \psi(z)\, dP_k(z) \to \int_{\mathbb{R}^n} \psi(z)\, dP(z)$$

as $k \to \infty$. Since $\psi \in \mathscr{S}(\mathbb{R}^n)$ was arbitrary, Theorem 1.5.3 in [6] entails that (P_k) converges to P in distribution (i.e. in law) as $k \to \infty$.

It remains to show that $\mathscr{L}_n^*(\psi) \in L^1(\mathbb{R}^n)$. If n is even, we have $(-\Delta)^{\frac{n-2}{2}}\psi \in \mathscr{S}(\mathbb{R}^n)$, so that $\mathscr{L}_n^*(\psi) = (-\Delta)^{\frac{1}{2}}((-\Delta)^{\frac{n-2}{2}}\psi) \in \mathscr{S}_{1/2}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n)$. If n is odd, then $\nabla\big((-\Delta)^{\frac{n-1}{2}}\psi\big)$ trivially belongs to $\mathscr{S}(\mathbb{R}^n, \mathbb{C}^n)$ which is a subset of $L^1(\mathbb{R}^n)$. In any case, we have $\mathscr{L}_n^*(\psi) \in L^1(\mathbb{R}^n)$. This concludes the proof. ■

2.10 A Glivenko-Cantelli result

When P is an absolutely continuous probability measure with a bounded density, [72] proved the following result. We extend it to any non-atomic probability measure P.

Theorem 2.10.1. *Let P be a non-atomic probability measure on $\mathbb{R}^n$. Let (P_k) be a sequence of probability measures on $\mathbb{R}^n$ that converges to P in distribution as $k \to \infty$. Then*

$$\sup_{x \in \mathbb{R}^n} \|R_{P_k}(x) - R_P(x)\| \to 0$$

as $k \to \infty$.

Proof of Theorem 2.10.1. The first part of this proof follows the same strategy as the proof of Lemma 2 in [72]. For any $\varepsilon > 0$ and $z \in \mathbb{R}^n$, let

$$S_\varepsilon(z) = \begin{cases} \frac{z}{\|z\|} & \text{if } \|z\| > \varepsilon, \\ \frac{z}{\varepsilon} & \text{if } \|z\| \leq \varepsilon. \end{cases}$$

For any probability measure Q over $\mathbb{R}^n$, let

$$R_{Q,\varepsilon}(x) := \int_{\mathbb{R}^n} S_\varepsilon(x - z)\, dQ(z)$$

for any $x \in \mathbb{R}^n$. For any $x \in \mathbb{R}^n$, any $\varepsilon > 0$ and $k \in \mathbb{N}$, let us write

$$R_{P_k}(x) - R_P(x) = (R_{P_k}(x) - R_{P_k,\varepsilon}(x)) + (R_{P_k,\varepsilon}(x) - R_{P,\varepsilon}(x)) + (R_{P,\varepsilon}(x) - R_P(x))$$
$$=: D^\varepsilon_{1,k}(x) + D^\varepsilon_{2,k}(x) + D^\varepsilon_{3,k}(x)$$

Let us show that $\sup_{x\in\mathbb{R}^n} \|D^\varepsilon_{2,k}(x)\| \to 0$ as $k \to \infty$ for any $\varepsilon > 0$. Let $\varepsilon > 0$ be fixed and

$$\mathcal{F}^\varepsilon = \{y \mapsto S_\varepsilon(x - y) : x \in \mathbb{R}^n\}.$$

It is clear that any $f \in \mathcal{F}^\varepsilon$ is continuous over $\mathbb{R}^n$ with $\|f\| \leq 1$ over $\mathbb{R}^n$. Assume that $\mathcal{F}^\varepsilon$ is equicontinuous at every point of $\mathbb{R}^n$. Since (P_k) converges in distribution to P as $k \to \infty$, Theorem 3.1 in [86] yields

$$\sup_{x\in\mathbb{R}^n} \|D^\varepsilon_{2,k}(x)\| = \sup_{f\in\mathcal{F}^\varepsilon} \left\| \int_{\mathbb{R}^n} f\, dP_k - \int_{\mathbb{R}^n} f\, dP \right\| \to 0 \tag{2.10.18}$$

as $k \to \infty$. It remains to show that $\mathcal{F}^\varepsilon$ is equicontinuous at every point of $\mathbb{R}^n$. It is clear that S_ε is $(1/\varepsilon)$-Lipschitz over $\overline{B_\varepsilon}$. We also have that S_ε is $(\sqrt{n}/\varepsilon)$-Lipschitz over $\mathbb{R}^n \setminus B_\varepsilon$ since the gradient of each of its components is bounded by $1/\varepsilon$ over $\mathbb{R}^n \setminus B_\varepsilon$. We deduce that S_ε is globally $(\sqrt{n}/\varepsilon)$-Lipschitz over $\mathbb{R}^n$. Indeed, let $x \in \overline{B_\varepsilon}$ and $y \in \mathbb{R}^n \setminus B_\varepsilon$. Let w belong to the segment $\{(1 - s)x + sy : s \in [0, 1]\}$ and be such that $\|w\| = \varepsilon$. We then have

$$\|S_\varepsilon(x) - S_\varepsilon(w)\| \leq \frac{1}{\varepsilon}\|x - w\| \leq \frac{\sqrt{n}}{\varepsilon}\|x - w\|$$

and

$$\|S_\varepsilon(w) - S_\varepsilon(y)\| \leq \frac{\sqrt{n}}{\varepsilon}\|w - y\|.$$

But $\|x - w\| + \|w - y\| = \|x - y\|$ since x, w and y are colinear. It follows that

$$\|S_\varepsilon(x) - S_\varepsilon(y)\| \leq \frac{\sqrt{n}}{\varepsilon}\|x - y\|.$$

We conclude that

$$\sup_{x\in\mathbb{R}^n} \|S_\varepsilon(x - y_1) - S_\varepsilon(x - y_2)\| \leq \frac{\sqrt{n}}{\varepsilon}\|y_1 - y_2\|,$$

for any $y_1, y_2 \in \mathbb{R}^n$. This yields the equicontinuity of $\mathcal{F}^\varepsilon$.

Now let us turn to $D^\varepsilon_{1,k}$. We have that

$$\|D^\varepsilon_{1,k}(x)\| = \left\| \int_{B_\varepsilon(x)} \frac{x - z}{\|x - z\|}\Big(1 - \frac{\|x - z\|}{\varepsilon}\Big)\mathbb{I}[z \neq x]\, dP_k(z) \right\| \leq P_k[B_\varepsilon(x)]$$

for any $x \in \mathbb{R}^n$, any $\varepsilon > 0$ and $k \in \mathbb{N}$. Let us fix $\delta > 0$. Since (P_k) converges in distribution, it is tight. Therefore, there exists $R_\delta > 0$ such that

$$\sup_{k\in\mathbb{N}} P_k[\mathbb{R}^n \setminus B_{R_\delta}] \leq \delta.$$

It follows that

$$\sup_{x\in\mathbb{R}^n\setminus B_{1+R_\delta}} \|D^\varepsilon_{1,k}(x)\| \leq \sup_{x\in\mathbb{R}^n\setminus B_{\varepsilon+R_\delta}} \|D^\varepsilon_{1,k}(x)\| \leq P_k[\mathbb{R}^n \setminus B_{R_\delta}] \leq \delta \tag{2.10.19}$$

for any $\varepsilon \in (0,1)$ and k. In order to have a global control on $D^\varepsilon_{1,k}$ it remains to bound $D^\varepsilon_{1,k}$ over $\overline{B_{1+R_\delta}}$. Since $\overline{B_{1+R_\delta}}$ is compact, there exist $N_\delta \in \mathbb{N}$ and $x_1, \dots, x_{N_\delta} \in B_{1+R_\delta}$ such that

$$\overline{B_{1+R_\delta}} \subset \cup_{j=1}^{N_\delta} B_\varepsilon(x_j).$$

This entails that

$$\sup_{x \in \overline{B_{1+R_\delta}}} \|D^\varepsilon_{1,k}(x)\| \leq \max_{j=1,\dots,N_\delta} \sup_{x \in B_\varepsilon(x_j)} P_k[B_\varepsilon(x)] \leq \max_{j=1,\dots,N_\delta} P_k[B_{2\varepsilon}(x_j)]. \tag{2.10.20}$$

For any $x \in \mathbb{R}^n$, let

$$\mathcal{A}_x := \{\varepsilon > 0 : P[\partial B_{2\varepsilon}(x)] > 0\}.$$

The set $\mathcal{A}_x$ is at most countable for any $x \in \mathbb{R}^n$. Let $\mathcal{E}_\delta := \cup_{j=1}^{N_\delta} \mathcal{A}_{x_j}$. The set $\mathcal{E}_\delta$ is at most countable. For any $\varepsilon \in (0,1) \setminus \mathcal{E}_\delta$, we have

$$P_k[B_{2\varepsilon}(x_j)] \to P[B_{2\varepsilon}(x_j)]$$

for any $j = 1, \dots, N_\delta$ as $k \to \infty$, since (P_k) converges to P in distribution as $k \to \infty$. It follows from (2.10.20) that

$$\limsup_{k\to\infty} \sup_{x \in \overline{B_{1+R_\delta}}} \|D^\varepsilon_{1,k}(x)\| \leq \max_{j=1,\dots,N} P[B_{2\varepsilon}(x_j)] \leq \sup_{x \in B_{1+R_\delta}} P[B_{2\varepsilon}(x)] \tag{2.10.21}$$

for any $\varepsilon \in (0,1) \setminus \mathcal{E}_\delta$. Let us show that

$$\lim_{\varepsilon\to 0} \sup_{x \in B_{1+R_\delta}} P[B_{2\varepsilon}(x)] = 0. \tag{2.10.22}$$

Assume, ad absurdum, that there exists a sequence $(\varepsilon_n) \searrow 0$ and $c > 0$ such that

$$\sup_{x \in B_{1+R_\delta}} P[B_{2\varepsilon_n}(x)] > c$$

for any n. This entails that for any n there exists $x_n \in B_{1+R_\delta}$ such that $P[B_{2\varepsilon_n}(x_n)] > c$. Since $\overline{B_{1+R_\delta}}$ is compact, we may assume (up to extraction of a subsequence) that (x_n) converges to some $x \in \overline{B_{1+R_\delta}}$. Let $r > 0$. We have

$$B_{2\varepsilon_n}(x_n) \subset B_r(x)$$

as soon as $\|x_n - x\| + 2\varepsilon_n < r$. It follows that

$$c \leq P[B_{2\varepsilon_n}(x_n)] \leq P[B_r(x)]$$

for any n large enough such that $\|x_n - x\| + 2\varepsilon_n < r$. Consequently, we have $P[B_r(x)] \geq c$ for any $r > 0$. Taking the limit as $r \searrow 0$ yields $P[\{x\}] \geq c$. This is a contradiction since P is non-atomic. Therefore, (2.10.22) is proved.

Putting (2.10.19), (2.10.21), and (2.10.22) together yields

$$\limsup_{\substack{\varepsilon\searrow 0 \\ \varepsilon \notin \mathcal{E}_\delta}} \limsup_{k\to\infty} \sup_{x \in \mathbb{R}^n} \|D^\varepsilon_{1,k}(x)\| \leq \delta. \tag{2.10.23}$$

Finally, we turn to $D^\varepsilon_{3,k}$. Let $\varepsilon \in (0,1)$ and $x \in \mathbb{R}^n$ be fixed. We have

$$\|D^\varepsilon_{3,k}(x)\| = \left\| \int_{B_\varepsilon(x)} \frac{x-z}{\|x-z\|}\Big(1 - \frac{\|x-z\|}{\varepsilon}\Big)\mathbb{I}[z \neq x]\, dP(z) \right\| \leq P[B_\varepsilon(x)].$$

The same reasoning we used for $D^\varepsilon_{1,k}$ yields

$$\lim_{\varepsilon \searrow 0} \limsup_{k\to\infty} \sup_{x\in\mathbb{R}^n} \|D^\varepsilon_{3,k}(x)\| = 0. \tag{2.10.24}$$

Putting (2.10.18), (2.10.23), and (2.10.24) together yields

$$\limsup_{k\to\infty} \sup_{x\in\mathbb{R}^n} \|R_{P_k}(x) - R_P(x)\| = \limsup_{\substack{\varepsilon\searrow 0\\ \varepsilon\notin\mathcal{E}_\delta}} \limsup_{k\to\infty} \sup_{x\in\mathbb{R}^n} \|R_{P_k}(x) - R_P(x)\| \leq \delta. \tag{2.10.25}$$

Since $\delta > 0$ was arbitrary, we conclude that

$$\sup_{x\in\mathbb{R}^n} \|R_{P_k}(x) - R_P(x)\| \to 0$$

as $k \to \infty$. ■

The previous result can not be strengthened in general without further assumptions. Indeed, if (P_k) is a sequence of non-atomic probability measures, then (R_{P_k}) is a sequence of continuous functions. If the sequence (R_{P_k}) converges to R_P uniformly over $\mathbb{R}^n$, then R_P must be continuous as well, i.e. P must be non-atomic. However, by imposing further requirements on the sequence (P_k) and the limit distribution P, we extend the previous result to atomic probability measures in the next proposition.

Proposition 2.10.2. *Let P be an atomic probability measure over $\mathbb{R}^n$. Let (P_k) be a sequence of atomic probability measures over $\mathbb{R}^n$ that converges to P in distribution as $k \to \infty$. If $P_k[\{x\}] \to P[\{x\}]$ for any $x \in \mathbb{R}^n$ such that $P[\{x\}] > 0$, then*

$$\sup_{x\in\mathbb{R}^n} \|R_{P_k}(x) - R_P(x)\| \to 0$$

as $k \to \infty$.

Proof of Proposition 2.10.2. Let $S' := \{x \in \mathbb{R}^n : P[\{x\}] > 0\}$ denote the support of P and

$$S_k := \{x \in \mathbb{R}^n : P_k[\{x\}] > 0\}$$

denote the support of P_k for any $k \in \mathbb{N}$. Since S_k is at most countable for any k, we may see (P_k) and P as atomic measures over the at most countable set $S := S' \cup (\cup_{k\in\mathbb{N}} S_k)$. Let us write $S := \{a_0, a_1, \ldots\}$.

For any $x \in \mathbb{R}^n$ and $k \in \mathbb{N}$, we have

$$R_P(x) = \int_{\mathbb{R}^n} \frac{x-z}{\|x-z\|}\mathbb{I}[z \neq x]\, dP(z) = \sum_{\ell\in\mathbb{N}} \frac{x-a_\ell}{\|x-a_\ell\|}\mathbb{I}[a_\ell \neq x]P[\{a_\ell\}]$$

and

$$R_{P_k}(x) = \sum_{\ell\in\mathbb{N}} \frac{x-a_\ell}{\|x-a_\ell\|}\mathbb{I}[a_\ell \neq x]P_k[\{a_\ell\}].$$

By assumption, we have $P_k[\{a_\ell\}] \to P[\{a_\ell\}]$ as $k \to \infty$ if $P[\{a_\ell\}] > 0$. If $P[\{a_l\}] = 0$, then $P_k[\{a_\ell\}] \to P[\{a_\ell\}]$ since (P_k) converges in distribution. It follows that $P_k[\{a_\ell\}] \to P[\{a_\ell\}]$ for any $\ell \in \mathbb{N}$ as $k \to \infty$. Noticing that

$$\sum_{\ell \in \mathbb{N}} P_k[\{a_\ell\}] = 1 = \sum_{\ell \in \mathbb{N}} P[\{a_\ell\}]$$

for any $k \in \mathbb{N}$, Scheffé's lemma entails that

$$\sum_{\ell \in \mathbb{N}} |P_k[\{a_\ell\}] - P[\{a_\ell\}]| \to 0$$

as $k \to \infty$. Next observe that

$$\sup_{x \in \mathbb{R}^n} \|R_{P_k}(x) - R_P(x)\| \le \sum_{\ell \in \mathbb{N}} |P_k[\{a_\ell\}] - P[\{a_\ell\}]|$$

for any k. It follows that

$$\sup_{x \in \mathbb{R}^n} \|R_{P_k}(x) - R_P(x)\| \to 0$$

as $k \to \infty$. ■

Corollary 2.10.3. *Let P be a probability measure over $\mathbb{R}^n$. Let $X_1, X_2, \dots$ be a random sample from P. For any k, let P_k denote the empirical measure*

$$P_k := \frac{1}{k} \sum_{j=1}^{k} \delta_{X_j}.$$

Then,

$$\sup_{x \in \mathbb{R}^n} \|R_{P_k}(x) - R_P(x)\| \to 0$$

with P-probability 1 as $k \to \infty$.

Proof of Theorem 2.10.3. Let $\mathcal{B}_n$ denote the Borel σ-algebra of $\mathbb{R}^n$. First, let us show that $P_k[E] \to P[E]$ for any $E \in \mathcal{B}_n$ with P-probability 1 as $k \to \infty$. Consider the semiring of sets

$$\mathcal{A} := \left\{ \prod_{j=1}^{n} (a_j, b_j] : -\infty < a_j \le b_j < \infty, a_j \in \mathbb{Q}, b_j \in \mathbb{Q} \right\}.$$

For any $E \in \mathcal{A}$, the strong law of large numbers entails that $P_k[E] \to P[E]$ with P-probability 1 as $k \to \infty$. Because $\mathcal{A}$ is a countable collection we have, with P-probability 1, that $P_k[E] \to P[E]$ for any $E \in \mathcal{A}$ as $k \to \infty$. Consequently, Theorem 2.5 in [4] entails that, with P-probability 1, we have $P_k[E] \to P[E]$ for any $E \in \mathcal{B}_n$ as $k \to \infty$. In particular, (P_k) converges to P in distribution with P-probability 1 as $k \to \infty$.

In what follows, we reason pathwise in the underlying probability space $(\Omega, \mathbb{P})$ on which the random sequence of measures (P_k) is defined. Let us fix $\omega \in \Omega$ such that

$P_k(\omega)[E] \to P[E]$ for any $E \in \mathcal{B}_n$ as $k \to \infty$. In particular, $(P_k)(\omega)$ converges to P in distribution as $k \to \infty$. For the sake of simplicity, we will omit the dependence of (P_k) on ω. If P is non-atomic, Theorem 2.10.1 entails that (R_{P_k}) converges to R_P uniformly over $\mathbb{R}^n$ as $k \to \infty$. If P is atomic, Proposition 2.10.2 entails that (R_{P_k}) converges to R_P uniformly over $\mathbb{R}^n$ as $k \to \infty$, since (P_k) is a sequence of atomic probability measures. Then let A be the set of atoms of P, and assume that $P[A] \in (0,1)$ (notice that $A \in \mathcal{B}_n$ because A is at most countable). Because $P_k[E] \to P[E]$ for any $E \in \mathcal{B}_n$ as $k \to \infty$, we have $P_k[A] \to P[A]$ and $P_k[A^c] \to P[A^c]$, where we let $A^c := \mathbb{R}^n \setminus A$. In particular we may assume, up to taking k large enough, that $P_k[A] > 0$ and $P_k[A^c] > 0$ for any k.

For any probability measure Q such that $Q[A] > 0$, let

$$Q_{|A}[E] := \frac{Q(A \cap E)}{Q[A]}$$

for any $E \in \mathcal{B}_n$ denote the probability measure Q conditioned on A. Then, we may write

$$P_k = P_k[A] P_{k|A} + P_k[A^c] P_{k|A^c},$$

for any k. It follows that

$$R_{P_k}(x) = P_k[A] R_{P_{k|A}}(x) + P_k[A^c] R_{P_{k|A^c}}(x)$$

for any $x \in \mathbb{R}^n$ and k.

Because $P_k[E] \to P[E]$ for any $E \in \mathcal{B}_n$ as $k \to \infty$, we have $P_{k|A}[E] \to P_{|A}[E]$ and $P_{k|A^c}[E] \to P_{|A^c}[E]$ for any $E \in \mathcal{B}_n$ as $k \to \infty$. In particular, $(P_{k|A})$ converges to $P_{|A}$ in distribution and $(P_{k|A^c})$ converges in distribution to $P_{|A^c}$ as $k \to \infty$. On the one hand, Proposition 2.10.2 entails that $(R_{P_{k|A}})$ converges to $R_{P_{|A}}$ uniformly over $\mathbb{R}^n$ as $k \to \infty$, since $P_{k|A}$ and $P_{|A}$ are atomic. On the other hand, Theorem 2.10.1 entails that $(R_{P_{k|A^c}})$ converges to $R_{P_{|A^c}}$ uniformly over $\mathbb{R}^n$ as $k \to \infty$, since $P_{|A^c}$ is non-atomic. Because $P_k[A] \to P[A]$ and $P_k[A^c] \to P[A^c]$ as $k \to \infty$, we conclude that (R_{P_k}) converges to R_P uniformly over $\mathbb{R}^n$ as $k \to \infty$. ∎

2.11 Appendix: proofs

Auxiliary proofs for Section 2.4.2

Proof of Corollary 2.4.9. Let $x_0 \in \Omega$ and $r > 0$ be such that $\overline{B(x_0,r)} \subset \Omega$. For any $x \in B_1$, let $\tilde{u}(x) := u(\frac{x-x_0}{r})$ and $\tilde{f}(x) := \frac{1}{r^2} f(\frac{x-x_0}{r})$. Since $\Delta u = f$ in the weak sense in $B(x_0,r)$, a direct computation entails that $\Delta\tilde{u} = \tilde{f}$ in the weak sense in B_1. Since $u \in H^1(B(x_0,r)) \cap L^\infty(B(x_0,r))$ and $f \in \mathscr{C}^{k,\alpha}(B(x_0,r))$, we have $\tilde{u} \in H^1(B_1) \cap L^\infty(B_1)$, and $\tilde{f} \in \mathscr{C}^{k,\alpha}(B_1)$. It follows from Proposition 2.4.8 that $\tilde{u} \in \mathscr{C}^{k+2,\alpha}(B_1)$. This implies that $u \in \mathscr{C}^{k+2,\alpha}(B(x_0,r))$. Now let $V \subset \Omega$ be an open subset such that $\overline{V} \subset \Omega$. Then V can be covered by a finite number of balls of the form $B(x_0,r)$ with $\overline{B(x_0,r)} \subset \Omega$. Since u is of class $\mathscr{C}^{k+2,\alpha}$ on each one of these balls, we have $u \in \mathscr{C}^{k+2,\alpha}(V)$. We conclude that $u \in \mathscr{C}^{k+2,\alpha}_{\mathrm{loc}}(\Omega)$. ∎

Auxiliary proofs for Section 2.5

PROOF OF PROPOSITION 2.5.1. Observe that the map $\xi \mapsto (1+|\xi|)^m|\xi|^{2s}(\mathscr{F}u)(\xi)$ is integrable over $\mathbb{R}^n$ for any $m \geq 0$ since $u \in \mathscr{S}(\mathbb{R}^n)$. This entails that $\mathscr{F}^{-1}(|\xi|^{2s}\mathscr{F}u) \in \mathscr{C}^m(\mathbb{R}^n)$ for any $m \geq 0$, whence $(-\Delta)^s u \in \mathscr{C}^\infty(\mathbb{R}^n)$. It remains to show that (2.5.3) holds for any $\alpha \in \mathbb{N}^n$. Observe that

$$\begin{aligned}\partial^\alpha(-\Delta)^s u &= \partial^\alpha \mathscr{F}^{-1}(|\xi|^{2s}\widehat{u}) \\ &= \mathscr{F}^{-1}\Big((2i\pi\xi)^\alpha|\xi|^{2s}\mathscr{F}u\Big) \\ &= \mathscr{F}^{-1}\Big(|\xi|^{2s}\mathscr{F}(\partial^\alpha u)\Big) \\ &= (-\Delta)^s(\partial^\alpha u)\end{aligned}$$

for any $\alpha \in \mathbb{N}^n$. Since $\partial^\alpha u$ belongs to $\mathscr{S}(\mathbb{R}^n)$ for any $\alpha \in \mathbb{N}^n$, it is enough to show that

$$\sup_{x\in\mathbb{R}^n} |(1+|x|^{n+2s})(-\Delta)^s u(x)| \lesssim |u|_{L^1(\mathbb{R}^n)} + \sup_{z\in\mathbb{R}^n}\Big((1+|z|)^{n+2}|\nabla^2 u(z)|\Big). \tag{2.11.26}$$

By simple changes of variable, it is easy to show that

$$(-\Delta)^s u(x) = -\frac{1}{2}c_{n,s}\int_{\mathbb{R}^n}\frac{u(x+y)+u(x-y)-2u(x)}{|y|^{n+2s}}\,dy$$

for any $x \in \mathbb{R}^n$; see, e.g., Lemma 3.2 in [78]. Notice that this last integral is not singular at $y=0$ anymore. Indeed, one can easily show that

$$\frac{u(x+y)+u(x-y)-2u(x)}{|y|^{n+2s}} = \frac{1}{|y|^{n+2s}}\int_{-1}^{1}(y, \nabla^2 u(x+ty)y)\,dt. \tag{2.11.27}$$

The r.h.s. of (2.11.27) is then bounded by

$$\frac{|\nabla^2 u|_{L^\infty(\mathbb{R}^n)}}{|y|^{n+2s-2}},$$

which is integrable near the origin. Let us first show that

$$\sup_{x\in\mathbb{R}^n}|(-\Delta)^s u(x)| \lesssim |u|_{L^1(\mathbb{R}^n)} + \sup_{z\in\mathbb{R}^n}\Big((1+|z|)^{n+2}|\nabla^2 u(z)|\Big). \tag{2.11.28}$$

Let us fix $x \in \mathbb{R}^n$ and write

$$\begin{aligned}-\frac{2}{c_{n,s}}(-\Delta)^s u(x) &= \int_{\mathbb{R}^n\setminus B_1}\frac{u(x+y)+u(x-y)-2u(x)}{|y|^{n+2s}}\,dy \\ &\qquad + \int_{B_1}\frac{u(x+y)+u(x-y)-2u(x)}{|y|^{n+2s}}\,dy \\ &=: I_1(x)+I_2(x).\end{aligned}$$

For I_1, we have

$$|I_1(x)| \leq 4|u|_{L^1(\mathbb{R}^n)}.$$

For I_2, recalling (2.11.27), we have

$$\begin{aligned}|I_2(x)| &\le \int_{B_1} \frac{1}{|y|^{n+2s-2}} \int_{-1}^{1} |\nabla^2 u(x+ty)|\,dt\,dy \\ &\lesssim \sup_{z\in\mathbb{R}^n} |\nabla^2 u(z)| \\ &\le \sup_{z\in\mathbb{R}^n} \Big((1+|z|)^{n+2}|\nabla^2 u(z)|\Big).\end{aligned}$$

This yields (2.11.28). Let us now show that

$$\sup_{x\in\mathbb{R}^n} \Big(|x|^{n+2s}|(-\Delta)^s u(x)|\Big) \lesssim |u|_{L^1(\mathbb{R}^n)} + \sup_{z\in\mathbb{R}^n} (1+|z|)^{n+2}|\nabla^2 u(z)|. \tag{2.11.29}$$

Let us fix $x \in \mathbb{R}^n$ and write

$$\begin{aligned}-\frac{2}{c_{n,s}}|x|^{n+2s}(-\Delta)^s u(x) &= |x|^{n+2s}\int_{\mathbb{R}^n\setminus B_{\frac{1}{2}|x|}} \frac{u(x+y)+u(x-y)-2u(x)}{|y|^{n+2s}}\,dy \\ &\quad +|x|^{n+2s}\int_{B_{\frac{1}{2}|x|}} \frac{u(x+y)+u(x-y)-2u(x)}{|y|^{n+2s}}\,dy \\ &=: J_1(x)+J_2(x).\end{aligned}$$

For J_1, we have

$$|J_1(x)| \le 4|u|_{L^1(\mathbb{R}^n)}2^{n+2s}.$$

For J_2, recalling (2.11.27), we have

$$\begin{aligned}|J_2(x)| &\le |x|^{n+2s}\int_{B_{\frac{1}{2}|x|}} \frac{1}{|y|^{n+2s-2}}\int_{-1}^{1} |\nabla^2 u(x+ty)|\,dt\,dy \\ &= \int_{B_{\frac{1}{2}|x|}} \frac{1}{|y|^{n+2s-2}}\int_{-1}^{1} \left(\frac{|x|}{|x+ty|}\right)^{n+2s} |x+ty|^{n+2s}|\nabla^2 u(x+ty)|\,dt\,dy.\end{aligned}$$

For any $t \in [-1,1]$ and y with $|y| \le \frac{1}{2}|x|$, we have

$$\frac{|x|}{|x+ty|} \le \frac{|x|}{|x|-|t||y|} \le 2.$$

For any k, let

$$C_k(u) := \sup_{x\in\mathbb{R}^n} (1+|z|)^k|\nabla^2 u(z)| < \infty.$$

We then have

$$|x+ty|^{n+2s}|\nabla^2 u(x+ty)| \le \frac{C_{N+n+2s}(u)}{(1+|x+ty|)^N} \le \frac{C_{N+n+2s}(u)}{(1+\frac{1}{2}|x|)^N}$$

for any N, $|y| \le \frac{1}{2}|x|$, and $t \in [-1,1]$. Let us fix $N = 2-2s$. We then have

$$|J_2(x)| \le 2^{n+2s}C_{n+2}(u)\frac{1}{(1+\frac{1}{2}|x|)^{2-2s}}\int_{B_{\frac{1}{2}|x|}} \frac{1}{|y|^{n+2s-2}}\,dy.$$

Furthermore, it is easy to see that

$$\int_{B_{\frac{1}{2}|x|}} \frac{1}{|y|^{n+2s-2}}\,dy \lesssim |x|^{2-2s}.$$

It follows that

$$\sup_{x\in\mathbb{R}^n} |J_2(x)| \lesssim \sup_{x\in\mathbb{R}^n} (1+|z|)^{n+2}|\nabla^2 u(z)|.$$

We deduce that

$$\sup_{x\in\mathbb{R}^n} \Big(|x|^{n+2s}|(-\Delta)^s u(x)|\Big) \lesssim |u|_{L^1(\mathbb{R}^n)} + \sup_{x\in\mathbb{R}^n} (1+|z|)^{n+2}|\nabla^2 u(z)|,$$

which establishes (2.11.29). Putting (2.11.28) and (2.11.29) together yields (2.11.26), which concludes the proof. ■

Proof of Proposition 2.5.4. 1. Since $u \in L^2(\mathbb{R}^n)$, we have $u \in \mathscr{S}_s(\mathbb{R}^n)'$. In particular, $(-\Delta)^s u$ is a well-defined tempered distribution. Let $\psi \in \mathscr{S}(\mathbb{R}^n)$. We have

$$\langle(-\Delta)^s u, \psi\rangle = \langle u, (-\Delta)^s\psi\rangle = (2\pi)^{2s}\int_{\mathbb{R}^n} u(x)\mathscr{F}^{-1}(|\xi|^{2s}\mathscr{F}\psi)(x)\,dx.$$

Since $u \in L^2(\mathbb{R}^n)$ and $|\xi|^{2s}\mathscr{F}\psi \in L^2(\mathbb{R}^n)$, we have

$$\langle(-\Delta)^s u, \psi\rangle = (2\pi)^{2s}\int_{\mathbb{R}^n} (\mathscr{F}^{-1}u)(\xi)|\xi|^{2s}(\mathscr{F}\psi)(\xi)\,d\xi$$

by interchanging the inverse Fourier transform under the integral. Since $u \in H^{2s}(\mathbb{R}^n)$, we have $|\xi|^{2s}\mathscr{F}^{-1}u \in L^2(\mathbb{R}^n)$. Since $\psi \in L^2(\mathbb{R}^n)$, interchanging the Fourier transform again yields

$$\langle(-\Delta)^s u, \psi\rangle = (2\pi)^{2s}\int_{\mathbb{R}^n} \mathscr{F}(|\xi|^{2s}(\mathscr{F}^{-1}u))(x)\psi(x)\,dx.$$

Observing that

$$\mathscr{F}(|\xi|^{2s}(\mathscr{F}^{-1}u)) = \mathscr{F}^{-1}(|\xi|^{2s}(\mathscr{F}u))$$

yields

$$\langle(-\Delta)^s u, \psi\rangle = \int_{\mathbb{R}^n} (2\pi)^{2s}\mathscr{F}^{-1}(|\xi|^{2s}(\mathscr{F}u))(x)\psi(x)\,dx.$$

Since $\psi \in \mathscr{S}(\mathbb{R}^n)$ was arbitrary, the conclusion follows.

2. Observe that $|\xi|^{2s}\mathscr{F}u \in \mathscr{S}(\mathbb{R}^n)'$ since $|\xi|^{2s}\mathscr{F}u \in L^1(\mathbb{R}^n)$. In particular, $\mathscr{F}^{-1}(|\xi|^{2s}\mathscr{F}u) \in \mathscr{S}(\mathbb{R}^n)'$ as well. Let $\psi \in \mathscr{S}(\mathbb{R}^n)$. We have

$$\begin{aligned}
\big\langle(2\pi)^{2s}\mathscr{F}^{-1}(|\xi|^{2s}\mathscr{F}u), \psi\big\rangle &= \big\langle|\xi|^{2s}\mathscr{F}u, (2\pi)^{2s}\mathscr{F}^{-1}\psi\big\rangle \\
&= \int_{\mathbb{R}^n} |\xi|^{2s}\mathscr{F}u(\xi)(2\pi)^{2s}(\mathscr{F}^{-1}\psi)(\xi)\,d\xi \\
&= \int_{\mathbb{R}^n} \mathscr{F}u(\xi)\mathscr{F}^{-1}\Big((2\pi)^{2s}\mathscr{F}\big(|\xi|^{2s}\mathscr{F}^{-1}\psi\big)\Big)(\xi)\,d\xi \\
&= \int_{\mathbb{R}^n} \mathscr{F}u(\xi)\mathscr{F}^{-1}((-\Delta)^s\psi)(\xi)\,d\xi.
\end{aligned}$$

Since $\psi \in \mathscr{S}(\mathbb{R}^n)$, we have $(-\Delta)^s\psi \in \mathscr{S}_s(\mathbb{R}^n)$. In particular, $(-\Delta)^s\psi \in L^1(\mathbb{R}^n)$. Since $\mathscr{F}u \in L^1_{\text{loc}}(\mathbb{R}^n)$ and $|\xi|^{2s}\mathscr{F}u$, it is clear that $\mathscr{F}u \in L^1(\mathbb{R}^n)$. By "hat skipping", we have

$$\left\langle (2\pi)^{2s}\mathscr{F}^{-1}(|\xi|^{2s}\mathscr{F}u), \psi \right\rangle = \int_{\mathbb{R}^n} \mathscr{F}^{-1}(\mathscr{F}u)(x)((-\Delta)^s\psi)(x)\, dx.$$

Observe that since the tempered distribution $\mathscr{F}u$ is a function of $L^1(\mathbb{R}^n)$, we have $\mathscr{F}^{-1}(\mathscr{F}u)$ is a continuous function. It is further easy to see that $\mathscr{F}^{-1}(\mathscr{F}u) = u$ almost everywhere on $\mathbb{R}^n$, although u might not be integrable. We conclude that

$$\left\langle (2\pi)^{2s}\mathscr{F}^{-1}(|\xi|^{2s}\mathscr{F}u), \psi \right\rangle = \int_{\mathbb{R}^n} u(x)((-\Delta)^s\psi)(x)\, dx = \langle (-\Delta)^s u, \psi \rangle$$

for any $\psi \in \mathscr{S}(\mathbb{R}^n)$.

3. This is proved in [94]; see Proposition 2.4. ∎

Auxiliary proofs for Section 2.6

Proof of Lemma 2.6.2. The proof follows the same lines as [53] and [5] (see Equation (1.1.1) in [53] and Theorem 56 in [5]). Let $g \in L^1(\mathbb{R}^n)$ be such that $g(x) = h(|x|)$ for any $x \in \mathbb{R}^n$. We first prove that

$$\widehat{g}(\xi) = \frac{1}{(2\pi)^{\frac{n}{2}}|\xi|^n} \int_0^\infty r^{\frac{n}{2}} h\Big(\frac{r}{2\pi|\xi|}\Big) J_{\frac{n-2}{2}}(r)\, dr \tag{2.11.30}$$

for any $\xi \in \mathbb{R}^n \setminus \{0\}$, where J_ν is the Bessel function of the first kind of order ν. Let us therefore fix $\xi \in \mathbb{R}^n \setminus \{0\}$. Observe that

$$\widehat{g}(\xi) = \int_{\mathbb{R}^n} g(x) e^{-2i\pi(x,\xi)}\, dx = \int_{\mathbb{R}^n} g(x) e^{-2i\pi(x,O\xi)}\, dx$$

for any $n \times n$ orthogonal matrix, since $g(x) = h(|x|)$ for any x. Let us then assume that $\xi = |\xi|(1, 0, \ldots, 0)$ and let us compute

$$\begin{aligned}
\int_{\mathbb{R}^n} g(x) e^{-2i\pi(x,\xi)}\, dx &= \int_{\mathbb{R}^n} h(|x|) e^{-2i\pi|\xi|x_1}\, dx \\
&= \int_{\mathbb{R}} e^{-2i\pi|\xi|x_1} \bigg(\int_{\mathbb{R}^{n-1}} h\Big(\sqrt{x_1^2 + |y|^2}\Big)\, dy \bigg)\, dx_1 \\
&= S_{n-2} \int_{\mathbb{R}} e^{-2i\pi|\xi|x_1} \bigg(\int_0^\infty t^{n-2} h\Big(\sqrt{x_1^2 + t^2}\Big)\, dt \bigg)\, dx_1,
\end{aligned}$$

where $S_{n-2} = 2\pi^{\frac{n-1}{2}}/\Gamma(\frac{n-1}{2})$ is the surface area of the $(n-2)$-dimensional unit sphere in $\mathbb{R}^{n-1}$. Let us now write (x_1, t) into spherical coordinates in the plane

$$(x_1, t) = (r\cos\theta, r\sin\theta),$$

with $r \in [0, \infty)$ and $\theta \in [0, \pi]$, since $t > 0$. We then have

$$\begin{aligned}
\widehat{g}(\xi) &= S_{n-2} \int_0^\infty \bigg(\int_0^\pi e^{-2i\pi|\xi|r\cos\theta} (r\sin\theta)^{n-2} h(r)\, r d\theta \bigg)\, dr \\
&= S_{n-2} \int_0^\infty r^{n-1} h(r) \bigg(\int_0^\pi e^{-2i\pi|\xi|r\cos\theta} (\sin\theta)^{n-2} d\theta \bigg)\, dr.
\end{aligned}$$

Let us now express $\int_0^\pi e^{-2i\pi|\xi|r\cos\theta}(\sin\theta)^{n-2}d\theta$ in terms of Bessel functions. For any $\nu \in \mathbb{C}$ with $\mathrm{Re}(\nu) > -1/2$, the Bessel function of the first kind of order ν can be computed as

$$J_\nu(x) = \frac{(\frac{x}{2})^\nu}{\sqrt{\pi}\,\Gamma(\nu+\frac{1}{2})}\int_{-1}^{1}(1-t^2)^{\nu-\frac{1}{2}}\cos(xt)\,dt = \frac{(\frac{x}{2})^\nu}{\sqrt{\pi}\,\Gamma(\nu+\frac{1}{2})}\int_{-1}^{1}(1-t^2)^{\nu-\frac{1}{2}}e^{-ixt}\,dt,$$

for any $x \in \mathbb{R}$ (see (10.9.4) in [82]). Substituting $t = \cos\theta$ leads to

$$J_\nu(x) = \frac{(\frac{x}{2})^\nu}{\sqrt{\pi}\,\Gamma(\nu+\frac{1}{2})}\int_0^\pi(\sin\theta)^{2\nu}e^{-ix\cos\theta}\,d\theta$$

for any $x \in \mathbb{R}$. It follows that

$$\int_0^\pi e^{-2i\pi|\xi|r\cos\theta}(\sin\theta)^{n-2}d\theta = \frac{\sqrt{\pi}\,\Gamma(\frac{n-1}{2})}{(\pi|\xi|r)^{\frac{n-2}{2}}}J_{\frac{n-2}{2}}(2\pi|\xi|r)$$

for any $r > 0$. We deduce that

$$\begin{aligned}\widehat{g}(\xi) &= \frac{2\pi}{|\xi|^{\frac{n-2}{2}}}\int_0^\infty r^{\frac{n}{2}}h(r)J_{\frac{n-2}{2}}(2\pi|\xi|r)\,dr\\ &= \frac{1}{(2\pi)^{\frac{n}{2}}|\xi|^n}\int_0^\infty r^{\frac{n}{2}}h\Big(\frac{r}{2\pi|\xi|}\Big)J_{\frac{n-2}{2}}(r)\,dr,\end{aligned}$$

for any $\xi \in \mathbb{R}^n \setminus \{0\}$, which yields (2.11.30). Let $\alpha \in (0, \frac{n+1}{2})$. For any $k \in \mathbb{N}$, let $g_{k,\alpha}(x) := |x|^{\alpha-n}\mathbb{I}[0 < |x| < k]$ for any $x \in \mathbb{R}^n$, and $h_{k,\alpha}(t) = t^{\alpha-n}\mathbb{I}[0 < t < k]$ for any $t > 0$ so that $g_{k,\alpha}(x) = h_{k,\alpha}(|x|)$ for any $x \in \mathbb{R}^n$. For any k, we have $g_{k,\alpha} \in L^1(\mathbb{R}^n)$. For any $\xi \in \mathbb{R}^n$, applying (2.11.30) to $g_{k,\alpha}$ leads to

$$\widehat{g_{k,\alpha}}(\xi) = \frac{1}{(2\pi)^{\alpha-\frac{n}{2}}|\xi|^\alpha}\int_0^{2\pi|\xi|k} r^{\alpha-\frac{n}{2}}J_{\frac{n-2}{2}}(r)\,dr.$$

According to (10.22.43) in [82], this integral converges to

$$\int_0^\infty r^{\alpha-\frac{n}{2}}J_{\frac{n-2}{2}}(r)\,dr = 2^{\alpha-\frac{n}{2}}\frac{\Gamma(\frac{\alpha}{2})}{\Gamma(\frac{n-\alpha}{2})}$$

as $k \to \infty$, since $\alpha \in (0, \frac{n+1}{2})$. It follows that

$$\lim_{k\to\infty}\widehat{g_{k,\alpha}}(\xi) = \frac{\pi^{\frac{n}{2}-\alpha}\Gamma(\frac{\alpha}{2})}{\Gamma(\frac{n-\alpha}{2})}\,\frac{1}{|\xi|^\alpha}$$

for any $\xi \in \mathbb{R}^n \setminus \{0\}$, with $|\widehat{g_{k,\alpha}}(\xi)| \lesssim \frac{1}{|\xi|^\alpha}$ for any $\xi \in \mathbb{R}^n \setminus \{0\}$, uniformly in k. Let $\psi \in \mathscr{S}(\mathbb{R}^n)$. Observe that

$$|\widehat{g_{k,\alpha}}(\xi)\psi(\xi)| \lesssim \frac{|\psi(\xi)|}{|\xi|^\alpha}$$

uniformly in k, where $|\xi|^{-\alpha}|\psi(\xi)| \in L^1(\mathbb{R}^n)$ since $\alpha < n$ (recall that $\alpha < \frac{n+1}{2}$ and that $n \geq 2$) and $\psi \in \mathscr{S}(\mathbb{R}^n)$. Since $\widehat{g_{k,\alpha}}$ converges almost everywhere over $\mathbb{R}^n$, the dominated convergence theorem entails that

$$\int_{\mathbb{R}^n}\widehat{g_{k,\alpha}}(\xi)\psi(\xi)\,d\xi \to \frac{\pi^{\frac{n}{2}-\alpha}\Gamma(\frac{\alpha}{2})}{\Gamma(\frac{n-\alpha}{2})}\int_{\mathbb{R}^n}\frac{1}{|\xi|^\alpha}\psi(\xi)\,d\xi$$

as $k \to \infty$. On the other hand, for any k, we have

$$\int_{\mathbb{R}^n} \widehat{g_{k,\alpha}}(\xi)\psi(\xi)\,d\xi = \int_{\mathbb{R}^n} g_{k,\alpha}(\xi)\widehat{\psi}(\xi)\,d\xi,$$

since $g_{k,\alpha} \in L^1(\mathbb{R}^n)$. Similarly, we have $|g_{k,\alpha}(x)\widehat{\psi}(x)| \leq |x|^{\alpha-n}|\psi(x)| \in L^1(\mathbb{R}^n)$, uniformly in k, and we have

$$\int_{\mathbb{R}^n} g_{k,\alpha}(\xi)\widehat{\psi}(\xi)\,d\xi \to \int_{\mathbb{R}^n} \frac{1}{|x|^{n-\alpha}}\widehat{\psi}(\xi)\,d\xi$$

by dominated convergence. It follows that

$$\int_{\mathbb{R}^n} \frac{1}{|x|^{n-\alpha}}\widehat{\psi}(\xi)\,d\xi = \frac{\pi^{\frac{n}{2}-\alpha}\Gamma(\frac{\alpha}{2})}{\Gamma(\frac{n-\alpha}{2})} \int_{\mathbb{R}^n} \frac{1}{|\xi|^\alpha}\psi(\xi)\,d\xi$$

for any $\psi \in \mathscr{S}\mathbb{R}^n)$. We deduce that

$$\mathscr{F}\Big(\frac{1}{|x|^{n-\alpha}}\Big)(\xi) = \frac{\pi^{\frac{n}{2}-\alpha}\Gamma(\frac{\alpha}{2})}{\Gamma(\frac{n-\alpha}{2})}\frac{1}{|\xi|^\alpha} \tag{2.11.31}$$

in $\mathscr{S}(\mathbb{R}^n)'$ for any $\alpha \in (0, \frac{n+1}{2})$. Taking the inverse Fourier transform on both sides of (2.11.31) yields

$$\mathscr{F}\Big(\frac{1}{|x|^\alpha}\Big)(\xi) = \frac{\Gamma(\frac{n-\alpha}{2})}{\pi^{\frac{n}{2}-\alpha}\Gamma(\frac{\alpha}{2})}\frac{1}{|\xi|^{n-\alpha}} \tag{2.11.32}$$

in $\mathscr{S}(\mathbb{R}^n)'$ for any $\alpha \in (0, \frac{n+1}{2})$. Now let $\beta \in (\frac{n-1}{2}, n)$ and let us write $\beta = n - \alpha$ for some $\alpha \in (0, \frac{n+1}{2})$. Then, (2.11.31) yields

$$\mathscr{F}\Big(\frac{1}{|x|^\beta}\Big)(\xi) = \frac{\pi^{\frac{n}{2}-\alpha}\Gamma(\frac{\alpha}{2})}{\Gamma(\frac{n-\alpha}{2})}\frac{1}{|\xi|^\alpha} = \frac{\Gamma(\frac{n-\beta}{2})}{\pi^{\frac{n}{2}-\beta}\Gamma(\frac{\beta}{2})}\frac{1}{|\xi|^{n-\beta}} \tag{2.11.33}$$

in $\mathscr{S}(\mathbb{R}^n)'$ for any $\beta \in (\frac{n-1}{2}, n)$. Putting (2.11.32) and (2.11.33) together yields the conclusion for any $\alpha \in (0, \frac{n+1}{2}) \cup (\frac{n-1}{2}, n) = (0, n)$. ■

Proof of Lemma 2.6.3. Let $\psi \in \mathscr{S}(\mathbb{R}^n)$ and let us compute

$$\begin{aligned}\big\langle \nabla(1/|x|^{n-1}), \psi\big\rangle &= -\big\langle 1/|x|^{n-1}, \nabla\psi\big\rangle \\ &= -\int_{\mathbb{R}^n} \frac{1}{|x|^{n-1}}\nabla\psi(x)\,dx \\ &= -\lim_{\substack{R\to\infty \\ \eta\to 0}} \int_{B_R\setminus B_\eta} \frac{1}{|x|^{n-1}}\nabla\psi(x)\,dx,\end{aligned}$$

where the last equality follows by dominated convergence since $x \mapsto 1/|x|^{n-1}$ is integrable near the origin in $\mathbb{R}^n$ and $\nabla\psi \in L^1(\mathbb{R}^n)$ since $\psi \in \mathscr{S}(\mathbb{R}^n)$. Fix $0 < \eta < R < \infty$ We have By Green's formula

$$\begin{aligned}\int_{B_R\setminus B_\eta} \frac{1}{|x|^{n-1}}\nabla\psi(x)\,dx = &-\int_{B_R\setminus B_\eta} \nabla\Big(\frac{1}{|x|^{n-1}}\Big)\psi(x)\,dx \\ &+\int_{\partial B_R} \frac{1}{|x|^{n-1}}\psi(x)\frac{x}{|x|}\,d\sigma_R(x) - \int_{\partial B_\eta} \frac{1}{|x|^{n-1}}\psi(x)\frac{x}{|x|}\,d\sigma_\eta(x),\end{aligned}$$

where σ_r is the surface area measure on the sphere of radius r. Letting $u = x/R$, we find

$$\int_{\partial B_R} \frac{1}{|x|^{n-1}}\psi(x)\frac{x}{|x|}\,d\sigma_R(x) = R^{n-1}\int_{S^{n-1}} \frac{1}{R^{n-1}}\psi(Ru)u\,d\sigma_1(u).$$

Since ψ is bounded and $\psi(x) \to 0$ as $|x| \to \infty$, the latter converges to 0 as $R \to \infty$. Similarly, we have

$$\int_{\partial B_\eta} \frac{1}{|x|^{n-1}}\psi(x)\frac{x}{|x|}\,d\sigma_\eta(x) = \eta^{n-1}\int_{S^{n-1}} \frac{1}{\eta^{n-1}}\psi(\eta u)u\,d\sigma_1(u).$$

As $\eta \to 0$, the last integral converges to $\psi(0)\int_{S^{n-1}} u\,d\sigma_1(u) = 0$. It follows that

$$\begin{aligned}
\big\langle \nabla(1/|x|^{n-1}), \psi \big\rangle &= \lim_{\substack{R\to\infty\\ \eta\to 0}} \int_{B_R\setminus B_\eta} \nabla\Big(\frac{1}{|x|^{n-1}}\Big)\psi(x)\,dx \\
&= -(n-1)\lim_{\substack{R\to\infty\\ \eta\to 0}} \int_{B_R\setminus B_\eta} \frac{x}{|x|^{n+1}}\psi(x)\,dx \\
&= -(n-1)\lim_{\eta\to 0} \int_{\mathbb{R}^n\setminus B_\eta} \frac{x}{|x|^{n+1}}\psi(x)\,dx \\
&= -(n-1)\ \text{P.V.}\Big(\frac{x}{|x|^{n+1}}\Big).
\end{aligned}$$

■

Chapter 3

Multivariate ρ-quantiles: a geometric approach

3.1 Introduction

The concept of quantile, which is of paramount importance in statistics, has long been limited to probability measures over $\mathbb{R}$. Defining a suitable quantile concept in $\mathbb{R}^d$ is a problem that is intrinsically difficult due to the lack of a canonical ordering in $\mathbb{R}^d$, $d > 1$. This has been an active research topic in the last decades; see, among many others, [44], [46], [88], and the references therein. One of the most successful multivariate quantile concepts is the concept of *geometric* quantiles from [17]; for a probability measure P over $\mathbb{R}^d$, the geometric quantile of order α in direction u is defined as the minimizer over $\mathbb{R}^d$ of the map

$$\mu \mapsto M_{\alpha,u}(\mu) = \int_{\mathbb{R}^d} \left\{ \|z-\mu\| - \|z\| - \alpha u'\mu \right\} dP(z), \tag{3.1.1}$$

with $\alpha \in [0,1)$ and $u \in \mathcal{S}^{d-1} := \{z \in \mathbb{R}^d : \|z\|^2 := z'z = 1\}$. The success of geometric quantiles is explained by several key distinctive properties, among which: geometric quantiles are easy to compute, even for large d ([73], [102]). The asymptotic behavior of their sample version is rather standard ([17], [106]). They can easily be extended into regression quantiles ([16], [20]) or turned into quantiles for functional data ([15], [23], [92]).

For $d = 1$, geometric quantiles, that minimize the L_1-objective function in (3.1.1), reduce to the usual univariate quantiles. In particular, the collection of intervals whose endpoints are the geometric quantiles of order α in direction $u = -1$ and $u = 1$ is a nested family of interquantile intervals, that all contain the univariate median (which is obtained with $\alpha = 0$, irrespective of the direction u). Similarly, *expectiles*, an L_2-analog of quantiles introduced in [77], provide a nested family of centrality intervals that all contain the mean of the distribution. Expectiles have met a big success, particularly so in financial risk assessment, where they provide coherent risk measures; see, e.g., [26], [52], [97]. Quantiles and expectiles belong to the class of L_p-quantiles (associated with L_p loss functions, with $p \geq 1$), or, more generally, of M-quantiles (associated with general convex loss functions); see [8], [18]. Recently, there has been a growing interest in such generalized quantiles, still with risk assessment as one of the main applications; see, e.g., [27], [37], or [100].

The success of geometric quantiles in $\mathbb{R}^d$ and the growing interest in L_p-quantiles in the univariate case $d = 1$ suggest to define a geometric concept of L_p-quantiles. For $p = 2$, this has actually recently been done in [47], but the resulting geometric expectiles remain less well understood than their geometric L_1-counterparts from [17]. To the best of our knowledge, a geometric L_p-quantile concept, or, more generally, a geometric M-quantile concept has not been investigated in the literature. In this work, we define such a general concept and we thoroughly study its properties. We adopt the following definition.

Definition 3.1.1. *Let $\rho : \mathbb{R}^+ \to \mathbb{R}^+$ be a convex function and P be a probability measure over $\mathbb{R}^d$. Fix $\alpha \in [0, 1)$ and $u \in \mathcal{S}^{d-1}$. We say that $\mu^\rho_{\alpha,u} = \mu^\rho_{\alpha,u}(P)$ is a* geometric ρ-quantile of order α in direction u for P *if and only if it minimizes the objective function*

$$\mu \mapsto M^\rho_{\alpha,u}(\mu) := \int_{\mathbb{R}^d} \big\{ H^\rho_{\alpha,u}(z - \mu) - H^\rho_{\alpha,u}(z) \big\}\, dP(z) \tag{3.1.2}$$

over $\mathbb{R}^d$, where we let

$$H^\rho_{\alpha,u}(z) := \rho(\|z\|)\left(1 + \alpha \frac{u'z}{\|z\|}\right)\xi_{z,0}, \tag{3.1.3}$$

with $\xi_{z_1,z_2} := \mathbb{I}[z_1 \neq z_2]$ (throughout, $\mathbb{I}[A]$ is the indicator function of A).

It might have been natural to write μ^ρ_v, with $v = \alpha u$ rather than $\mu^\rho_{\alpha,u}$, to emphasize the indexation of geometric ρ-quantiles on the unit ball, but we favour the notation $\mu^\rho_{\alpha,u}$ that stresses the heterogenous roles α and u will play in the sequel. The multivariate L_p-quantiles we consider in this work are obtained with $\rho(t) = t^p$, $p \geq 1$. Clearly, these reduce for $p = 1$ to the minimizers of (3.1.1), that is, to the geometric quantiles from [17]. If P has finite second-order moments, then our L_p-quantiles reduce for $p = 2$ to the expectiles introduced in [47]; as we will show, however, our formulation above only requires that P has finite first-order moments. Note that for $\alpha = 0$, the ρ-quantile $\mu^\rho_{\alpha,u}$, irrespective of u (the direction u does not play any role for $\alpha = 0$), is an M-functional of location, that, for $p = 1$ and $p = 2$, provides the celebrated geometric median ([9]) and the mean vector of P, respectively. The ρ-quantiles from Definition 3.1.1 extend this M-functional of location in the same way the geometric quantiles from [17] extend the geometric median: they are thus *M-quantiles*, in the sense of [8] or [50] (it should be noted that the only intersection between the M-quantiles in Definition 3.1.1 and those from [50] are the geometric quantiles from [17]).

Above, the motivation to consider L_p-quantiles was linked to their relevance for risk assessment, and there is no doubt that geometric L_p-quantiles are natural tools to define suitable risk measures in situations where multidimensional portfolios are considered. The main focus in this work, however, is on a careful study of the probabilistic properties of these L_p-quantiles and, more generally, of the corresponding ρ-quantiles. Quite remarkably, many of these properties crucially depend on the loss function ρ. We provide two examples. (i) Convexity of the objective function in (3.1.2) for any order α and direction u is a key property for the study of ρ-quantiles and for their evaluation at empirical distributions, and it may be expected that this convexity is inherited from the convexity of ρ. Our results, however, will show that, in the class of L_p-quantiles, this is the case if and only if $p \leq 2$. For $p > 2$, we will show that convexity holds for $\alpha \leq \alpha_p$ only, where, quite remarkably, α_p is very close to one for any p but does not depend on p monotonically. (ii) The geometric quantiles from [17] have recently been criticised because they exit any compact set as $\alpha \to 1$ even for a compactly supported probability measure P; see [42]. As our results will show, this behavior of extreme geometric quantiles is shared by L_p-quantiles with $p \leq 2$, but not by those with $p > 2$. While we discussed these results here for L_p-quantiles only, we will throughout study properties of ρ-quantiles for a virtually arbitrary convex loss function ρ, which will allow us to consider, e.g., exponential loss functions or the celebrated Huber loss functions.

The outline of this chapter is as follows. In Section 3.2, we provide the assumptions under which the objective function $M^{\rho}_{\alpha,u}(\mu)$ in (3.1.2) is well-defined for any μ, and we discuss existence of ρ-quantiles. In Section 3.3, we obtain a necessary and sufficient condition for convexity of $M^{\rho}_{\alpha,u}(\mu)$, we characterize the orders α for which convexity fails when this condition is not satisfied, and we exploit this to derive uniqueness results for ρ-quantiles. In Section 3.4, we refine these convexity and uniqueness results in the particular case for which the underlying probability measure is spherically symmetric. In Section 3.5, we study first- and second-order differentiability of the objective function $M^{\rho}_{\alpha,u}(\mu)$, which will play a key role in the subsequent sections. In Section 3.6, we exploit Robert Serfling's DOQR paradigm to define ρ-depth functions, ρ-outlyingness functions and ρ-rank functions associated with our ρ-quantile functions. We also identify conditions under which ρ-quantile functions are homeomorphisms from the open unit ball (quantiles are indexed by $(\alpha, u) \in [0, 1) \times \mathcal{S}^{d-1}$ or, equivalently, by αu in the open unit ball of $\mathbb{R}^d$) to the whole Euclidean space $\mathbb{R}^d$. This will play a major role when studying in Section 3.7 the behavior of extreme ρ-quantiles. In Section 3.8, we derive Bahadur representation results for sample ρ-quantiles and deduce their asymptotic distribution.

3.2 Existence

Throughout, we assume that the loss function ρ belongs to the class $\mathcal{C}$ of functions from $\mathbb{R}^+$ to $\mathbb{R}^+$ that are convex, piecewise twice continuously differentiable on $(0, \infty)$, and satisfy $\rho(t) = 0$ only for $t = 0$. Here, *ρ is piecewise twice continuously differentiable on* $(0, \infty)$ means that either (i) there exist a nonnegative integer K and $(0 =: t_0 <) t_1 < t_2 < \ldots < t_K < t_{K+1} := \infty$ such that ρ is twice continuously differentiable on each open interval (t_k, t_{k+1}), $k = 0, \ldots, K$, or (ii) there exists a monotone strictly increasing sequence $(t_0 := 0, t_1, t_2, \ldots)$ in $\mathbb{R}^+$ diverging to infinity such that ρ is twice continuously differentiable on each open interval (t_k, t_{k+1}), $k \in \mathbb{N}$. We let $\mathcal{D}_\rho = (0, \infty) \setminus \{t_1, t_2, \ldots, t_K\}$ in case (i) and $\mathcal{D}_\rho = (0, \infty) \setminus \{t_1, t_2, \ldots\}$ in case (ii). Examples of loss functions in $\mathcal{C}$ are the power loss functions $\rho(t) = t^p$, with $p \geq 1$, the exponential functions $\rho(t) = \exp(ct) - 1$, with $c > 0$, and the Huber loss functions $\rho(t) = (t^2/2)\mathbb{I}[0 < t < c] + c(t - (c/2))\mathbb{I}[t \geq c]$, with $c > 0$. For the Huber loss functions, $\mathcal{D}_\rho = (0, \infty) \setminus \{c\}$, whereas $\mathcal{D}_\rho = (0, \infty)$ for power and exponential loss functions.

For any $\rho \in \mathcal{C}$, we denote as $\mathcal{P}^{\rho}_d$ the class of probability measures P over $\mathbb{R}^d$ such that for any $\mu \in \mathbb{R}^d$, there exists $\delta > 0$ for which

$$\int_{\mathbb{R}^d} \psi_-(\|z-\mu\| + \delta)\, dP(z) < \infty; \tag{3.2.4}$$

throughout, ψ_- and ψ_+ will denote the left- and right-derivative of ρ, respectively (convexity of ρ ensures existence of these one-sided derivatives). For the power loss function $\rho(t) = t^p$, with $p \geq 1$, $P \in \mathcal{P}_d^\rho$ if and only if P has finite moments of order $p-1$ (that is, $\mathrm{E}[\|Z\|^{p-1}] < \infty$, where Z is a random d-vector with distribution P). For $\rho(t) = t$, $\mathcal{P}_d^\rho$ thus collects all probability measures on $\mathbb{R}^d$, which is also the case for Huber loss functions.

We then have the following existence result (see the appendix for a proof).

Theorem 3.2.1. *Let $\rho \in \mathcal{C}$ and $P \in \mathcal{P}_d^\rho$. Fix $\alpha \in [0,1]$ and $u \in \mathcal{S}^{d-1}$. Then, (i) $M^\rho_{\alpha,u}(\mu)$ is well-defined for any $\mu \in \mathbb{R}^d$; (ii) if $\alpha < 1$, then P admits at least one ρ-quantile of order α in direction u.*

The existence result in Theorem 3.2.1(ii) is obtained by establishing that, for any $\alpha \in [0,1)$ and $u \in \mathcal{S}^{d-1}$, the map $\mu \mapsto M^\rho_{\alpha,u}(\mu)$ is both coercive (in the sense that $M^\rho_{\alpha,u}(\mu)$ diverges to infinity as $\|\mu\|$ does) and continuous on $\mathbb{R}^d$. We require that $\rho \in \mathcal{C}$ in Theorem 3.2.1 to avoid introducing many different collections of loss functions in the sequel, but inspection of the proof reveals that the result actually holds without any differentiability assumption on ρ.

Theorem 3.2.1(ii) shows that, for $\rho(t) = t^p$, with $p \geq 1$, ρ-quantiles exist for any $\alpha \in [0,1)$ and $u \in \mathcal{S}^{d-1}$ as soon as P has finite moments of order $p-1$ (while subtracting $H^\rho_{\alpha,u}(z)$ in the integrand of (3.1.2) in principle has no impact on the corresponding quantile minimizers, not doing so would guarantee existence of a minimizer only under the stronger condition of finite moments of order p). In particular, taking $p = 1$ and $p = 2$, this shows that the quantiles from [17] always exist, whereas their expectile analogs from [47] only require that P has finite first-order moments (as already mentionned, finite second-order moments are imposed in [47]). The quantiles associated with Huber loss functions also always exist for any $\alpha \in [0,1)$ and $u \in \mathcal{S}^{d-1}$.

In Section 3.7 below, we will study extreme ρ-quantiles, that is, the ρ-quantiles indexed by an order α that is arbitrarily close to one. As we will see, the behavior of such quantiles crucially depends on the existence of the boundary ρ-quantiles indexed by an order $\alpha = 1$ (the term "boundary" results from the fact that Definition 3.1.1 imposes that $\alpha \in [0,1)$). Our interest in such boundary quantiles explains why we will be investigating the properties of the map $M^\rho_{\alpha,u}(\mu)$ also for $\alpha = 1$, as we already did in Theorem 3.2.1(ii). At this stage, we stress that Theorem 3.2.1(ii) remains silent about the existence of such boundary ρ-quantiles, the reason being that coercivity of $\mu \mapsto M^\rho_{\alpha,u}(\mu)$ may fail for $\alpha = 1$. For $\rho(t) = t$ and $d \geq 2$, for instance, no such boundary quantiles exist when P is non-atomic and is not supported on a line of $\mathbb{R}^d$; see [42], Proposition 2.1.

We conclude this section with the following orthogonal- and translation-equivariance result, that will be particularly relevant in Section 3.4 when considering the particular case for which P is spherically symmetric (the proof readily follows from Definition 3.1.1).

Proposition 3.2.2. *Let $\rho \in \mathcal{C}$ and $P \in \mathcal{P}_\rho^d$. Fix $\alpha \in [0,1)$ and $u \in \mathcal{S}^{d-1}$. Let O be a $d \times d$ orthogonal matrix, b be a d-vector, and denote as $P_{O,b}$ the distribution of $OZ + b$ when Z has distribution P. Then, if μ is a ρ-quantile of P of order α in direction u, then $O\mu + b$ is a ρ-quantile of $P_{O,b}$ of order α in direction Ou.*

Spatial ρ-quantiles are not affine-equivariant, that is, they fail to be equivariant under general affine transformations. However, they can be made affine-equivariant through a transformation-retransformation approach; see, e.g., [89] in the case of the [17] geometric quantiles.

3.3 Convexity and uniqueness

The quantiles studied in this work are defined as minimizers of the map $\mu \mapsto M_{\alpha,u}^\rho(\mu)$ in (3.1.2). Convexity of this map is a most desirable property, that is expected to play a key role when investigating uniqueness of these quantiles and when evaluating them for empirical probability measures. In this section, we therefore study under which conditions on the loss function $\rho \in \mathcal{C}$ the map $\mu \mapsto M_{\alpha,u}^\rho(\mu)$ is convex for any $P \in \mathcal{P}_d^\rho$.

First note that convexity of $H_{\alpha,u}^\rho$ trivially implies convexity of $M_{\alpha,u}^\rho$, and that, if $H_{\alpha,u}^\rho$ is not convex, then there exists $P \in \mathcal{P}_d^\rho$ for which $M_{\alpha,u}^\rho$ fails to be convex (simply consider a Dirac probability measure). Therefore, we may focus on studying convexity of $H_{\alpha,u}^\rho$. Since any $\rho \in \mathcal{C}$ clearly makes $H_{\alpha,u}^\rho$ convex for $d = 1$, we tacitly restrict throughout this section to the case $d \geq 2$. We start with the following preliminary result showing that the larger α, the fewer the functions ρ making $H_{\alpha,u}^\rho$ convex for any $u \in \mathcal{S}^{d-1}$.

Lemma 3.3.1. *For any $\alpha \in [0,1]$, denote as $\mathcal{C}_\alpha$ the collection of functions $\rho \in \mathcal{C}$ such that $H_{\alpha,u}^\rho$ is convex for any $u \in \mathcal{S}^{d-1}$. Then, we have the following: (i) $\mathcal{C}_0 = \mathcal{C}$; (ii) if $\alpha_1, \alpha_2 \in [0,1]$ satisfy $\alpha_1 < \alpha_2$, then $\mathcal{C}_{\alpha_2} \subseteq \mathcal{C}_{\alpha_1}$.*

This result suggests considering $\alpha_\rho := \max\{\alpha \in [0,1] : \rho \in \mathcal{C}_\alpha\}$, the largest value of α for which ρ makes $H_{\alpha,u}^\rho$ convex for any $u \in \mathcal{S}^{d-1}$ (it is trivial to prove that the maximum exists for any $\rho \in \mathcal{C}$). Ideally, we would like to have that $\alpha_\rho = 1$, as this would ensure that $H_{\alpha,u}^\rho$ is convex for any $\alpha \in [0,1]$ and $u \in \mathcal{S}^{d-1}$. The following result provides a necessary and sufficient condition for $\alpha_\rho = 1$.

Theorem 3.3.2. *Let $\rho \in \mathcal{C}$. Then, irrespective of $d \geq 2$, $\alpha_\rho = 1$ if and only if the map $t \mapsto t^2/\rho(t)$ is concave on $(0,\infty)$.*

As a corollary, the power loss function $\rho(t) = t^p$ makes $H_{\alpha,u}^\rho$ convex for any $\alpha \in [0,1)$ and any $u \in \mathcal{S}^{d-1}$ if and only if $p \in [1,2]$. For $p > 2$, it is then of interest to determine the corresponding value of $\alpha_\rho (< 1)$. More generally, the following result allows one to determine α_ρ for any loss function ρ that does not satisfy the necessary and sufficient condition in Theorem 3.3.2.

Theorem 3.3.3. *Let $\rho \in \mathcal{C}$ be such that the map $t \mapsto t^2/\rho(t)$ is not concave on $(0,\infty)$. Then, irrespective of $d \geq 2$,*

$$\alpha_\rho = \inf_{t \in \mathcal{D}_\rho^{\mathrm{cv}}} \sqrt{\frac{q_t(4p_t^2 - 4p_t - q_t)}{4(p_t - 1)^2(q_t + 1)}} < 1,$$

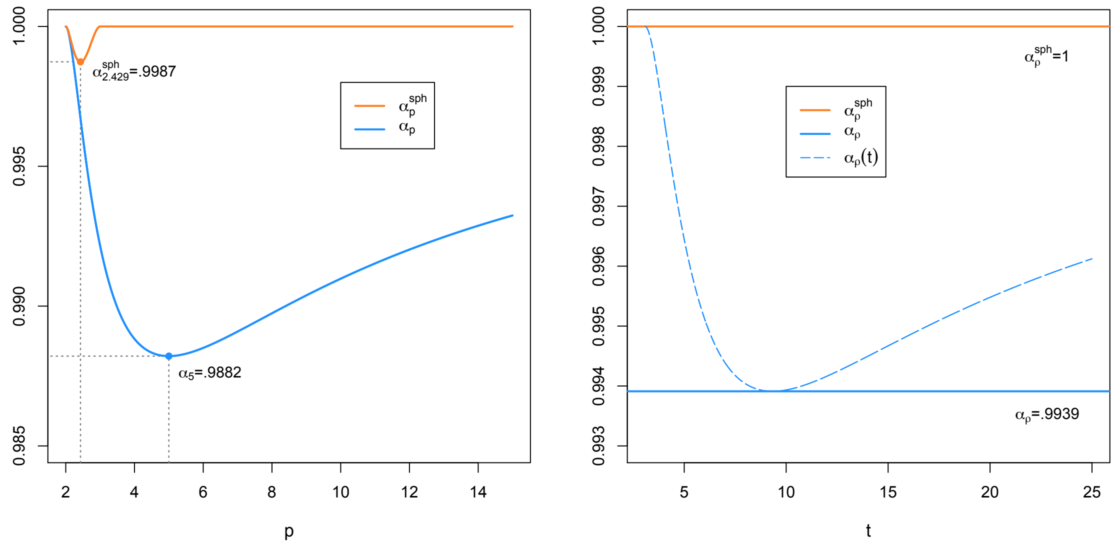

Figure 3.1: (Left:) For power loss functions $\rho(t) = t^p$, plot of α_ρ (see (3.3.5)) and α_ρ^{sph} (see (3.4.6)) (note that (3.3.5) shows that $\alpha_\rho \to 1$ as $p \to \infty$). (Right:) For the loss function $\rho(t) = \exp(t) - 1$, plot of the quantity, $\alpha_\rho(t)$ say, of which the infimum is taken in Theorem 3.3.3; the blue line marks the resulting infimum $\alpha_\rho = .9939$, whereas the orange line stresses that $\alpha_\rho^{\text{sph}} = 1$ (see Section 3.4).

with

$$p_t := \frac{t\psi_-(t)}{\rho(t)} \quad \textit{and} \quad q_t := \frac{t^2\psi'_-(t)}{t\psi_-(t) - \rho(t)},$$

where we let $\mathcal{D}_\rho^{\text{cv}} := \{t \in \mathcal{D}_\rho : (t^2/\rho(t))'' > 0\}$ *and where* ψ'_- *is the left-derivative of* ψ_- *(in this result,* $\psi_-(t)$ *and* $\psi'_-(t)$ *are used only for* $t \in \mathcal{D}_\rho$*, so that we could write* $\psi_-(t) = \rho'(t)$ *and* $\psi'_-(t) = \rho''(t)$ *above).*

For $\rho(t) = t^p$ with $p > 2$, one readily checks that $\mathcal{D}_\rho^{\text{cv}} = (0, \infty)$ and

$$\alpha_\rho := \sqrt{\frac{p^2(4p-5)}{4(p-1)^2(p+1)}}. \tag{3.3.5}$$

Remarkably, α_ρ in (3.3.5) exhibits a non-monotonic pattern in p: for $p \in [2, 5]$, it decreases monotonically from one to its minimal value $\sqrt{125/128}$ (slightly above .9882), then increases monotonically to one again for $p \in [5, \infty)$; see the left panel of Figure 4.1. For $p > 2$, it is thus only for most extreme quantile orders α that convexity fails. This is even more the case for the exponential loss function $\rho(t) = \exp(ct) - 1$, for which $\mathcal{D}_\rho^{\text{cv}} = (3.0861/c, \infty)$ and $\alpha_\rho = .9939$. It can be shown that, if the loss function ρ is such that $t \mapsto \rho(t)/t$ is convex, then $\alpha_\rho \geq \sqrt{2/3} \approx .8165$ (for the sake of completeness, we prove this in the appendix; see Corollary 3.9.9). Like the power loss functions $\rho(t) = t^p$, $p \in (1, 2)$, the Huber loss functions provide a compromise between the L_1 and L_2 loss functions, but since Theorem 3.3.3 entails that $\alpha_\rho = 0$ for the Huber loss functions, power loss functions clearly should be favoured in terms of convexity.

An important corollary of convexity is the following uniqueness result.

Theorem 3.3.4. *Let $\rho \in \mathcal{C}$ and $P \in \mathcal{P}_\rho^d$. Assume that there is no open interval in $(0, \infty)$ on which ψ_- is constant or that P is not concentrated on a line. Then, for any $\alpha \in [0, \alpha_\rho) \cup \{0\}$ and $u \in \mathcal{S}^{d-1}$ (union with $\{0\}$ is needed when $\alpha_\rho = 0$), the map $\mu \mapsto M_{\alpha,u}^\rho(\mu)$ is strictly convex on $\mathbb{R}^d$, so that the ρ-quantile $\mu_{\alpha,u}^\rho$ is unique.*

This covers the well-known result stating that, for $\rho(t) = t$, all ρ-quantiles are unique provided that P is not supported on a line. Remarkably, Theorem 3.3.4 shows that this structural constraint on P is not needed for the ρ-quantiles associated with other power loss functions: under the corresponding moment assumptions, all ρ-quantiles are unique for $\rho(t) = t^p$, with $p \in (1, 2]$, whereas, for $p > 2$, all ρ-quantiles with an order α that is below (3.3.5) are unique. Similarly, for exponential loss functions, ρ-quantiles are unique for any order $\alpha < .9939$.

3.4 The spherical case

In this section, we consider the special case for which $P(\in \mathcal{P}_d^\rho)$ is spherically symmetric about some location $\mu_0(\in \mathbb{R}^d)$, in the sense that, for any d-Borel set B and any $d \times d$ orthogonal matrix O, the P-probability of $\mu_0 + OB$ does not depend on O. Since ρ-quantiles are translation-equivariant, we will actually restrict, without any loss of generality, to the case $\mu_0 = 0$ (translation-equivariance here means that if μ is a ρ-quantile of P of order α in direction u, then, for any $h \in \mathbb{R}^d$, $\mu + h$ is a ρ-quantile of P_h of order α in direction u, where P_h is the distribution of $Z + h$ when Z has distribution P).

Note that it follows from Proposition 3.2.2 that, if P is spherically symmetric about the origin of $\mathbb{R}^d$ and satisfies $P[\{0\}] < 1$, with $d \geq 2$ say, then any quantile contour $\{\mu_{\alpha,u} : u \in \mathcal{S}^{d-1}\}$, with $\alpha \in [0, \alpha_\rho) \cup \{0\}$, is a hypersphere (uniqueness of these quantiles follows from Theorem 3.3.4 since P is then not supported on a line). Proposition 3.2.2 then also implies that, for an arbitrary order $\alpha \in [0, 1)$ and any direction $u \in \mathcal{S}^{d-1}$, the ρ-quantiles of P of order α in direction u form a set that is invariant under all rotations fixing u. In particular, if $\mu_{\alpha,u}^\rho$ is unique, then it belongs to the line spanned by u, which is most natural. For $\alpha \geq \alpha_\rho$, however, uniqueness is not guaranteed, so that it is unclear whether or not quantiles meet this natural property in the spherical case. This motivates the following result.

Theorem 3.4.1. *Let $\rho \in \mathcal{C}$ and $P \in \mathcal{P}_d^\rho$. Assume that P is spherically symmetric about the origin of $\mathbb{R}^d$. Then, (i) for $\alpha = 0$ and any $u \in \mathcal{S}^{d-1}$, the unique ρ-quantile $\mu_{\alpha,u}^\rho$ is the origin of $\mathbb{R}^d$; (ii) for $\alpha \in (0, 1)$ and $u \in \mathcal{S}^{d-1}$, any ρ-quantile $\mu_{\alpha,u}^\rho$ belongs to the halfline $\{\lambda u : \lambda \geq 0\}$.*

In case (ii), the origin of $\mathbb{R}^d$ may be a ρ-quantile of order $\alpha > 0$ in direction u. Actually, it can be shown that (a) for $\rho(t) = t$, the origin is a ρ-quantile of order $\alpha > 0$ in direction u if and only $\alpha \leq P[\{0\}]$. Moreover, (b) provided that $\psi_+(0)P[\{0\}] + P[\|Z\| \in (0, \infty) \setminus \mathcal{D}_\rho] = 0$ where Z has distribution P (a condition that always holds for $\rho(t) = t^p$ with $p > 1$), the origin cannot be a ρ-quantile of order $\alpha > 0$ in direction u, so that all these quantiles then belong to $\{\lambda u : \lambda > 0\}$ (for the sake of completeness, we prove (a)–(b) in the appendix; see Proposition 3.9.12).

If P is spherically symmetric about the origin and satisfies $P[\{0\}] < 1$, Theorem 3.3.4 shows that ρ-quantiles are unique for any $\alpha < \alpha_\rho$ (as mentioned above) but it remains silent on the case $\alpha \geq \alpha_\rho$. Interestingly, we will be able to say more under sphericity, thanks to the fact that Theorem 3.4.1 entails that uniqueness will hold if $t \mapsto M^\rho_{\alpha,u}(tu)$ is strictly convex over $[0, \infty)$ for all $u \in \mathcal{S}^{d-1}$, which in turn will hold if $t \mapsto H^\rho_{\alpha,u}(z - tu)$ is convex for any $z \in \mathbb{R}^d$ and any $u \in \mathcal{S}^{d-1}$. Accordingly, for any $\alpha \in [0, 1]$, let $\mathcal{C}^{\rm sph}_\alpha$ be the collection of functions $\rho \in \mathcal{C}$ such that $t \mapsto H^\rho_{\alpha,u}(z - tu)$ is convex for any $z \in \mathbb{R}^d$ and $u \in \mathcal{S}^{d-1}$. Since $\mathcal{C}^{\rm sph}_0 = \mathcal{C}$ and $\mathcal{C}^{\rm sph}_{\alpha_2} \subseteq \mathcal{C}^{\rm sph}_{\alpha_1}$ for any $\alpha_1 < \alpha_2$ (see the proof of Theorem 3.4.2 below), we let $\alpha^{\rm sph}_\rho := \max\{\alpha \in [0, 1] : \rho \in \mathcal{C}^{\rm sph}_\alpha\}$, parallel to what we did for α_ρ in Section 3.3. We have the following result.

Theorem 3.4.2. *Let $\rho \in \mathcal{C}$ and $P \in \mathcal{P}^\rho_d$. Assume that P is spherically symmetric about the origin of $\mathbb{R}^d$. Then, for any $\alpha \in [0, \alpha^{\rm sph}_\rho) \cup \{0\}$ and $u \in \mathcal{S}^{d-1}$ (again, union with $\{0\}$ is needed when $\alpha^{\rm sph}_\rho = 0$), the map $t \mapsto M^\rho_{\alpha,u}(tu)$ is strictly convex on $[0, \infty)$, and the ρ-quantile $\mu^\rho_{\alpha,u}$ is unique.*

Note that, in the present spherical setup, this uniqueness result may only strengthen the one in Theorem 3.3.4, since the fact that $\mathcal{C}_\alpha \subseteq \mathcal{C}^{\rm sph}_\alpha$ for any $\alpha \in [0, 1]$ implies that $\alpha^{\rm sph}_\rho \geq \alpha_\rho$. Of course, it is natural to wonder under which conditions on ρ all ρ-quantiles are unique ($\alpha^{\rm sph}_\rho = 1$) and, when these conditions are not met, what are the orders α for which uniqueness is guaranteed (that is, what is then the value of $\alpha^{\rm sph}_\rho < 1$). The following result provides a complete answer to these questions.

Theorem 3.4.3. *Let $\rho \in \mathcal{C}$. Then, (i) $\alpha^{\rm sph}_\rho = 1$ if and only if $4p_t + q_t - p_t q_t \leq 6$ for any $t \in \mathcal{D}_\rho$, where p_t and q_t are as in Theorem 3.3.3; (ii) if $\alpha^{\rm sph}_\rho < 1$, then, letting $\mathcal{D}^{\rm sph}_\rho := \{t \in \mathcal{D}_\rho : 4p_t + q_t - p_t q_t > 6\} (\subseteq \mathcal{D}^{\rm cv}_\rho)$,*

$$\alpha^{\rm sph}_\rho = \inf_{t \in \mathcal{D}^{\rm sph}_\rho} \sqrt{\beta_{p_t,q_t}},$$

where

$$\beta_{p,q} := \frac{2(pq - p - q)^3(\sqrt{c_{p,q}} - q(2p - 3)/3)^2}{3(p - 1)^2(3 - p)(\sqrt{c_{p,q}} - (2p - q))(\sqrt{c_{p,q}} - q(2p - 3))^2}$$

involves $c_{p,q} := \frac{q}{3}(3 - 2p)(2pq - 8p + q)$ (if q makes $\beta_{p,q}$ undefined in the expression above, then we let $\beta_{p,q} := \lim_{r \to q} \beta_{p,r}$).

Parallel to α_ρ in Theorems 3.3.2–3.3.3, $\alpha^{\rm sph}_\rho$ does not depend on $d (\geq 2)$. For the power loss functions $\rho(t) = t^p$ with $p \geq 1$, it follows from Theorem 3.4.3 that

$$\alpha^{\rm sph}_\rho = \begin{cases} \sqrt{\frac{p^2(p-2)^3(b_p - (p - \frac{3}{2})^2)}{(p-1)^2(3-p)(b_p - \frac{3}{2})(b_p - 3(p - \frac{3}{2})^2)}}\,(<1) & \text{if } p \in (2, 3) \\ 1 & \text{otherwise,} \end{cases} \tag{3.4.6}$$

with $b_p := \big(3(p - \frac{3}{2})(\frac{7}{2} - p)\big)^{1/2}$. For $p \in [1, 2]$, the result is just a corollary of Theorem 3.3.2 since we then have $\alpha^{\rm sph}_\rho \geq \alpha_\rho = 1$. For $p > 3$, (3.4.6) implies that all ρ-quantiles are uniquely defined under sphericity, while there is no guarantee that this is the case in general (since $\alpha_\rho < 1$ for such values of p). As shown in Figure 4.1, the values of $\alpha^{\rm sph}_p$ for $p \in (2, 3)$ are remarkably close to one (the minimal value, achieved at $p \approx 2.429$, is about .9987), which implies that, also for $p \in (2,3)$, essentially all ρ-quantiles are uniquely defined under sphericity. For the exponential loss functions $\rho(t) = \exp(ct) - 1$, all ρ-quantiles are also unique under sphericity ($\alpha_\rho^{\rm sph} = 1$), while "only" quantiles of order $\alpha < \alpha_\rho = .9939$ are guaranteed to be unique in general.

3.5 Differentiability of the objective function

For any $\alpha < \alpha_\rho$, the ρ-quantiles $\mu^\rho_{\alpha,u}$ are minimizers of the convex objective function $\mu \mapsto M^\rho_{\alpha,u}(\mu)$. If this objective function is smooth, then ρ-quantiles are characterized by the first-order condition $\nabla M^\rho_{\alpha,u}(\mu^\rho_{\alpha,u}) = 0$. Such a gradient condition will actually play a key role when deriving further properties of ρ-quantiles in the next sections. This provides a strong motivation to study smoothness of the map $\mu \mapsto M^\rho_{\alpha,u}(\mu)$. We start with the following result.

Proposition 3.5.1. *Let $\rho \in \mathcal{C}$ and $P \in \mathcal{P}^\rho_d$. Fix $\alpha \in [0,1]$ and $u \in \mathcal{S}^{d-1}$. Let Z be a random d-vector with distribution P and write $Z_\mu := Z - \mu$, for any $\mu \in \mathbb{R}^d$. Then, for any $\mu \in \mathbb{R}^d$ and $v \in \mathbb{R}^d \setminus \{0\}$, the directional derivative*

$$\frac{\partial M^\rho_{\alpha,u}}{\partial v}(\mu) = \lim_{t \overset{>}{\to} 0} \frac{M^\rho_{\alpha,u}(\mu + tv) - M^\rho_{\alpha,u}(\mu)}{t}$$

exists and is given by

$$\frac{\partial M^\rho_{\alpha,u}}{\partial v}(\mu) = \psi_+(0)(\|v\| - \alpha u'v)P[\{\mu\}] - \alpha v' \mathrm{E}\bigg[\frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\bigg(I_d - \frac{Z_\mu Z_\mu'}{\|Z_\mu\|^2}\bigg)\xi_{Z,\mu}\bigg]u$$
$$-v'\mathrm{E}\bigg[\{\psi_-(\|Z_\mu\|)\mathbb{I}[v'Z_\mu > 0] + \psi_+(\|Z_\mu\|)\mathbb{I}[v'Z_\mu < 0]\}\bigg(1 + \alpha\frac{u'Z_\mu}{\|Z_\mu\|}\bigg)\frac{Z_\mu}{\|Z_\mu\|}\bigg],$$

where I_d is the $d \times d$ identity matrix and ξ_{z_1,z_2} is as in Definition 3.1.1.

The objective function thus admits directional derivatives in all directions (hence, is continuous over $\mathbb{R}^d$), but it is not necessarily differentiable. For instance, the classical geometric quantiles obtained with $\rho(t) = t$ provide

$$\frac{\partial M^\rho_{\alpha,u}}{\partial v}(\mu) = (\|v\| - \alpha u'v)P[\{\mu\}] + v'\mathrm{E}\bigg[\bigg(\frac{\mu - Z}{\|\mu - Z\|} - \alpha u\bigg)\xi_{Z,\mu}\bigg],$$

so that $M^\rho_{\alpha,u}$ fails to be differentiable at atoms of P. Clearly, it follows from Theorem 3.5.1 that a necessary condition for this objective function to be differentiable at μ is $\psi_+(0)P[\{\mu\}] = 0$. The next result provides a necessary and sufficient condition and gives an expression for the corresponding gradient.

Theorem 3.5.2. *Let $\rho \in \mathcal{C}$ and $P \in \mathcal{P}^\rho_d$. Fix $\alpha \in [0,1]$ and $u \in \mathcal{S}^{d-1}$. Then, (i) $\mu \mapsto M^\rho_{\alpha,u}(\mu)$ is differentiable at $\mu_0 (\in \mathbb{R}^d)$ if and only if $\psi_+(0)P[\{\mu_0\}] + P[\|Z - \mu_0\| \in (0,\infty) \setminus \mathcal{D}_\rho] = 0$, in which case the corresponding gradient is*

$$\nabla M^\rho_{\alpha,u}(\mu_0) = v(\mu_0) - \alpha T(\mu_0)u, \quad \textit{with} \quad v(\mu) := -\mathrm{E}\bigg[\psi_-(\|Z_\mu\|)\frac{Z_\mu}{\|Z_\mu\|}\xi_{Z,\mu}\bigg]$$

and

$$T(\mu) := \mathrm{E}\Bigg[\Bigg\{\frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\bigg(I_d - \frac{Z_\mu Z_\mu'}{\|Z_\mu\|^2}\bigg) + \psi_-(\|Z_\mu\|)\frac{Z_\mu Z_\mu'}{\|Z_\mu\|^2}\Bigg\}\xi_{Z,\mu}\Bigg],$$

where $Z_\mu := Z - \mu$ *is based on a random d-vector* Z *with distribution* P. *(ii) If* $\psi_+(0)P[\{\mu\}] + P[\|Z - \mu\| \in (0,\infty) \setminus \mathcal{D}_\rho] = 0$ *for any* μ *in an open set* $\mathcal{N}$, *then* $\mu \mapsto M^\rho_{\alpha,u}(\mu)$ *is continuously differentiable on* $\mathcal{N}$.

It follows from this result that, in contrast with $\rho(t) = t$, the power loss functions $\rho(t) = t^p$ with $p > 1$ make the objective function $M^\rho_{\alpha,u}$ (continuously) differentiable even in the atomic case. The corresponding quantiles $\mu^\rho_{\alpha,u}$ are thus the solutions of the first-order equations $\nabla M^\rho_{\alpha,u}(\mu) = 0$, which rewrite

$$-p\mathrm{E}\bigg[\|Z_\mu\|^{p-1}\frac{Z_\mu}{\|Z_\mu\|}\xi_{Z,\mu}\bigg] = \alpha\mathrm{E}\bigg[\|Z_\mu\|^{p-1}\bigg(I_d + (p-1)\frac{Z_\mu Z_\mu'}{\|Z_\mu\|^2}\bigg)\xi_{Z,\mu}\bigg]u.$$

In particular, geometric expectiles ($p = 2$) of order α in direction u are the unique (Theorem 3.3.4) solutions of

$$2(\mu - \mathrm{E}[Z]) = \alpha\mathrm{E}\bigg[\|Z - \mu\|\bigg(I_d + \frac{(Z-\mu)(Z-\mu)'}{\|Z-\mu\|^2}\xi_{Z,\mu}\bigg)\bigg]u.$$

This is compatible with the fact that the corresponding "median" (that is, the quantile of order $\alpha = 0$, in an arbitrary direction u) is the mean vector $\mathrm{E}[Z]$.

We turn to second-order differentiability, which will be relevant when studying the asymptotic behavior of sample ρ-quantiles in Section 3.8.

Theorem 3.5.3. *Let* $\rho \in \mathcal{C}$ *and* $P \in \mathcal{P}^\rho_d$. *Fix* $\alpha \in [0,1]$, $u \in \mathcal{S}^{d-1}$, *and* $\mu_0 \in \mathbb{R}^d$. *Assume that* $P[\|Z - \mu\| \in [0,\infty) \setminus \mathcal{D}_\rho] = 0$ *for any* μ *in an open neighbourhood of* μ_0 *(hence, in particular, that* P *is non-atomic in this neighbourhood). Let further one of the following assumptions hold:*

(A) ψ_- *is concave on* $(0,\infty)$ *and*

$$\int_{\mathbb{R}^d\setminus\{\mu_0\}} \frac{\psi_-(\|z-\mu_0\|)}{\|z-\mu_0\|}\,dP(z) < \infty;$$

(A′) ψ_- *is convex on* $(0,\infty)$, $\psi_+(0) = 0$, *and there exists* $r > 0$ *such that*

$$\int_{\mathbb{R}^d} \psi_-'(\|z-\mu_0\| + r)\,dP(z) < \infty$$

(recall that ψ_-' *is the left-derivative of* ψ_-*).*

Then, for any $v \in \mathbb{R}^d \setminus \{0\}$,

$$\lim_{t \overset{>}{\to} 0} \frac{\nabla M^\rho_{\alpha,u}(\mu_0 + tv) - \nabla M^\rho_{\alpha,u}(\mu_0)}{t} = \nabla^2 M^\rho_{\alpha,u}(\mu_0)v,$$

where the Hessian matrix $\nabla^2 M^{\rho}_{\alpha,u}(\mu)$ *is given by*

$$\begin{aligned}\nabla^2 M^{\rho}_{\alpha,u}(\mu) &= \big(\partial_i\partial_j M^{\rho}_{\alpha,u}(\mu)\big)_{i,j=1,\ldots,d} \\ &= \mathrm{E}\bigg[\bigg(\psi'_-(\|Z_\mu\|) - \frac{2\psi_-(\|Z_\mu\|)}{\|Z_\mu\|} + \frac{2\rho(\|Z_\mu\|)}{\|Z_\mu\|^2}\bigg)\bigg(1+\alpha\frac{u'Z_\mu}{\|Z_\mu\|}\bigg)\frac{Z_\mu Z'_\mu}{\|Z_\mu\|^2}\xi_{Z,\mu} \\ &\quad +\frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|^2}\bigg(I_d - \frac{Z_\mu Z'_\mu}{\|Z_\mu\|^2}\bigg)\xi_{Z,\mu} + \frac{\|Z_\mu\|\psi_-(\|Z_\mu\|) - \rho(\|Z_\mu\|)}{\|Z_\mu\|^2} \\ &\quad \times\bigg\{\bigg(1+\alpha\frac{u'Z_\mu}{\|Z_\mu\|}\bigg)\bigg(I_d - \frac{Z_\mu Z'_\mu}{\|Z_\mu\|^2}\bigg) + 2\frac{Z_\mu Z'_\mu}{\|Z_\mu\|^2} + \alpha\frac{Z_\mu u' + Z'_\mu u}{\|Z_\mu\|}\bigg\}\xi_{Z,\mu}\bigg]\end{aligned}$$

(*as in the previous results,* $Z_\mu := Z - \mu$*, where* Z *is a random* d*-vector with distribution* P).

While they may seem complex at first, the assumptions of Theorem 3.5.3 turn out to be simple (and very weak) when considering specific loss functions ρ. For instance, for $\rho(t) = t^p$ with $p \geq 1$, they only require that $P \in \mathcal{P}^{\rho}_d$ is non-atomic in a neighborhood of μ_0 and is such that $\mathrm{E}[\|Z-\mu_0\|^{p-2}] < \infty$ when Z has distribution P. Note that this last assumption, that cannot be avoided since this expectation is involved in the Hessian matrix $\nabla^2 M^{\rho}_{\alpha,u}(\mu)$, is superfluous for $p \geq 2$. Under the assumptions of Theorems 3.3.4 and 3.5.3, this Hessian matrix is positive definite for any $\alpha \in [0,\alpha_\rho) \cup \{0\}$ and any $u \in \mathcal{S}^{d-1}$; since this will be needed in the sequel, we prove it in the appendix (see Lemma 3.9.21).

3.6 A ρ-version of Robert Serfling's DOQR paradigm

In a series of papers, Robert Serfling introduced the *DOQR paradigm*, that presents *Depth*, *Outlyingness*, *Quantile* and *Rank* functions as interrelated, yet distinct, objects of interest for multivariate nonparametric statistics; see, e.g., [89], [90], [93] and the references therein. While this paradigm in principle applies to any multivariate quantile concept, the primary focus when considering this paradigm in the aforementioned papers was on geometric quantiles. This makes it natural to study the paradigm for the generalized geometric quantiles considered in this work, which leads to introducing ρ-depth, ρ-outlyingness, ρ-quantile and ρ-rank functions. As we will see later, some of these functions play a key role to understand the nature of extreme ρ-quantiles.

We start by formally defining ρ-quantile functions. Restricting to the interesting case for which $\alpha_\rho > 0$, Theorems 3.2.1 and 3.3.4 imply that ρ-quantiles exist and are unique for any $\alpha \in [0,\alpha_\rho)$ and $u \in \mathcal{S}^{d-1}$, which allows us to adopt the following definition.

Definition 3.6.1. *Let* $\rho \in \mathcal{C}$ *and* $P \in \mathcal{P}^d_\rho$*. Assume that there is no open interval in* $(0,\infty)$ *on which* ψ_- *is constant or that* P *is not concentrated on a line. Write* $\mathcal{B}^d_r = \{z \in \mathbb{R}^d : \|z\| < r\}$*. Then, the* ρ-quantile function *of* P *is the map* $Q = Q^{\rho}_P : \mathcal{B}^d_{\alpha_\rho} \to \mathbb{R}^d$ *that is defined through* $Q(\alpha u) = \mu^{\rho}_{\alpha,u}$.

In dimension $d = 1$ and $\rho(t) = t$, this provides the (centered-outward version of the) usual quantile function. This standard quantile function, that is defined on $\mathcal{B}^1_1 =$

$(-1, 1)$, may of course fail to be continuous (it is discontinuous for empirical probability measures). The multivariate case $d \geq 2$ is different.

Proposition 3.6.2. *Let $\rho \in \mathcal{C}$ and $P \in \mathcal{P}_\rho^d$, with $d \geq 2$. Assume that there is no open interval in $(0, \infty)$ on which ψ_- is constant or that P is not concentrated on a line. Then, the quantile function $Q = Q_P^\rho : \mathcal{B}_{\alpha_\rho}^d \to \mathbb{R}^d$ is continuous.*

Following [89], we associate with the ρ-quantile function Q corresponding concepts of *rank function R*, *depth function D* and *outlyingness function O*. We start with the rank function.

Definition 3.6.3. *Let $\rho \in \mathcal{C}$ and assume that $P \in \mathcal{P}_\rho^d$ is not a Dirac probability measure. Then, the* rank function *of P is the map $R = R_P^\rho : \mathbb{R}^d \to \mathbb{R}^d$ defined through $R(\mu) = (T(\mu))^{-1} v(\mu)$, where the $d \times d$ matrix $T(\mu)$ and the d-vector $v(\mu)$ were introduced in Theorem 3.5.2.*

In the setup of this definition, $T(\mu)$ is positive definite, hence invertible, for any $\mu \in \mathbb{R}^d$ (for the sake of completeness, we prove this in the appendix; see Lemma 3.9.16). The natural assumptions under which to study the rank function are those of Theorem 3.5.2 complemented by conditions ensuring uniqueness of ρ-quantiles (which provides the assumptions in Theorem 3.6.4 below). Under these assumptions, $\mu \mapsto M_{\alpha,u}^\rho(\mu)$ is continuously differentiable on $\mathbb{R}^d$, with gradient

$$\nabla M_{\alpha,u}^\rho(\mu) = T(\mu)(R(\mu) - \alpha u),$$

so that $\mu = \mu_{\alpha,u} = Q(\alpha u)$ (for $\alpha < \alpha_\rho$) if and only if $R(\mu) = \alpha u$ (recall that, under the assumptions considered, quantiles of order $\alpha \in [0, \alpha_\rho)$ in direction $u \in \mathcal{S}^{d-1}$ are indeed uniquely determined by the gradient condition $\nabla M_{\alpha,u}^\rho(\mu) = 0$). This provides a clear interpretation of the rank function as the inverse map of the quantile function. We have the following result.

Theorem 3.6.4. *Let $\rho \in \mathcal{C}$ and $P \in \mathcal{P}_d^\rho$. Assume that there is no open interval in $(0, \infty)$ on which ψ_- is constant or that P is not concentrated on a line. Assume further that $\psi_+(0) P[\{\mu\}] + P[\|Z - \mu\| \in (0, \infty) \setminus \mathcal{D}_\rho] = 0$ for any $\mu \in \mathbb{R}^d$. Write $\mathcal{Z}_\rho = Q_P^\rho(\mathcal{B}_{\alpha_\rho}^d)$. Then, $Q = Q_P^\rho : \mathcal{B}_{\alpha_\rho}^d \to \mathcal{Z}_\rho$ is a homeomorphism, with inverse $R_{P|\mathcal{Z}_\rho}^\rho : \mathcal{Z}_\rho \to \mathcal{B}_{\alpha_\rho}^d$ (the restriction of R_P^ρ to $\mathcal{Z}_\rho$).*

If $t \mapsto t^2/\rho(t)$ is concave on $(0, \infty)$, then we are in the important particular case $\alpha_\rho = 1$ (Theorem 3.3.2), for which the quantile function Q is defined on the open unit ball $\mathcal{B}^d = \mathcal{B}_1^d$. We then have the following result.

Theorem 3.6.5. *Let $\rho \in \mathcal{C}$ be such that $t \mapsto t^2/\rho(t)$ is concave on $(0, \infty)$. Assume that P is not concentrated on a line and that $\psi_+(0) P[\{\mu\}] + P[\|Z - \mu\| \in (0, \infty) \setminus \mathcal{D}_\rho] = 0$ for any $\mu \in \mathbb{R}^d$. Then, $\mathcal{Z}_\rho = Q_P^\rho(\mathcal{B}^d) = \mathbb{R}^d$, so that $Q = Q_P^\rho : \mathcal{B}^d \to \mathbb{R}^d$ is a homeomorphism, with inverse $R = R_P^\rho : \mathbb{R}^d \to \mathcal{B}^d$.*

This result shows in particular that for any power loss function $\rho(t) = t^p$ with $p \in [1, 2]$, any non-atomic probability measure that is not concentrated on a line provides ρ-quantiles that span the whole Euclidean space $\mathbb{R}^d$ (the non-atomicity condition is actually needed for $p = 1$ only), whereas the result remains silent for the case $p > 2$. This will have important implications when studying extreme quantiles in Section 3.7.

Let us turn to depth and outlyingness functions. Clearly, central or "deep" quantiles are indexed by a small order $\alpha \in [0, 1)$, whereas exterior or "outlying" ones are rather indexed by a large order α. A natural outlyingness measure for $\mu \in \mathbb{R}^d$ is then the order α of the quantile $\mu_{\alpha,u}$ for which $\mu = \mu_{\alpha,u}$, that is, the oultlyingness of μ is $\|R(\mu)\|$. Any decreasing function of this outlyingness measure is then a natural depth measure. We adopt the following definition.

Definition 3.6.6. *Let $\rho \in \mathcal{C}$ and $P \in \mathcal{P}^d_\rho$. Then, (i) the* outlyingness function *of P is the map $O = O^\rho_P$ from $\mathbb{R}^d$ to $[0, 1]$ defined through $O(\mu) = \min(\|R(\mu)\|, 1)$, where $R = R^\rho_P$ is the rank function of P. (ii) The* depth function *of P is the map $D = D^\rho_P$ from $\mathbb{R}^d$ to $[0, 1]$ defined through $D(\mu) = 1 - O(\mu)$.*

The deepest location, the only one that receives the maximal depth value one, is the ρ-median $\mu^\rho_{0,u}$ of P (the direction u plays no role for $\alpha = 0$). For any direction $u \in \mathcal{S}^{d-1}$, depth decreases along the quantile curves $\{\mu^\rho_{\alpha,u} : \alpha \in [0, \alpha_\rho)\}$ originating from the ρ-median. For $\rho(t) = t$, this depth reduces to the celebrated *geometric depth*; see, e.g., [102]. The depths associated with our ρ-quantiles extend this classical depth; in particular, an "expectile geometric depth", whose deepest point is the mean vector of P, is obtained for $\rho(t) = t^2$. For any depth function, the depth regions collecting locations with depth exceeding a given threshold are of interest. The depth regions

$$\mathcal{R}^\rho_\alpha = \mathcal{R}^\rho_{P,\alpha} := \{\mu \in \mathbb{R}^d : D^\rho_P(\mu) \geq \alpha\}$$

are nested "centrality regions"; see, e.g., [70] and the references therein. The corresponding depth contours, i.e. the boundaries $\partial\mathcal{R}^\rho_\alpha$ of these depth regions, collect the ρ-quantiles associated with a fixed order α.

For each combination of $\alpha \in \{.25, .50, .75\}$ and $p \in \{1, 1.5, 2, 4\}$, we plot in Figure 4.2 the depth contours of order α, based on $\rho(t) = t^p$, for the empirical probability measure P_n of six random samples of size $n = 200$ (these were obtained from a uniform grid of 50 directions on the unit circle $\mathcal{S}^{d-1}$, and each quantile was evaluated through the descent method involving the backtracking line search in Section 9.2 of [7]; R code is available on request). These samples were generated from (i) the bivariate standard normal distribution, (ii)–(iii) the standard t-distributions with $\nu = 4$ and $\nu = 1$ degrees of freedom, (iv) the centered bivariate normal distribution with covariance matrix $\Sigma = \binom{2\ 1}{1\ 1}$, (v) the bivariate distribution whose marginals are independent exponential distributions with mean one, (vi) the standard skew-t distribution with 4 degrees of freedom and slant vector $\alpha = (10, 10)$; see [3]. Figure 4.2 shows that larger values of p provide contours that are more concentrated about the corresponding median; the only exception is the Cauchy distribution, for which these large-p contours are the most spread ones due to their lack of robustness with respect to extreme observations. As expected, the various medians differ when the underlying distribution is skewed, as it is the case in (v)–(vi).

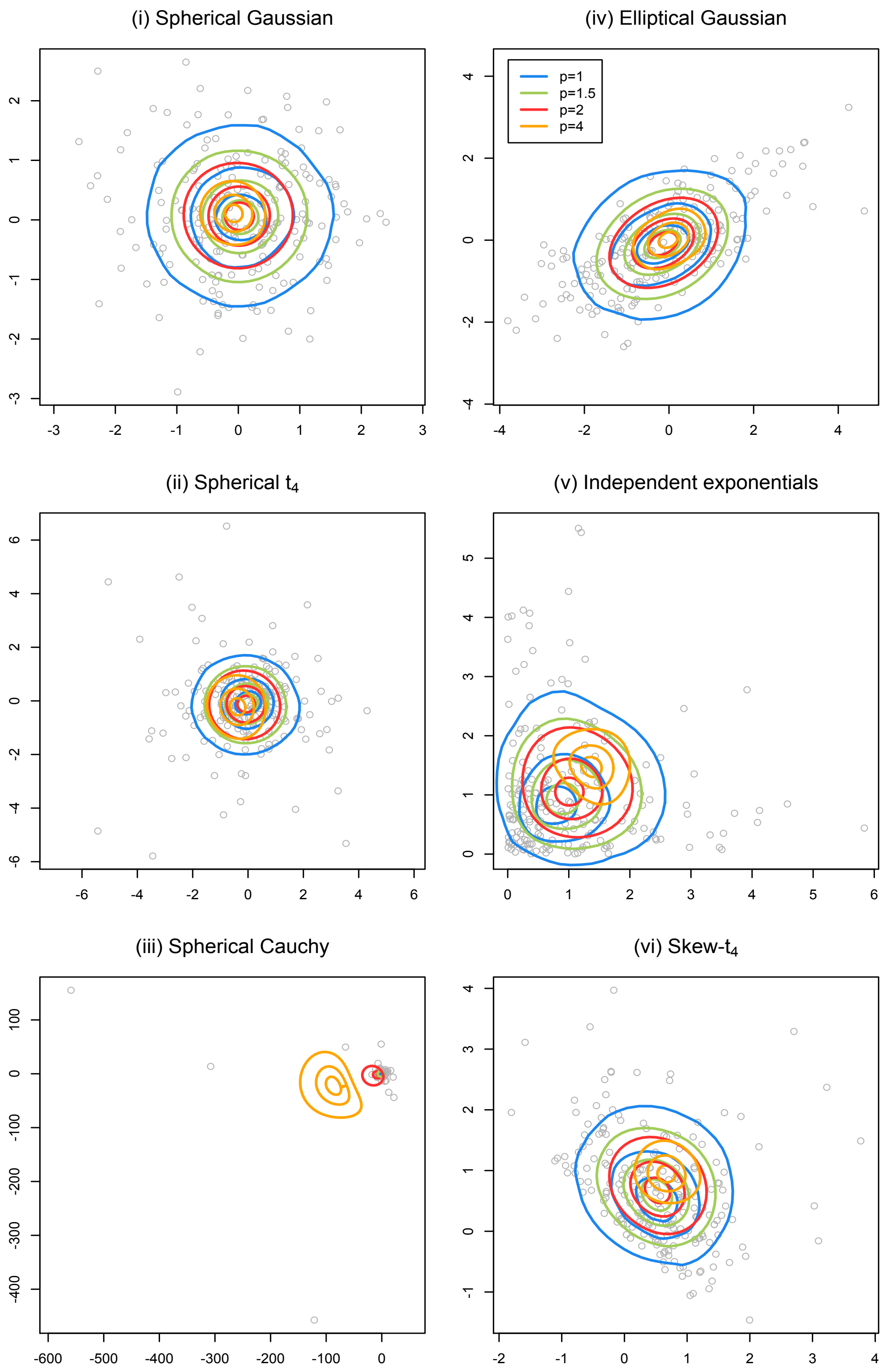

Figure 3.2: For $\rho(t) = t^p$ with $p = 1, 1.5, 2, 4$, ρ-depth contours of order $\alpha = .25, .50, .75$ for random samples of size $n = 200$ drawn from six bivariate distributions; see Section 3.6 for details.

3.7 Extreme quantiles

Recently, [41, 42] studied the geometric quantiles from [17] with a focus on extreme quantiles, that is, those associated with an order α that is close to one. In particular, [42] derived striking results on extreme quantiles showing that (i) geometric quantiles exit any compact set as $\alpha \to 1$ and that (ii) they do so in a direction that eventually coincides with the direction u in which quantiles are computed. Surprisingly, this typically also happens when the underlying distribution P is compactly supported. As shown in [83], the result even holds under atomic probability measures P, so that this unexpected behavior also shows in the sample case (provided that not all observations lie on a line of $\mathbb{R}^d$).

Of course, it is natural to ask whether or not this behavior of extreme quantiles shows for other ρ-quantiles. We tackle this question in the present section. Our first result is the following.

Theorem 3.7.1. *Let $\rho \in \mathcal{C}$ be such that $t \mapsto t^2/\rho(t)$ is concave on $(0, \infty)$. Assume that P is not concentrated on a line and that $\psi_+(0)P[\{\mu\}]+P[\|Z-\mu\| \in (0,\infty)\setminus\mathcal{D}_\rho] = 0$ for any $\mu \in \mathbb{R}^d$. Let (α_n) be a sequence in $[0, 1)$ that converges to one and (u_n) be a sequence in $\mathcal{S}^{d-1}$. Then, (i) $\|\mu^\rho_{\alpha_n,u_n}\| \to \infty$; (ii) if $u_n \to u$, then $\mu^\rho_{\alpha_n,u_n}/\|\mu^\rho_{\alpha_n,u_n}\| \to u$.*

This result shows that all ρ-quantiles for which $\alpha_\rho = 1$, hence in particular those associated with $\rho(t) = t^p$ for $p \in (1, 2]$, will show the behavior of the extreme quantiles from [17] described above. Note that for $p \in (1, 2]$, we have $\psi_+(0) = 0$ and $\mathcal{D}_\rho = (0,\infty)$, so that Theorem 3.7.1 does not require that P is non-atomic, hence also allows for empirical distributions. We illustrate this in Figure 4.3 for $P = P_n$, the empirical distribution of a random sample of size $n = 10$ drawn from the bivariate standard normal distribution. For $\rho(t) = t^p$, with $p \in \{1, 1.5, 2, 2.25, 3, 4\}$, the figure shows the ρ-quantiles $\mu^\rho_{\alpha,u}$, for $\alpha \in [0, 1)$ and $u = (\cos(\pi\ell/6), \sin(\pi\ell/6))$, with $\ell = 0, 1, 2, 3$. Clearly, for the values of p that are covered by Theorem 3.7.1, namely $p = 1, 1.5, 2$, quantiles exit any compact set and do so eventually in the corresponding direction u. In contrast, the figure suggests that, for $p > 2$, the Euclidean norm of extreme ρ-quantiles remains bounded. This is indeed the case, as the following result shows.

Theorem 3.7.2. *Let $\rho \in \mathcal{C}$ such that $\rho(t)/t^2 \to \infty$ as $t \to \infty$. Assume that $P \in \mathcal{P}^\rho_d$ (a) is not concentrated on a line of $\mathbb{R}^d$ and (b) satisfies $\int_{\mathbb{R}^d} \rho(\|z\|)\, dP(z) < \infty$ (if $\rho(t)/t^3$ is bounded away from 0 as $t \to \infty$, then Condition (b) is superfluous). Then, there exists a bounded set $S \subset \mathbb{R}^d$ such that, for any $\alpha \in [0, 1)$ and $u \in \mathcal{S}^{d-1}$, all ρ-quantiles of order α in direction u belong to S (moreover, $D(\mu) = 0$ for any $\mu \in \mathbb{R}^d \setminus S$).*

Under the condition of this result, all ρ-quantiles of order α in direction u may fail to be unique for $\alpha \in (\alpha_\rho, 1)$, which is the reason why Theorem 3.7.2 states that all ρ-quantiles of order α in direction u belong to S. The result implies that for $\rho(t) = t^p$ with $p \geq 3$, extreme ρ-quantiles are bounded as soon as $P \in \mathcal{P}^\rho_d$ is not concentrated on a line and that, for $\rho(t) = t^p$ with $p \in (2, 3)$, the same holds provided that P further has finite moments of order p rather than finite moments of order $p-1$ only (we conjecture that this stronger moment assumption for $p \in (2, 3)$ is actually superfluous, but we were not able to avoid this assumption when proving Theorem 3.7.2). Note

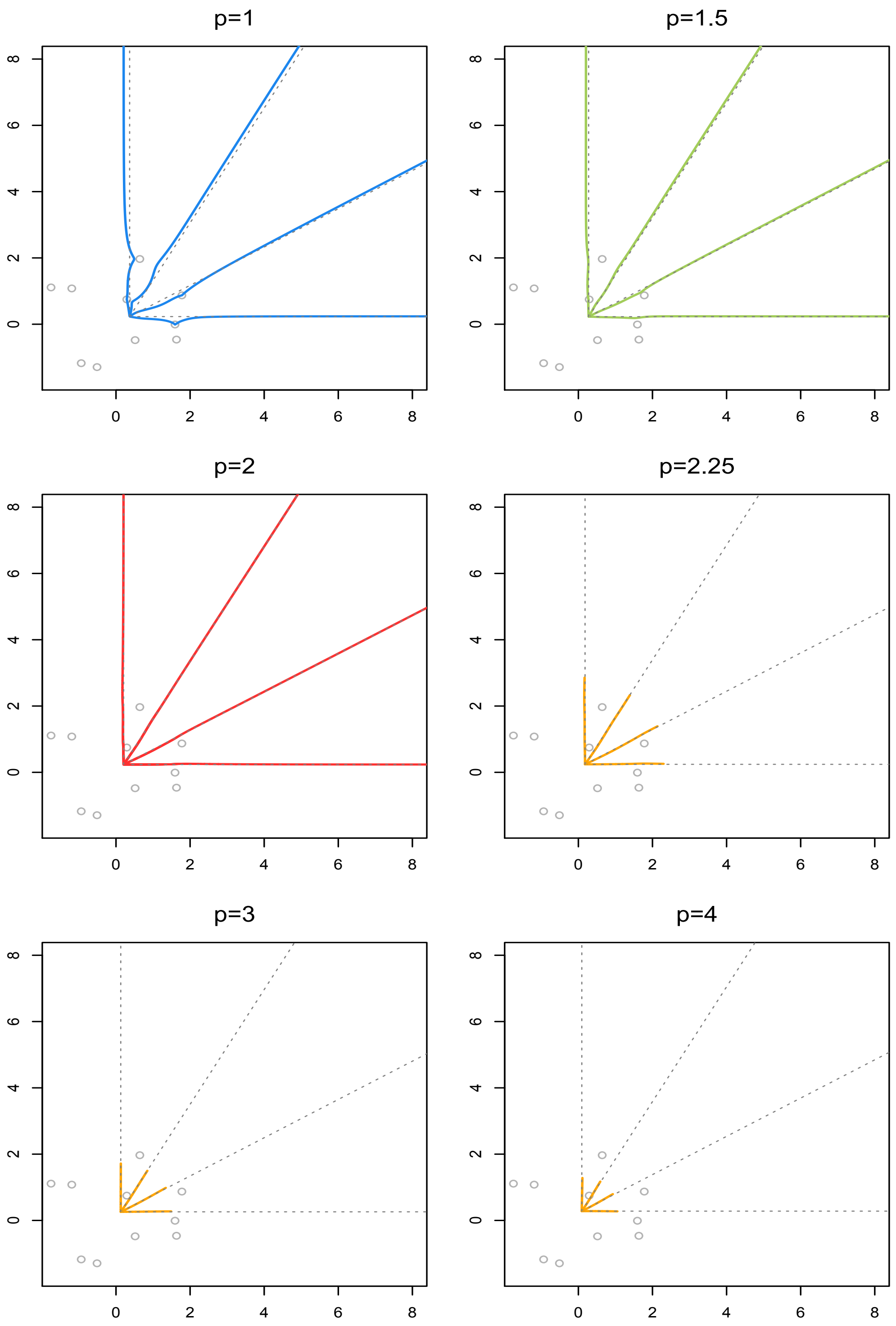

Figure 3.3: For the loss functions $\rho(t) = t^p$ with $p = 1, 1.5, 2, 2.25, 3, 4$, the plots show the ρ-quantiles $\mu^{\rho}_{\alpha,u}$ for $\alpha \in [0, 1)$ and $u = (\cos(\pi\ell/6), \sin(\pi\ell/6))$, with $\ell = 0, 1, 2, 3$; the underlying probability measure P is the empirical distribution P_n associated with a random sample of size $n = 10$ from the bivariate standard normal distribution. Dashed lines are showing the halflines with the corresponding directions u originating from the median $\mu^{\rho}_{0,u}$.

that Theorem 3.7.2 confirms in particular that, in Figure 4.3, the ρ-quantiles associated with $p > 2$ form a bounded set.

As mentioned in Section 3.2, ρ-quantiles in principle are not defined for $\alpha = 1$, but of course Definition 3.1.1 may still be adopted to define possible quantiles for order $\alpha = 1$. We have the following existence result.

Proposition 3.7.3. *Let the assumptions of Theorem 3.7.2 hold. Then, for any $u \in \mathcal{S}^{d-1}$, there exists a quantile $\mu^{\rho}_{1,u}$.*

In contrast, it directly follows from Theorem 3.6.5 that, under the assumptions of Theorem 3.7.1, there is no $u \in \mathcal{S}^{d-1}$ for which a quantile $\mu^{\rho}_{1,u}$ exist (this result is already known for $\rho(t) = t$; see Proposition 2.1 in [42]).

The following corollary of Theorem 3.7.2 extends in some sense the continuity of the quantile function (Proposition 3.6.2) to the framework where quantiles of order $\alpha = 1$ exist.

Corollary 3.7.4. *Let the assumptions of Theorem 3.7.2 hold. Let (α_n) be a sequence in $[0, 1)$ that converges to $\alpha \in [0, 1]$ and (u_n) be a sequence in $\mathcal{S}^{d-1}$ that converges to $u(\in \mathcal{S}^{d-1})$. Fix an arbitrary sequence $(\mu^{\rho}_{\alpha_n,u_n})$ of ρ-quantiles. Then, (i) any converging subsequence of $(\mu^{\rho}_{\alpha_n,u_n})$ converges to a ρ-quantile $\mu^{\rho}_{\alpha,u}$; (ii) if $\mu^{\rho}_{\alpha,u}$ is unique, then $\mu^{\rho}_{\alpha_n,u_n} \to \mu^{\rho}_{\alpha,u}$.*

This result further confirms that the quantile functions—hence also the rank, depth and outlyingness functions—associated with the loss functions ρ covered by Theorem 3.7.1 and Theorem 3.7.2 are very different in nature. In particular, in the framework of Theorem 3.7.2, the depth of μ will be exactly zero if $\|\mu\|$ is large enough. Some recent research efforts in the statistical depth literature aimed at defining depth functions—or at modifying existing depth functions—that do not show this "vanishing property"; see, e.g., [34] and the many references therein. This vanishing property is indeed undesirable in some inferential applications, such as, e.g., supervised classification based on the max-depth approach; see [34], [39], and [58]. Quite nicely, the ρ-depths associated with loss functions ρ compatible with Theorem 3.7.1 will not exhibit this vanishing property. Yet, as in [42], some might find it shocking that the corresponding ρ-quantiles span the whole Euclidean space even when P is compactly supported. This can be avoided by adopting a loss function ρ meeting the conditions of Theorem 3.7.2. As a conclusion, while Theorems 3.7.1–3.7.2 discriminate between two fundamentally different classes of DOQR functions, none of these two worlds is "the good one" and the choice of ρ, hence the choice among both worlds, should be performed based on the inferential problem at hand.

3.8 Asymptotics for point estimation

We now consider estimation of the ρ-quantiles $\mu^{\rho}_{\alpha,u} = \mu^{\rho}_{\alpha,u}(P)$ based on a random sample $Z_1, \ldots, Z_n$ from P. As usual, the natural estimator is obtained by replacing P with the corresponding empirical probability measure. In this section, we study the asymptotic properties of the resulting sample ρ-quantiles. We start with the following consistency result.

Theorem 3.8.1. *Fix $\rho \in \mathcal{C}$ and assume that there is no open interval in $(0, \infty)$ on which ψ_- is constant or that $P \in \mathcal{P}_\rho^d$ is not concentrated on a line. Denote as P_n the empirical probability measure associated with a random sample of size n from P. Fix $\alpha \in [0, \alpha_\rho) \cup \{0\}$, $u \in \mathcal{S}^{d-1}$, and write $\hat{\mu}^\rho_{\alpha,u} = \mu^\rho_{\alpha,u}(P_n)$. Then,*

$$\hat{\mu}^\rho_{\alpha,u} \to \mu^\rho_{\alpha,u}$$

almost surely as $n \to \infty$.

The sample geometric median, that is, the median obtained with the loss function $\rho(t) = t$, satisfies a classical asymptotic normality result (see, e.g., [71]), which, as usual, allows one to perform hypothesis testing or to build confidence zones for the population geometric median. This is an important advantage over competing multivariate medians, whose asymptotic distributions are too complicated to base inference on (this is in particular the case for the celebrated Tukey median; see [68]). Quite nicely, all sample ρ-quantiles enjoy a standard asymptotic normality result, relying on a neat Bahadur representation result (that typically may itself have further applications, such as the derivation of LIL results). We have the following result.

Theorem 3.8.2. *Let $\rho \in \mathcal{C}$ and $P \in \mathcal{P}^\rho_d$. Assume that there is no open interval in $(0, \infty)$ on which ψ_- is constant or that P is not concentrated on a line. Fix $\alpha \in [0, \alpha_\rho) \cup \{0\}$ and $u \in \mathcal{S}^{d-1}$. Assume that*

$$\int_{\mathbb{R}^d} \psi_-^2(\|z - \mu^\rho_{\alpha,u}\|)\, dP(z) < \infty$$

and that $P[\|Z - \mu\| \in [0, \infty) \setminus \mathcal{D}_\rho] = 0$ for any μ in an open neighborhood of $\mu^\rho_{\alpha,u}$ (hence, in particular, that P is non-atomic in this neighborhood). Let further one of the following assumptions hold:

(A) ψ_- is concave on $(0, \infty)$ and

$$\int_{\mathbb{R}^d \setminus \{\mu^\rho_{\alpha,u}\}} \frac{\psi_-(\|z - \mu^\rho_{\alpha,u}\|)}{\|z - \mu^\rho_{\alpha,u}\|}\, dP(z) < \infty;$$

(A′) ψ_- is convex on $(0, \infty)$, $\psi_+(0) = 0$, and there exists $r > 0$ such that

$$\int_{\mathbb{R}^d} \psi'_-(\|z - \mu^\rho_{\alpha,u}\| + r)\, dP(z) < \infty$$

(recall that ψ'_- is the left-derivative of ψ_-).

Let $\hat{\mu}^\rho_{\alpha,u} = \mu^\rho_{\alpha,u}(P_n)$, where P_n is the empirical probability measure associated with a random sample $Z_1, \ldots, Z_n$ of size n from P. Then,

$$\begin{aligned}
&\sqrt{n}(\hat{\mu}^\rho_{\alpha,u} - \mu^\rho_{\alpha,u}) \\
&\qquad = \frac{1}{\sqrt{n}} A^{-1} \sum_{i=1}^n \nabla H^\rho_{\alpha,u}(Z_i - \mu^\rho_{\alpha,u}) \mathbb{I}[\|Z_i - \mu^\rho_{\alpha,u}\| \in \mathcal{D}_\rho] + o_{\mathrm{P}}(1) \\
&\qquad \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, V)
\end{aligned}$$

as $n \to \infty$, where $V := A^{-1} B A^{-1}$ involves $A := \nabla^2 M^\rho_{\alpha,u}(\mu^\rho_{\alpha,u})$ and $B := \mathrm{E}[(\nabla H^\rho_{\alpha,u}(Z_1 - \mu^\rho_{\alpha,u}))(\nabla H^\rho_{\alpha,u}(Z_1 - \mu^\rho_{\alpha,u}))' \mathbb{I}[\|Z_1 - \mu^\rho_{\alpha,u}\| \in \mathcal{D}_\rho]]$.

We stress that this result requires very mild assumptions only. In particular, for the power loss functions $\rho(t) = t^p$ with $p \geq 2$, it only requires that P is non-atomic in a neighborhood of $\mu^\rho_{\alpha,u}$ and admits finite moments of order $2(p-1)$ (for the median obtained with $p = 2$, namely the mean, this is the usual finite second-order moment assumption, and the result only restate the usual multivariate central limit theorem, but for the mild local non-atomicity assumption). For $p \in [1, 2)$, Theorem 3.8.2 further requires that $\mathrm{E}[\|Z - \mu^\rho_{0,u}\|^{p-2}]$ exists and is finite (note that, for the geometric median ($p = 1$), [71] derives the result under assumptions that are more stringent, since it is imposed there that $\mathrm{E}[\|Z - \mu^\rho_{0,u}\|^{-r}]$ exists and is finite for any $r \in [0, 2)$. Invertibility of A is always guaranteed; see Lemma 3.9.21 in the appendix.

To illustrate the result, we focus on ρ-medians ($\alpha = 0$) under sphericity. If P is spherically symmetric about the origin of $\mathbb{R}^d$, then all ρ-medians $\mu^\rho_{0,u}$ are equal to each other (they coincide with the origin of $\mathbb{R}^d$; see Theorem 3.4.1), which makes it valid to compare the asymptotic variances of sample ρ-medians. We consider the power loss functions $\rho(t) = t^p$ with $p \geq 1$, for which

$$\nabla H^\rho_{\alpha,u}(x)(\nabla H^\rho_{\alpha,u}(x))'\mathbb{I}[x \in \mathcal{D}_\rho] = p^2\|x\|^{2(p-1)}\frac{xx'}{\|x\|^2}\xi_{x,0}$$

and

$$\nabla^2 H^\rho_{\alpha,u}(x)\mathbb{I}[x \in \mathcal{D}_\rho] = \|x\|^{p-2}\left\{pI_d + p(p-2)\frac{xx'}{\|x\|^2}\right\}\xi_{x,0};$$

see Lemma 3.9.13 in the appendix. If P is spherically symmetric about the origin of $\mathbb{R}^d$, then $\|Z\|$ and $Z/\|Z\|$ are mutually independent, with $Z/\|Z\|$ uniformly distributed over $\mathcal{S}^{d-1}$, which yields

$$B = \frac{p^2}{d}\mathrm{E}[\|Z\|^{2(p-1)}]I_d \quad \text{and} \quad A = \frac{p(d+p-2)}{d}\mathrm{E}[\|Z\|^{p-2}\xi_{Z,0}]I_d.$$

Thus, the asymptotic covariance matrix V is given by

$$V = A^{-1}BA^{-1} = \frac{d\mathrm{E}[\|Z\|^{2(p-1)}]}{(d+p-2)^2(\mathrm{E}[\|Z\|^{p-2}])^2}I_d =: v_p(P)I_d. \tag{3.8.7}$$

For $p = 1$, this reduces to the asymptotic covariance matrix of the geometric median (see [71]), whereas, for $p = 2$, this provides the asymptotic covariance matrix $V = \mathrm{E}[ZZ']$ of the sample mean. Let us consider various spherical distributions. If $P = P^t_\nu$ is the d-variate t-distribution with ν degrees of freedom, then $\|Z\|^2/d$ is Fisher–Snedecor with d and ν degrees of freedom, which yields

$$v_p(P^t_\nu) = \frac{\Gamma(\frac{d+2}{2})\Gamma(\frac{d+2p-2}{2})\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu-2p+2}{2})}{\Gamma^2(\frac{d+p}{2})\Gamma^2(\frac{\nu-p+2}{2})} \tag{3.8.8}$$

for $\nu > 2(p-1)$, whereas if $P = P^e_\eta$ is the d-variate power-exponential distribution with tail parameter $\eta(> 0)$, then

$$v_p(P^e_\eta) = \frac{2^{(1-\eta)/\eta}\Gamma(\frac{d+2\eta}{2\eta})\Gamma(\frac{d+2p-2}{2\eta})}{\eta\Gamma^2(\frac{d+p+2\eta-2}{2\eta})}; \tag{3.8.9}$$

the power-exponential distribution with tail parameter η refers to the distribution admitting the density $z \mapsto f^e_\eta(z) := c_{d,\eta} \exp(-\|z\|^{2\eta}/2)$ with respect to the Lebesgue measure over $\mathbb{R}^d$ ($c_{d,\eta}$ is a normalizing constant). The asymptotic variance at the standard d-variate normal distribution is obtained by taking $\nu \to \infty$ in (3.8.8) or, alternatively, by taking $\eta = 1$ in (3.8.9).

The factors $v_p(P^t_\nu)$ and $v_p(P^e_\eta)$, that completely characterize the asymptotic covariance matrix of the sample ρ-median associated with $\rho(t) = t^p$ under the corresponding distributions, are plotted in Figure 4.4. For heavy tails, the medians associated with a small value of p dominate their competitors, whereas the opposite happens for light tails (lighter-than-normal tails are obtained for $\eta > 1$ in the power-exponential case). Note that the sample ρ-median associated with $\rho(t) = t^p$ is the maximum likelihood estimator of the symmetry center in the location family generated by power-exponential distributions with parameter $\eta = p/2$, which explains that large values of p ($p > 2$) will behave well under lighter-than-normal tails. All in all, the median associated with $p = 1.5$ seems to provide a nice balance between the geometric median and sample mean associated with $p = 1$ and $p = 2$, respectively. While these considerations are specific to the spherical case, the efficiency of ρ-medians in the elliptical case could be studied following the analysis in [65], where the focus was exclusively on the geometric median ($p = 1$).

To check correctness of Theorem 3.8.2, we performed a Monte-Carlo study involving the bivariate ($d = 2$) t-distributions with ν degrees of freedom with $\nu \in \{3, 5, 7, \ldots, 21\}$, and the bivariate power-exponential distributions with parameter $\eta \in \{.8, 1.2, 1.6, \ldots, 4\}$. For each of these distributions, we generated $M = 10{,}000$ random samples of size $n = 200$ and evaluated the ρ-medians $\hat{\mu}^\rho_{0,u} = \hat{\mu}^\rho_{0,u}(m)$ associated with $\rho(t) = t^p$ for $p \in \{1, 1.5, 2, 4\}$ in each sample $m = 1, \ldots, M$. In Figure 4.4, we report the quantities

$$\hat{v}_p := n\bigg(\frac{1}{M}\sum_{m=1}^{M} \big(\hat{\mu}^\rho_{0,u}(m) - \bar{\mu}^\rho_{0,u}\big)\big(\hat{\mu}^\rho_{0,u}(m) - \bar{\mu}^\rho_{0,u}\big)'\bigg)_{11},$$

with

$$\bar{\mu}^\rho_{0,u} := \frac{1}{M}\sum_{m=1}^{M} \hat{\mu}^\rho_{0,u}(m).$$

These quantities estimate the upper-left entry in the corresponding asymptotic covariance matrix V, namely the corresponding factor $v_p(P)$ in (3.8.7). Clearly, the results are in perfect agreement with Theorem 3.8.2. It is only for $p = 4$ and t-distributions that some deviation from the asymptotic theory is seen, but this deviation vanishes for larger sample sizes (for $p = 4$ and t-distributions, Figure 4.4 also provides the results for sample size $n = 1{,}000$).

Obviously, using Theorem 3.8.2 to conduct inference based on ρ-quantiles (i.e., performing hypothesis testing or building confidence zones) requires estimating consistently the corresponding asymptotic covariance matrix V. A natural estimator is of course $\hat{V}_n = \hat{A}_n^{-1}\hat{B}_n\hat{A}_n^{-1}$, with

$$\hat{A}_n := \frac{1}{n}\sum_{i=1}^{n} \nabla^2 H^\rho_{\alpha,u}(Z_i - \hat{\mu}^\rho_{\alpha,u})\mathbb{I}[\|Z_i - \hat{\mu}^\rho_{\alpha,u}\| \in \mathcal{D}_\rho]$$

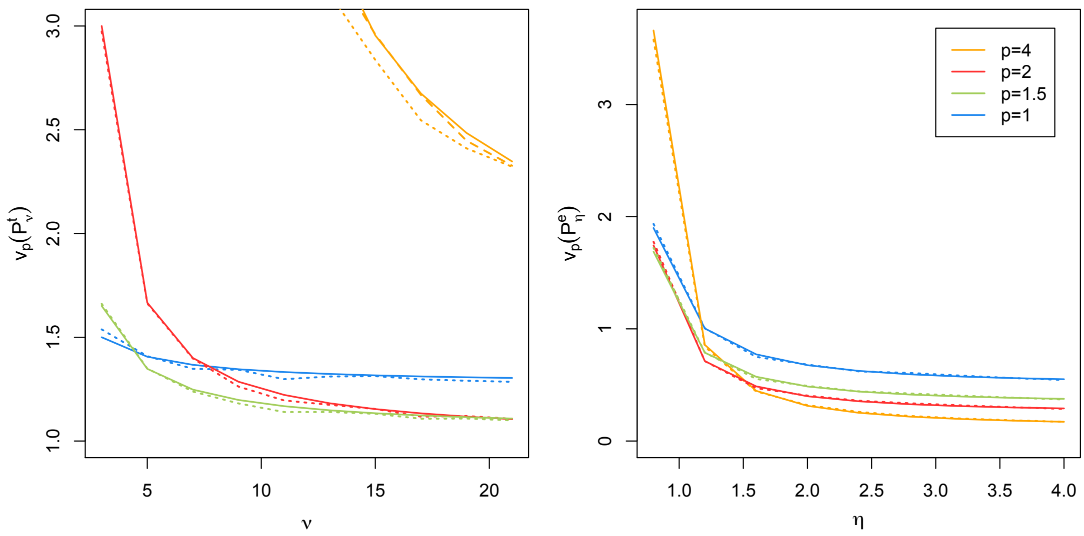

Figure 3.4: (Left:) For $p \in \{1, 1.5, 2, 4\}$, plots of $\nu \mapsto v_p(P^t_\nu)$ for $\nu = 3, 5, 7, \ldots, 21$, where $v_p(P^t_\nu)$ (see (3.8.8)) is the factor characterizing the asymptotic covariance matrix, at the bivariate t-distribution with ν degrees of freedom, of the sample ρ-median based on $\rho(t) = t^p$. Dotted lines are estimates of $v_p(P^t_\nu)$ computed from $M = 10{,}000$ random samples of size $n = 200$. For $p = 4$, the dashed line provides the same result for random samples of size $n = 1{,}000$. (Right:) Still for $p \in \{1, 1.5, 2, 4\}$, plots of $\eta \mapsto v_p(P^e_\eta)$ for $\eta = .8, 1.2, 1.6, \ldots, 4$, where $v_p(P^e_\eta)$ (see (3.8.9)) is the factor characterizing the asymptotic covariance matrix, at the bivariate power-exponential distribution with tail parameter η, of the sample ρ-median based on $\rho(t) = t^p$. Dotted lines are estimates of $v_p(P^e_\eta)$ computed from $M = 10{,}000$ random samples of size $n = 200$; see Section 3.8 for details.

and

$$\hat{B}_n := \frac{1}{n}\sum_{i=1}^{n}(\nabla H^{\rho}_{\alpha,u}(Z_i - \hat{\mu}^{\rho}_{\alpha,u}))(\nabla H^{\rho}_{\alpha,u}(Z_i - \hat{\mu}^{\rho}_{\alpha,u}))'\mathbb{I}[\|Z_i - \hat{\mu}^{\rho}_{\alpha,u}\| \in \mathcal{D}_\rho].$$

One may proceed as in [43] to establish that $\hat{V}_n$ converges in probability to V as the sample size n diverges to infinity.

3.9 Appendix: proofs

Some preliminary results

In this section, we provide results that will be repeatedly used in the proofs of the next sections.

Lemma 3.9.1. *Let ν be a finite measure on $\mathbb{R}^d$. Let (μ_n) be a sequence in $\mathbb{R}^d$ that either (i) is such that $\|\mu_n\|$ diverges to infinity or (ii) converges to $\mu \in \mathbb{R}^d$ but satisfies $\mu_n \neq \mu$ for any n. Then $\nu(\{\mu_n\}) \to 0$ as $n \to \infty$.*

Proof of Lemma 3.9.1. Assume, ad absurdum, that $\nu(\{\mu_n\})$ does not converge to 0 as $n \to \infty$. Then there exist $\varepsilon > 0$ and a subsequence (n_ℓ) such that $\nu(\{\mu_{n_\ell}\}) \geq \varepsilon$ for any ℓ. By using Assumptions (i)–(ii), we may assume, up to extraction of a further subsequence, that (μ_{n_ℓ}) has pairwise different terms. We then have

$$\nu(\mathbb{R}^d) \geq \sum_{\ell=0}^{\infty} \nu(\{\mu_{n_\ell}\}) \geq \sum_{\ell=0}^{\infty} \varepsilon = \infty,$$

which is a contradiction. ■

Lemma 3.9.2. *Let $v, w \in \mathbb{R}^d \setminus \{0\}$. Then*

$$\left\| \frac{v}{\|v\|} - \frac{w}{\|w\|} \right\| \leq 2 \min\left(\frac{\|v-w\|}{\|v\|}, \frac{\|v-w\|}{\|w\|} \right).$$

Proof of Lemma 4.8.5. A direct computation provides

$$\left\| \frac{v}{\|v\|} - \frac{w}{\|w\|} \right\| \leq \left\| \frac{\|w\|v - \|v\|v - \|v\|(w-v)}{\|v\|\|w\|} \right\| \leq \frac{\big|\|w\| - \|v\|\big|}{\|w\|} + \frac{\|w-v\|}{\|w\|} \leq 2\frac{\|v-w\|}{\|w\|}.$$

Since one may interchange the roles of v and w in these inequalities, the result follows. ■

The following result is a version of the mean-value theorem for one-sided derivatives; see, e.g., [54].

Lemma 3.9.3. *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. (i) Assume that f is left-differentiable on (a, b), with left-derivative f'_-. Then,*

$$f'_-(c_1) \leq \frac{f(b) - f(a)}{b - a} \leq f'_-(c_2)$$

for some $c_1, c_2 \in (a,b)$. *(ii) Assume that* f *is right-differentiable on* (a,b)*, with right-derivative* f'_+*. Then,*

$$f'_+(c_1) \leq \frac{f(b)-f(a)}{b-a} \leq f'_+(c_2)$$

for some $c_1, c_2 \in (a,b)$*.*

We end this section with a result that states structural properties of any loss function $\rho \in \mathcal{C}$.

Lemma 3.9.4. *Let* $\rho \in \mathcal{C}$*. Then, (i)* $t\psi_-(t) \geq \rho(t)$ *for any* $t > 0$*. (ii)* $t \mapsto \psi_-(t)$*,* $t \mapsto \psi_+(t)$*, and* $t \mapsto \rho(t)/t$ *are monotone non-decreasing on* $(0,\infty)$*. (iii)* $\psi_-(t) > 0$ *for any* $t > 0$*. (iv)* ρ *is monotone strictly increasing on* $[0,\infty)$*.*

Proof of Lemma 3.9.4. (i) Convexity of ρ implies that, for any $t > 0$,

$$\frac{\rho(t)}{t} = \frac{\rho(t)-\rho(0)}{t-0} \leq \psi_-(t),$$

which establishes the result. (ii) This trivially follows from the convexity of ρ. (iii) Assume ad absurdum that $\psi_-(t_0) = 0$ for some $t_0 > 0$. Then, Parts (i)–(ii) of the result imply that $\psi_-(t) = 0$ for any $t \in (0, t_0)$. Lemma 3.9.3(i) then entails that $\rho(t_0) = 0$, which contradicts the fact that $\rho(t) = 0$ only for $t = 0$. (iv) The result follows from Part (iii) and Lemma 3.9.3(i). ■

Proofs for Section 3.2

We first prove Part (i) of Theorem 3.2.1.

Proof of Theorem 3.2.1(i). We need to show that

$$\mathcal{I} := \int_{\mathbb{R}^d} \left| H^\rho_{\alpha,u}(z-\mu) - H^\rho_{\alpha,u}(z) \right| dP(z) < \infty \tag{3.9.10}$$

for any $\mu \in \mathbb{R}^d$. Since the result trivially holds for $\mu = 0$, we may assume that $\mu \neq 0$. Recalling that $\rho(0) = 0$, note then that

$$\begin{aligned}
\mathcal{I} \;&\leq\; \int_{\mathbb{R}^d} \left|\rho(\|z-\mu\|) - \rho(\|z\|)\right| \left(1 + \alpha \frac{u'(z-\mu)}{\|z-\mu\|}\right) \xi_{z,\mu} \, dP(z) \\
&\qquad + \int_{\mathbb{R}^d} \rho(\|z\|) \left| \left(1 + \alpha \frac{u'(z-\mu)}{\|z-\mu\|}\right)\xi_{z,\mu} - \left(1 + \alpha \frac{u'z}{\|z\|}\right)\xi_{z,0} \right| dP(z) \\
&\leq\; (1+\alpha) \int_{\mathbb{R}^d} \left|\rho(\|z-\mu\|) - \rho(\|z\|)\right| dP(z) \\
&\qquad + (1+\alpha)\rho(\|\mu\|)P[\{\mu\}] + \alpha \int_{\mathbb{R}^d \setminus \{0,\mu\}} \rho(\|z\|) \left| \frac{u'(z-\mu)}{\|z-\mu\|} - \frac{u'z}{\|z\|} \right| dP(z) \\
&=:\; (1+\alpha)\mathcal{I}_1 + (1+\alpha)\rho(\|\mu\|)P[\{\mu\}] + \alpha \mathcal{I}_2,
\end{aligned}$$

say. Lemma 3.9.3 implies that there exists c between $\|z-\mu\|$ and $\|z\|$ such that $|\rho(\|z-\mu\|)-\rho(\|z\|)| \leq \psi_-(c)|\|z-\mu\|-\|z\||$. Since ψ_- is monotone non-decreasing and nonnegative over $(0,\infty)$ (Lemma 3.9.4), we obtain

$$\begin{aligned}\big|\rho(\|z-\mu\|)-\rho(\|z\|)\big| &\leq \|\mu\|\{\psi_-(\|z-\mu\|)+\psi_-(\|z\|)\}\\ &\leq \|\mu\|\{\psi_-(\|z-\mu\|+\delta_\mu)+\psi_-(\|z\|+\delta_0)\},\end{aligned}$$

where δ_μ and δ_0 are as in is as in (3.2.4). The finiteness of $\mathcal{I}_1$ therefore follows from the assumption that ρ belongs to $\mathcal{C}$. Now, Lemma 4.8.5 shows that any $z \in \mathbb{R}^d \setminus \{0,\mu\}$ satisfies

$$\left|\frac{u'(z-\mu)}{\|z-\mu\|}-\frac{u'z}{\|z\|}\right| \leq \left\|\frac{z-\mu}{\|z-\mu\|}-\frac{z}{\|z\|}\right\| \leq 2\frac{\|\mu\|}{\|z\|}.$$

Since convexity of ρ entails that $\rho(\|z\|)/\|z\| \leq \psi_-(\|z\|) \leq \psi_-(\|z\|+\delta_0)$ for any $z \in \mathbb{R}^d \setminus \{0\}$, we then have

$$\mathcal{I}_2 \leq 2\|\mu\| \int_{\mathbb{R}^d\setminus\{0,\mu\}} \frac{\rho(\|z\|)}{\|z\|}\, dP(z) \leq 2\|\mu\| \int_{\mathbb{R}^d} \psi_-(\|z\|+\delta_0)\, dP(z) < \infty.$$

This proves (3.9.10), hence establishes the result. ■

The proof of Theorem 3.2.1(ii) requires Lemmas 3.9.5–3.9.6 below.

Lemma 3.9.5. *Let $\rho \in \mathcal{C}$ and $P \in \mathcal{P}_d^\rho$. Then, for any $(\mu_0,\alpha_0,u_0) \in \mathbb{R}^d \times [0,1] \times \mathcal{S}^{d-1}$, the map $(\mu,\alpha,u) \mapsto M^\rho_{\alpha,u}(\mu)$ is Lipschitz at (μ_0,α_0,u_0), in the sense that there exist a positive constant C and a neighbourhood $\mathcal{N}$ of (μ_0,α_0,u_0) such that, for any $(\mu,\alpha,u) \in \mathcal{N}$,*

$$\big|M^\rho_{\alpha,u}(\mu)-M^\rho_{\alpha_0,u_0}(\mu_0)\big| \leq C\big\{\|\mu-\mu_0\|+|\alpha-\alpha_0|+\|u-u_0\|\big\}.$$

In particular, the map $(\mu,\alpha,u) \mapsto M^\rho_{\alpha,u}(\mu)$ is continuous over $\mathbb{R}^d \times [0,1] \times \mathcal{S}^{d-1}$.

Proof of Lemma 3.9.5. Fix $(\mu_0,\alpha_0,u_0) \in \mathbb{R}^d \times [0,1] \times \mathcal{S}^{d-1}$. We need to prove that there exist a neighborhood $\mathcal{N}$ of (μ_0,α_0,u_0) and a positive constant C such that

$$\big|M^\rho_{\alpha,u}(\mu)-M^\rho_{\alpha_0,u_0}(\mu_0)\big| \leq C\big\{\|\mu-\mu_0\|+|\alpha-\alpha_0|+\|u-u_0\|\big\}$$

for any $(\mu,\alpha,u) \in \mathcal{N} \cap (\mathbb{R}^d \times [0,1] \times \mathcal{S}^{d-1})$. To that end, write, with $v=\alpha u$ and $v_0 = \alpha_0 u_0$,

$$\begin{aligned}&M^\rho_{\alpha,u}(\mu)-M^\rho_{\alpha_0,u_0}(\mu_0)\\ &\quad= \int_{\mathbb{R}^d} \big\{[H^\rho_{\alpha,u}(z-\mu)-H^\rho_{\alpha_0,u_0}(z-\mu_0)]-[H^\rho_{\alpha,u}(z)-H^\rho_{\alpha_0,u_0}(z)]\big\}\, dP(z)\\ &\quad= \int_{\mathbb{R}^d} \{T_1(z)+T_2(z)+T_3(z)\}\, dP(z),\end{aligned}$$

where

$$T_1(z) := \{\rho(\|z-\mu\|)-\rho(\|z-\mu_0\|)\}\bigg(1+\frac{v'(z-\mu)}{\|z-\mu\|}\bigg)\xi_{z,\mu},$$

$$T_2(z) := \{\rho(\|z-\mu_0\|) - \rho(\|z\|)\}\bigg\{\Big(1+\frac{v'z}{\|z\|}\Big)\xi_{z,0} - \Big(1+\frac{v_0'z}{\|z\|}\Big)\xi_{z,0}\bigg\},$$

and

$$T_3(z) := \rho(\|z-\mu_0\|)\bigg\{\Big(1+\frac{v'(z-\mu)}{\|z-\mu\|}\Big)\xi_{z,\mu} - \Big(1+\frac{v_0'(z-\mu_0)}{\|z-\mu_0\|}\Big)\xi_{z,\mu_0}$$
$$-\Big(1+\frac{v'z}{\|z\|}\Big)\xi_{z,0} + \Big(1+\frac{v_0'z}{\|z\|}\Big)\xi_{z,0}\bigg\}.$$

Assume that there exists a positive constant C such that

$$|T_\ell(z)| \le C\{1+\psi_-(\|z\|)+\psi_-(\|z-\mu\|)+\psi_-(\|z-\mu_0\|)\}\{\|\mu-\mu_0\|+\|v-v_0\|\} \quad (3.9.11)$$

for any $\ell = 1,2,3$ and any $z \in \mathbb{R}^d$ (in the rest of this proof, C may change from line to line). Monotonicity of ψ_- then ensures that, for μ close enough to $\mu_0 \in \mathbb{R}^d$, we have (with δ_μ as in (3.2.4))

$$|T_\ell(z)| \le C\{1+\psi_-(\|z\|+\delta_0)+2\psi_-(\|z-\mu_0\|+\delta_{\mu_0})\}\{\|\mu-\mu_0\|+\|v-v_0\|\}$$

for any $\ell = 1,2,3$ and any $z \in \mathbb{R}^d$. Since $\rho \in \mathcal{C}$ and $\|v-v_0\| \le |\alpha-\alpha_0| + \|u-u_0\|$, this provides

$$\begin{aligned}|M^\rho_{\alpha,u}(\mu) - M^\rho_{\alpha_0,u_0}(\mu_0)| &\le \sum_{\ell=1}^3 \int_{\mathbb{R}^d} |T_\ell(z)|\}\, dP(z)\\ &\le C(\|\mu-\mu_0\|+|\alpha-\alpha_0|+\|u-u_0\|),\end{aligned}$$

as was to be shown. It is therefore sufficient to prove that there indeed exists a positive constant C such that (3.9.11) holds for any $\ell = 1,2,3$ and any $z \in \mathbb{R}^d$.

Using Lemma 3.9.3 and the fact that ψ_- is non-decreasing, we obtain that, for some c between $\|z-\mu\|$ and $\|z-\mu_0\|$,

$$\begin{aligned}|T_1(z)| &\le 2\psi_-(c)\big|\|z-\mu\|-\|z-\mu_0\|\big|\\ &\le 2\{\psi_-(\|z-\mu\|)+\psi_-(\|z-\mu_0\|)\}\|\mu-\mu_0\|,\end{aligned}$$

which shows that (3.9.11) holds for $T_1(z)$. Noting that $T_2(z)$ rewrites

$$T_2(z) = \{\rho(\|z-\mu_0\|) - \rho(\|z\|)\}\frac{(v-v_0)'z}{\|z\|}\xi_{z,0},$$

we obtain in the same way (here, c is between $\|z-\mu_0\|$ and $\|z\|$)

$$\begin{aligned}|T_2(z)| &\le \psi_-(c)\big|\|z-\mu_0\|-\|z\|\big|\|v-v_0\|\\ &\le \{\psi_-(\|z-\mu_0\|+\psi_-(\|z\|)\}\|\mu_0\|\|v-v_0\|,\end{aligned}$$

so that (3.9.11) also holds for $T_2(z)$. We may thus focus on $T_3(z)$. Note that, if $z \notin$

$\{0, \mu, \mu_0\}$, then Lemma 4.8.5 yields

$$\begin{aligned}
|T_3(z)| &\leq \rho(\|z-\mu_0\|)\left|\frac{v'(z-\mu)}{\|z-\mu\|} - \frac{v_0'(z-\mu_0)}{\|z-\mu_0\|} - \frac{v'z}{\|z\|} + \frac{v_0'z}{\|z\|}\right| \\
&\leq \rho(\|z-\mu_0\|)\left\|(v-v_0)'\left(\frac{z-\mu_0}{\|z-\mu_0\|} - \frac{z}{\|z\|}\right) + v'\left(\frac{z-\mu}{\|z-\mu\|} - \frac{z-\mu_0}{\|z-\mu_0\|}\right)\right\| \\
&\leq \rho(\|z-\mu_0\|)\left(2\|v-v_0\|\frac{\|\mu_0\|}{\|z-\mu_0\|} + 2\|v\|\frac{\|\mu-\mu_0\|}{\|z-\mu_0\|}\right) \\
&\leq C\psi_-(\|z-\mu_0\|)\big(\|v-v_0\| + \|\mu-\mu_0\|\big),
\end{aligned}$$

where we used the fact that $\rho(t)/t \leq \psi_-(t)$ for any $t \in (0, \infty)$. Obviously, if $z = \mu_0$, then $T_3(z) = 0$, whereas if $z = \mu(\neq \mu_0)$, then

$$|T_3(z)| \leq 4\rho(\|\mu-\mu_0\|) \leq 4\psi_-(\|\mu-\mu_0\|)\|\mu-\mu_0\| = 4\psi_-(\|z-\mu_0\|)\|\mu-\mu_0\|.$$

Finally, if $z = 0 \notin \{\mu, \mu_0\}$, then Lemma 4.8.5 provides

$$\begin{aligned}
T_3(z) &\leq \rho(\|\mu_0\|)\left|\frac{v_0'\mu_0}{\|\mu_0\|} - \frac{v'\mu}{\|\mu\|}\right| \leq \rho(\|\mu_0\|)\left|(v_0-v)'\frac{\mu_0}{\|\mu_0\|} + v'\left(\frac{\mu_0}{\|\mu_0\|} - \frac{\mu}{\|\mu\|}\right)\right| \\
&\leq \rho(\|\mu_0\|)\left(\|v-v_0\| + \left\|\frac{\mu}{\|\mu\|} - \frac{\mu_0}{\|\mu_0\|}\right\|\right) \leq \rho(\|\mu_0\|)\left(\|v-v_0\| + 2\frac{\|\mu-\mu_0\|}{\|\mu_0\|}\right) \\
&\leq \rho(\|\mu_0\|)\|v-v_0\| + 2\psi_-(\|\mu_0\|)\|\mu-\mu_0\| \\
&\leq C\Big(\|v-v_0\| + \|\mu-\mu_0\|\Big),
\end{aligned}$$

which shows that (3.9.11) holds for $T_3(z)$, too. The result is proved. ■

Note that if the assumption that $P \in \mathcal{P}_d^\rho$ is reinforced into the assumption that

$$\int_{\mathbb{R}^d} \psi_-(\|z\| + c)\, dP(z) < \infty$$

for any $c > 0$, then the proof of Lemma 3.9.5 shows that $(\mu, \alpha, u) \mapsto M^\rho_{\alpha,u}(\mu)$ is actually Lipschitz over $K \times [0,1] \times \mathcal{S}^{d-1}$ for any compact set $K \subset \mathbb{R}^d$. In particular, for $\rho(t) = t^p$ with $p \geq 1$, if P has finite moments of order $p-1$, then $(\mu, \alpha, u) \mapsto M^\rho_{\alpha,u}(\mu)$ is not only locally Lipschtiz (Lemma 3.9.5) but satisfies this global Lipschtiz property over compact sets.

The next result states a coercivity property for the objective function.

Lemma 3.9.6. *Let $\rho \in \mathcal{C}$ and $P \in \mathcal{P}_d^\rho$. Fix sequences $(\alpha_\ell) \in [0,1]$, (u_ℓ) in $\mathcal{S}^{d-1}$, and (μ_ℓ) in $\mathbb{R}^d$ such that $c := \limsup_{\ell\to\infty} \alpha_\ell u_\ell'(\mu_\ell/\|\mu_\ell\|)\xi_{\mu_\ell,0} < 1$ and $\|\mu_\ell\| \to \infty$. Then,*

$$\liminf_{\ell\to\infty} \frac{M^\rho_{\alpha_\ell,u_\ell}(\mu_\ell)}{\|\mu_\ell\|} > 0.$$

In particular, $M^\rho_{\alpha_\ell,u_\ell}(\mu_\ell) \to \infty$.

Proof of Lemma 3.9.6. Of course, we may assume without any loss of generality that $\mu_\ell \neq 0$ for any ℓ. Note that if $P[\{0\}] = 1$, then we have $M^\rho_{\alpha_\ell,u_\ell}(\mu_\ell) = H^\rho_{\alpha_\ell,u_\ell}(-\mu_\ell) = \rho(\|\mu_\ell\|)(1 - \alpha_\ell u'_\ell \mu_\ell/\|\mu_\ell\|)$, so that the fact that $t \mapsto \rho(t)/t$ is monotone non-decreasing (Lemma 3.9.4) yields

$$\liminf_{\ell\to\infty} \frac{M^\rho_{\alpha_\ell,u_\ell}(\mu_\ell)}{\|\mu_\ell\|} \geq (1-c) \liminf_{\ell\to\infty} \frac{\rho(\|\mu_\ell\|)}{\|\mu_\ell\|} \geq (1-c)\rho(1) > 0.$$

We may thus assume that $P[\{0\}] < 1$. Write then

$$\frac{M^\rho_{\alpha_\ell,u_\ell}(\mu_\ell)}{\|\mu_\ell\|} = \mathcal{I}_1(\alpha_\ell, u_\ell, \mu_\ell) + \mathcal{I}_2(\alpha_\ell, u_\ell, \mu_\ell) + \mathcal{I}_3(\alpha_\ell, u_\ell, \mu_\ell), \tag{3.9.12}$$

with

$$\mathcal{I}_1(\alpha, u, \mu) = \int_{\mathbb{R}^d} \frac{\rho(\|z-\mu\|) - \rho(\|z\|)}{\|\mu\|} \left(1 + \alpha \frac{u'(z-\mu)}{\|z-\mu\|}\right) \xi_{z,\mu}\, dP(z),$$

$$\begin{aligned}\mathcal{I}_2(\alpha, u, \mu) &= \int_{\mathbb{R}^d\setminus\{0,\mu\}} \frac{\rho(\|z\|)}{\|\mu\|} \left\{ \left(1 + \alpha \frac{u'(z-\mu)}{\|z-\mu\|}\right)\xi_{z,\mu} - \left(1 + \alpha\frac{u'z}{\|z\|}\right)\xi_{z,0} \right\} dP(z) \\ &= \int_{\mathbb{R}^d\setminus\{0\}} \alpha \frac{\rho(\|z\|)}{\|\mu\|} \left(\frac{u'(z-\mu)}{\|z-\mu\|} - \frac{u'z}{\|z\|}\right) \xi_{z,\mu}\, dP(z),\end{aligned} \tag{3.9.13}$$

and

$$\begin{aligned}\mathcal{I}_3(\alpha, u, \mu) &= \int_{\{0,\mu\}} \frac{\rho(\|z\|)}{\|\mu\|} \left\{ \left(1 + \alpha \frac{u'(z-\mu)}{\|z-\mu\|}\right)\xi_{z,\mu} - \left(1 + \alpha\frac{u'z}{\|z\|}\right)\xi_{z,0} \right\} dP(z) \\ &= -\frac{\rho(\|\mu\|)}{\|\mu\|}\left(1 + \alpha\frac{u'\mu}{\|\mu\|}\right) P[\{\mu\}]\xi_{\mu,0}.\end{aligned}$$

Consider first $\mathcal{I}_1(\alpha_\ell, u_\ell, \mu_\ell)$. If $\|z-\mu_\ell\| > \|z\|$, then Lemma 3.9.3 provides

$$\rho(\|z-\mu_\ell\|) - \rho(\|z\|) \geq \psi_-(c_\ell^-)\big(\|z-\mu_\ell\| - \|z\|\big) \geq \psi_-(\|z\|)\big(\|z-\mu_\ell\| - \|z\|\big),$$

with c_ℓ^- between $\|z-\mu_\ell\|$ and $\|z\|$, whereas if $\|z-\mu_\ell\| < \|z\|$, then the same result yields

$$\rho(\|z\|) - \rho(\|z-\mu_\ell\|) \leq \psi_-(c_\ell^+)\big(\|z\| - \|z-\mu_\ell\|\big) \leq \psi_-(\|z\|)\big(\|z\| - \|z-\mu_\ell\|\big),$$

still with c_ℓ^+ between $\|z-\mu_\ell\|$ and $\|z\|$. Thus, for any $z \in \mathbb{R}^d$, we have

$$\rho(\|z-\mu_\ell\|) - \rho(\|z\|) \geq \psi_-(\|z\|)\big(\|z-\mu_\ell\| - \|z\|\big),$$

which provides

$$\mathcal{I}_1(\alpha_\ell, u_\ell, \mu_\ell) \geq \tilde{\mathcal{I}}_1(\alpha_\ell, u_\ell, \mu_\ell) = \int_{\mathbb{R}^d} h_{\alpha_\ell,u_\ell,\mu_\ell}(z)\, dP(z),$$

where we let

$$h_{\alpha,u,\mu}(z) = \psi_-(\|z\|) \frac{\|z-\mu\| - \|z\|}{\|\mu\|}\left(1 + \alpha\frac{u'(z-\mu)}{\|z-\mu\|}\right)\xi_{z,\mu}.$$

Since the function $z \mapsto |h_{\alpha_\ell,u_\ell,\mu_\ell}(z)|$ is upper-bounded by the P-integrable function $z \mapsto 2\psi_-(\|z\|)$ uniformly in ℓ, Fatou's lemma shows that

$$\liminf_{\ell\to\infty} \mathcal{I}_1(\alpha_\ell, u_\ell, \mu_\ell) \geq \liminf_{\ell\to\infty} \tilde{\mathcal{I}}_1(\alpha_\ell, u_\ell, \mu_\ell) \geq \int_{\mathbb{R}^d} \liminf_{\ell\to\infty} h_{\alpha_\ell,u_\ell,\mu_\ell}(z)\, dP(z)$$

$$\geq \int_{\mathbb{R}^d} \liminf_{\ell\to\infty} \Big(\psi_-(\|z\|)\frac{\|z-\mu_\ell\|-\|z\|}{\|\mu_\ell\|}\xi_{z,\mu_\ell}\Big) \liminf_{\ell\to\infty} \Big(1+\alpha_\ell \frac{u_\ell'(z-\mu_\ell)}{\|z-\mu_\ell\|}\Big) dP(z)$$

$$= (1-c)\int_{\mathbb{R}^d} \psi_-(\|z\|)\, dP(z) \geq (1-c)\int_{\mathbb{R}^d\setminus\{0\}} \psi_-(\|z\|)\, dP(z) > 0,$$

where the last inequality follows from the fact that $P[\mathbb{R}^d \setminus \{0\}] > 0$ and that $\psi_-(t) > 0$ for any $t > 0$ (Lemma 3.9.4).

Let us turn to $\mathcal{I}_2(\alpha_\ell, u_\ell, \mu_\ell)$. For any $z \in \mathbb{R}^d \setminus \{0\}$, we have $\rho(\|z\|) \leq \|z\|\psi_-(\|z\|)$ and

$$\Big|\frac{u_\ell'(z-\mu_\ell)}{\|z-\mu_\ell\|} - \frac{u_\ell' z}{\|z\|}\Big|\xi_{z,\mu_\ell} \leq \frac{2\|\mu_\ell\|}{\|z\|}$$

(Lemma 4.8.5). Therefore, the absolute value of the integrand in (3.9.13) is upper-bounded by the P-integrable function $z \mapsto 2\psi_-(\|z\|)$. Since this function does not depend on ℓ, Lebesgue's Dominated Convergence Theorem (DCT) shows that

$$\lim_{\ell\to\infty} \mathcal{I}_2(\alpha_\ell, u_\ell, \mu_\ell) = \int_{\mathbb{R}^d\setminus\{0\}} \lim_{\ell\to\infty} \alpha_\ell \frac{\rho(\|z\|)}{\|\mu_\ell\|}\Big(\frac{u_\ell'(z-\mu_\ell)}{\|z-\mu_\ell\|} - \frac{u_\ell' z}{\|z\|}\Big)\xi_{z,\mu_\ell}\, dP(z) = 0.$$

Finally, using again the identity $\rho(\|z\|) \leq \|z\|\psi_-(\|z\|)$, we obtain

$$\liminf_{\ell\to\infty} \mathcal{I}_3(\alpha_\ell, u_\ell, \mu_\ell) \geq -2\limsup_{\ell\to\infty}\Big(\psi_-(\|\mu_\ell\|)P[\{\mu_\ell\}]\Big) = 0,$$

where the limsup vanishes by Lemma 3.9.1 applied to the measure attributing to any d-Borel set B the measure $\nu(B) = \int_B \psi_-(\|z\|)dP(z)$ (which is finite by assumption). Therefore,

$$\liminf_{\ell\to\infty} \frac{M^\rho_{\alpha_\ell,u_\ell}(\mu_\ell)}{\|\mu_\ell\|} \geq \liminf_{\ell\to\infty} \mathcal{I}_1(\alpha_\ell, u_\ell, \mu_\ell) > 0,$$

which establishes the result. ■

We can now prove Part (ii) of Theorem 3.2.1.

Proof of Theorem 3.2.1(ii). Theorem 3.2.1(i) ensures that the map $\mu \mapsto M^\rho_{\alpha,u}(\mu)$ is well-defined for any $\mu \in \mathbb{R}^d$. Pick $R > 0$ such that $M^\rho_{\alpha,u}(\mu) > M^\rho_{\alpha,u}(0)$ for any $\|\mu\| > R$ (existence follows from Lemma 3.9.6 since $\alpha < 1$). From continuity (Lemma 3.9.5), the map $\mu \mapsto M^\rho_{\alpha,u}(\mu)$ admits a minimum over the compact set $K = \{\mu \in \mathbb{R}^d : \|\mu\| \leq R\}$. Since $M^\rho_{\alpha,u}(\mu) > M^\rho_{\alpha,u}(0)$ for any $\|\mu\| > R$, this minimum over K is actually a minimum over $\mathbb{R}^d$. ■

Proofs for Section 3.3

PROOF OF LEMMA 3.3.1. (i) By definition, $\mathcal{C}_0 \subseteq \mathcal{C}$. Now, fix $\rho \in \mathcal{C}$. Since ρ is monotone non-decreasing and convex, we have $\rho(\|(1-\lambda)x + \lambda y\|) \leq \rho((1-\lambda)\|x\| + \lambda\|y\|) \leq (1-\lambda)\rho(\|x\|) + \lambda\rho(\|y\|)$ for any $x, y \in \mathbb{R}^d$ and any $\lambda \in [0,1]$. This shows that $z \mapsto H^{\rho}_{0,u}(z) = \rho(\|z\|)$ is convex for any $u \in \mathcal{S}^{d-1}$, hence that $\rho \in \mathcal{C}_0$. Therefore, we also have $\mathcal{C} \subseteq \mathcal{C}_0$. (ii) For any $\alpha \in [0,1]$, define

$$V_\alpha := \big\{\rho \in \mathcal{C} : g^{\rho}_{\alpha}(x,y) := a(x,y) - \alpha\|V(x,y)\| \geq 0 \ \forall x, y \in \mathbb{R}^d\big\},$$

with

$$a(x,y) := \rho(\|x\|) + \rho(\|y\|) - 2\rho\bigg(\frac{\|x+y\|}{2}\bigg)$$

and

$$V(x,y) := \rho(\|x\|)\frac{x}{\|x\|}\xi_{x,0} + \rho(\|y\|)\frac{y}{\|y\|}\xi_{y,0} - 2\rho\bigg(\frac{\|x+y\|}{2}\bigg)\frac{x+y}{\|x+y\|}\xi_{x+y,0}.$$

Note that $H^{\rho}_{\alpha,u}(x) + H^{\rho}_{\alpha,u}(y) - 2H^{\rho}_{\alpha,u}((x+y)/2) = a(x,y) + \alpha u'V(x,y)$. Since we trivially have that $V_{\alpha_2} \subseteq V_{\alpha_1}$ for any $\alpha_1 < \alpha_2$, it is sufficient to prove that $\mathcal{C}_\alpha = V_\alpha$ for any $\alpha \in [0,1]$.

Fix first $\rho \in \mathcal{C}_\alpha$. Then, for any $x, y \in \mathbb{R}^d$ and any $u \in \mathcal{S}^{d-1}$, we have $a(x,y) + \alpha u'V(x,y) \geq 0$. If $V(x,y) = 0$, then $g^{\rho}_{\alpha}(x,y) = a(x,y) \geq 0$ since $x \mapsto \rho(\|x\|)$ is convex (see Part (i) of the result), whereas if $V(x,y) \neq 0$, then taking $u_0 = -V(x,y)/\|V(x,y)\|$ yields $g^{\rho}_{\alpha}(x,y) = a(x,y) + \alpha u_0'V(x,y) \geq 0$ since $x \mapsto H^{\rho}_{\alpha,u_0}(x)$ is convex. Thus, $g^{\rho}_{\alpha}(x,y) \geq 0$ for any $x, y \in \mathbb{R}^d$, so that $\rho \in V_\alpha$. Now, fix $\rho \in V_\alpha$. Then, the Cauchy–Schwarz inequality ensures that, for any $x, y \in \mathbb{R}^d$ and any $u \in \mathcal{S}^{d-1}$, one has $a(x,y) + \alpha u'V(x,y) \geq g^{\rho}_{\alpha}(x,y) \geq 0$. This shows that, for any $u \in \mathcal{S}^{d-1}$, the map $H^{\rho}_{\alpha,u}$ is midpoint convex, hence convex (from continuity). In other words, $\rho \in \mathcal{C}_\alpha$. ■

The following result plays a key role in the proof of Theorem 3.3.2.

Lemma 3.9.7. *Let $\rho \in \mathcal{C}$ and fix $t \in \mathcal{D}_\rho$. Then, the Hessian matrix $\nabla^2 H^{\rho}_{1,u}(x)$ is positive semi-definite for any $u \in \mathcal{S}^{d-1}$ and any $x \in \mathbb{R}^d$ with $\|x\| = t$ if and only if the second-order derivative of $s \mapsto s^2/\rho(s)$ at t is nonpositive.*

PROOF OF LEMMA 3.9.7. Throughout the proof, we write $y = x/\|x\|$. For any $u, v \in \mathcal{S}^{d-1}$ and any x such that $\|x\| = t$, Lemma 3.9.13 (whose proof is independent) entails that

$$
\begin{aligned}
& v'\nabla^2 H^{\rho}_{1,u}(x)v \\
&= \frac{t^2\psi'_-(t) - 2t\psi_-(t) + 2\rho(t)}{t^2}(1+u'y)(v'y)^2 + \frac{\rho(t)}{t^2}(1-(v'y)^2) \\
&\quad + \frac{t\psi_-(t) - \rho(t)}{t^2}\{(1+u'y)(1-(v'y)^2) + 2(v'y)^2 + 2(v'y)(u'v)\} \\
&= \psi'_-(t)(v'y)^2 + \frac{\psi_-(t)}{t}(1-(v'y)^2) \\
&\quad + u'\Big\{\psi'_-(t)(v'y)^2 y + \frac{t\psi_-(t)-\rho(t)}{t^2}(1-(v'y)^2)y \\
&\qquad + \frac{2(t\psi_-(t)-\rho(t))}{t^2}(v'y)\{v-(v'y)y\}\Big\} = a + u'\{(a-b)y + cy^\perp\},
\end{aligned}
$$

with

$$
a = \psi'_-(t)(v'y)^2 + \frac{\psi_-(t)}{t}(1-(v'y)^2), \quad b = \frac{\rho(t)}{t^2}(1-(v'y)^2),
$$

$$
c = \frac{2(t\psi_-(t)-\rho(t))}{t^2}(v'y), \quad \text{and} \quad y^\perp = (I_d - yy')v.
$$

Clearly, $v'\nabla^2 H^{\rho}_{1,u}(ty)v \geq 0$ for any $u, v, y \in \mathcal{S}^{d-1}$ if and only if $a - \|(a-b)y + cy^\perp\| \geq 0$ for any $v, y \in \mathcal{S}^{d-1}$. Since $y'y^\perp = 0$ and since $a \geq 0$ (recall that ρ is convex and non-decreasing), this is in turn equivalent to requiring that $2ab - b^2 - c^2\|y^\perp\|^2 \geq 0$, that is,

$$
\begin{aligned}
& 2\Big(\psi'_-(t)(v'y)^2 + \frac{\psi_-(t)}{t}(1-(v'y)^2)\Big)\frac{\rho(t)}{t^2} \\
&\qquad - \frac{(\rho(t))^2}{t^4}(1-(v'y)^2) - \frac{4(t\psi_-(t)-\rho(t))^2}{t^4}(v'y)^2 \geq 0
\end{aligned}
$$

for any $v, y \in \mathcal{S}^{d-1}$ with $v \notin \{\pm y\}$, or again that

$$
(1-(v'y)^2)\Big(2\psi_-(t) - \frac{\rho(t)}{t}\Big)\frac{\rho(t)}{t^3} + (v'y)^2\frac{2}{t^4}\Big(t^2\rho(t)\psi'_-(t) - 2(t\psi_-(t)-\rho(t))^2\Big) \geq 0
$$

for any such v, y. Since the assumptions on ρ imply that $2\psi_-(t) - \rho(t)/t \geq \psi_-(t) \geq 0$, this is equivalent to requiring that $t^2\rho(t)\psi'_-(t) - 2(t\psi_-(t)-\rho(t))^2 \geq 0$, that is,

$$
-(\rho(t))^3 \frac{d^2}{ds^2}\frac{s^2}{\rho(s)}\bigg|_{s=t} = t^2\rho(t)\psi'_-(t) + 4t\rho(t)\psi_-(t) - 2t^2(\psi_-(t))^2 - 2(\rho(t))^2 \geq 0.
$$

This establishes the result. ■

The proof of Theorem 3.3.2 further requires the following result.

Lemma 3.9.8. *Let $g : [a,b] \to \mathbb{R}$ be a continuous function that is differentiable in $[a,b] \setminus \{s_1, \ldots, s_k\}$, with $s_0 := a < s_1 < \ldots < s_k < b =: s_{k+1}$ ($k \in \mathbb{N}$). Assume that,* *for any* $j = 1, \dots, k+1$, g' *is monotone non-decreasing in* (s_{j-1}, s_j), *and that, for any* $j = 1, \dots, k$,

$$\ell_j := \lim_{s \overset{<}{\to} s_j} g'(s) \quad \textit{and} \quad r_j := \lim_{s \overset{>}{\to} s_j} g'(s)$$

exist, are finite, and satisfy $\ell_j \leq r_j$. *Then,* g *is convex on* $[a, b]$.

Proof of Lemma 3.9.8. For any $j = 1, \dots, k$, L'Hospital's rule ensures that

$$\lim_{s \overset{<}{\to} s_j} \frac{g(s) - g(s_j)}{s - s_j} = \ell_j,$$

which shows that g is actually left-differentiable in (a, b). The monotonicity assumption on g' and the assumption that $\ell_j \leq r_j$ for any $j = 1, \dots, k$, then entail that, g'_-, the left-derivative of g, is monotone non-decreasing in $(0, 1)$.

Now, assume ad absurdum that g is not convex $[a, b]$. Then, there exist $u, v, w \in [a, b]$, with $u < v < w$, such that

$$\frac{g(v) - g(u)}{v - u} > \frac{g(w) - g(v)}{w - v}.$$

Lemma 3.9.3(i) then ensures that there exist $\eta \in (u, v)$ and $\xi \in (v, w)$ such that $g'(\eta) > g'(\xi)$. Therefore, g'_- is not monotone non-decreasing in $(0, 1)$, a contradiction. ■

Proof of Theorem 3.3.2. Assume first that $\alpha_\rho = 1$, so that $H^\rho_{1,u}$ is convex for any $u \in \mathcal{S}^{d-1}$. Fix a positive integer k and let $\mathcal{U}_k = (t_{k-1}, t_k)$, where the t_ℓ's are the endpoints of the intervals on which ρ is twice continuously differentiable (see the beginning of Section 3.2). Thus, $v'\nabla^2 H^\rho_{1,u}(x)v \geq 0$ for any $u, v \in \mathcal{S}^{d-1}$ and $x \in E_k = \{x \in \mathbb{R}^d : \|x\| \in \mathcal{U}_k\}$. Letting $f(t) := \frac{d}{dt}(t^2/\rho(t))$ for any $t \in \mathcal{U}_k$, Lemma 3.9.7 then yields that f is monotone non-increasing on $\mathcal{U}_k$. Now, convexity of ρ implies that

$$\lim_{t \overset{<}{\to} t_k} f(t) = \frac{2t_k\rho(t_k) - t_k^2\psi_-(t_k)}{(\rho(t_k))^2} \geq \frac{2t_k\rho(t_k) - t_k^2\psi_+(t_k)}{(\rho(t_k))^2} = \lim_{t \overset{>}{\to} t_k} f(t)$$

(recall that ψ_- and ψ_+ are the left- and right-derivatives of ρ, respectively). Since this holds for any positive integer k, we conclude that f is monotone non-increasing on $(0, \infty)$, hence that $t \mapsto t^2/\rho(t)$ is concave on $(0, \infty)$.

Assume now that $t \mapsto t^2/\rho(t)$ is concave on $(0, \infty)$. Fix $x, y \in \mathbb{R}^d$ linearly independent and let $\Gamma(s) = (1 - s)x + sy$, $s \in [0, 1]$. Then, $\Gamma(s) \neq 0$ for any $s \in [0, 1]$. Let $s_1, \dots, s_r$ be the values in $(0, 1)$ for which $\|\Gamma(s)\| \notin \mathcal{D}_\rho$. Since $s \mapsto \|\Gamma(s)\|^2$ is convex, r is finite. Letting $s_0 = 0$ and $s_{r+1} = 1$,

$$s \mapsto g(s) = H^\rho_{1,u}(\Gamma(s)) = \rho(\|\Gamma(s)\|)\left(1 + \frac{u'\Gamma(s)}{\|\Gamma(s)\|}\right)$$

is twice continuously differentiable on (s_j, s_{j+1}) for any $j = 0, 1, \dots, r$. Fix then such a value of j. Since $\Gamma''(s) = 0$ for any $s \in (s_j, s_{j+1})$, we have

$$g''(s) = (\Gamma'(s))'\nabla^2 H^\rho_{1,u}(\Gamma(s))\Gamma'(s) \geq 0$$

for any $s \in (s_j, s_{j+1})$ (non-negativity follows from Lemma 3.9.7 since the concavity assumption ensures that $\frac{d^2}{dt^2}(t^2/\rho(t)) \leq 0$ for any $t \in \mathcal{D}_\rho$). This implies that g is convex on any interval (s_j, s_{j+1}). Now, noting that the derivative of $s \mapsto \|\Gamma(s)\|$ is $(y-x)'\Gamma(s)/\|\Gamma(s)\|$, we have, for any $j = 1, \ldots, r$,

$$\lim_{s \overset{>}{\to} s_j} g'(s) - \lim_{s \overset{<}{\to} s_j} g'(s)$$

$$= (\psi_+(\|\Gamma(s_j)\|) - \psi_-(\|\Gamma(s_j)\|)) \frac{(y-x)'\Gamma(s_j)}{\|\Gamma(s_j)\|}\left(1 + \frac{u'\Gamma(s_j)}{\|\Gamma(s_j)\|}\right)$$

if $(y-x)'\Gamma(s_j) \geq 0$, and

$$\lim_{s \overset{>}{\to} s_j} g'(s) - \lim_{s \overset{<}{\to} s_j} g'(s)$$

$$= (\psi_-(\|\Gamma(s_j)\|) - \psi_+(\|\Gamma(s_j)\|)) \frac{(y-x)'\Gamma(s_j)}{\|\Gamma(s_j)\|}\left(1 + \frac{u'\Gamma(s_j)}{\|\Gamma(s_j)\|}\right)$$

if $(y-x)'\Gamma(s_j) < 0$. In both cases, convexity of ρ implies that

$$\lim_{s \overset{>}{\to} s_j} g'(s) \geq \lim_{s \overset{<}{\to} s_j} g'(s)$$

for any $j = 1, \ldots, r$. Therefore, Lemma 3.9.8 entails that g is convex on $[0,1]$. This yields that

$$H^\rho_{1,u}((1-\lambda)x + \lambda y) \leq (1-\lambda)H^\rho_{1,u}(x) + \lambda H^\rho_{1,u}(y), \quad \forall \lambda \in (0,1), \tag{3.9.14}$$

for any $x, y \in \mathbb{R}^d$ that are linearly independent. Since (3.9.14) also holds if x, y are linearly dependent (it is easy to check that, for any $y \in \mathcal{S}^{d-1}$, the map $t \mapsto H^\rho_{1,u}(ty) = \rho(|t|)(1 + \mathrm{Sign}(t)u'y)$ is convex over $\mathbb{R}$), we conclude that $H^\rho_{1,u}$ is convex on $\mathbb{R}^d$ for any $u \in \mathcal{S}^{d-1}$. The result then follows from Lemma 3.3.1. ∎

Proof of Theorem 3.3.3. Assume first that ρ is twice continuously differentiable on $(0,\infty)$, that is, assume that $\mathcal{D}_\rho = (0,\infty)$. Fix $t \in \mathcal{D}_\rho$ such that $(t^2/\rho(t))'' > 0$. Recall that convexity of ρ implies that $t\psi_-(t) \geq \rho(t)$. If $t\psi_-(t) = \rho(t)$, then we have

$$\left(\frac{t^2}{\rho(t)}\right)'' = -\frac{t^2\rho(t)\psi'_-(t) + 4t\rho(t)\psi_-(t) - 2t^2(\psi_-(t))^2 - 2(\rho(t))^2}{(\rho(t))^3} = -\frac{t^2\psi'_-(t)}{(\rho(t))^2},$$

which implies that $\psi'_-(t) < 0$. Since this is incompatible with the convexity of ρ, we must have $t\psi_-(t) > \rho(t)$. In other words,

$$\begin{aligned} \mathcal{D}^{\mathrm{cv}}_\rho &= \{t \in \mathcal{D}_\rho : (t^2/\rho(t))'' > 0\} \\ &= \{t \in \mathcal{D}_\rho : t\psi_-(t) > \rho(t) \text{ and } (t^2/\rho(t))'' > 0\}, \end{aligned}$$

which allows us to partition $\mathcal{D}_\rho$ into $\mathcal{D}^{\mathrm{cv}}_\rho \cup \mathcal{E}_\rho \cup \mathcal{F}_\rho$, with

$$\mathcal{E}_\rho := \{t \in \mathcal{D}_\rho : t\psi_-(t) > \rho(t) \text{ and } (t^2/\rho(t))'' \leq 0\}$$

and

$$\mathcal{F}_\rho := \{t \in \mathcal{D}_\rho : t\psi_-(t) = \rho(t)\}.$$

Now, note that, for $t \in \mathcal{F}_\rho$, Lemma 3.9.13 provides

$$v'\nabla^2 H^\rho_{\alpha,u}(ty)v = \psi'_-(t)(1+\alpha u'y)(v'y)^2 + \frac{\rho(t)}{t^2}(1-(v'y)^2) \geq 0, \tag{3.9.15}$$

for any $\alpha \in [0,1]$ and $u, v, y \in \mathcal{S}^{d-1}$. For $t \in \mathcal{D}_\rho \setminus \mathcal{F}_\rho$, the quantity q_t is well-defined, and we have $(p_t, q_t) \in (1,\infty) \times [0,\infty)$; here, p_t and q_t refer to the quantities defined in the statement of the theorem. For $t \in \mathcal{D}_\rho \setminus \mathcal{F}_\rho$, Lemma 3.9.13 then yields

$$\begin{aligned}\frac{t^2}{\rho(t)} v'\nabla^2 H^\rho_{\alpha,u}(ty)v &= (p_t-1)(q_t-2)(1+\alpha u'y)(v'y)^2 + 1 - (v'y)^2 \\ &+(p_t-1)\big\{(1+\alpha u'y)(1-(v'y)^2) + 2(v'y)^2 + 2\alpha(v'y)(u'v)\big\} =: g_{t,\alpha}(u,v,y).\end{aligned}$$

With an obvious abuse of notation, write

$$\begin{aligned}g_{t,\alpha}(\theta,\omega) &= (p_t-1)(q_t-2)(1+\alpha\cos\theta)(\cos\omega)^2 + 1 - (\cos\omega)^2 \\ &+(p_t-1)\big\{(1+\alpha\cos\theta)(1-(\cos\omega)^2) + 2(\cos\omega)^2 + 2\alpha(\cos\omega)\cos(\theta-\omega)\big\},\end{aligned}$$

where $\omega = \arccos(v'y) \in [0,\pi]$ is the angle between v and y and $\theta = s_{u,v,y}\arccos(u'y) \in [-\pi,\pi]$ is the signed angle between u and y; here, $s_{u,v,y} = 1$ (resp., $s_{u,v,y} = -1$) if, in the plane containing u, v, y, we have that u and v are not separated (resp., are separated) by the line through $\pm y$.

Let

$$\tilde{\alpha}_\rho := \inf_{t \in \mathcal{D}^{\rm cv}_\rho} \tilde{\alpha}_{t,\rho}, \quad \text{with } \tilde{\alpha}_{t,\rho} := \sqrt{\frac{q_t(4p_t^2 - 4p_t - q_t)}{4(p_t-1)^2(q_t+1)}}$$

(the assumption that $t \mapsto t^2/\rho(t)$ is not concave on $(0,\infty)$ ensures that $\mathcal{D}^{\rm cv}_\rho$ is non-empty, so that $\tilde{\alpha}_\rho$ is well-defined). We will show that

- (i)$_a$ for any $t \in \mathcal{D}^{\rm cv}_\rho$ and (θ,ω), $g_{t,\tilde{\alpha}_\rho}(\theta,\omega) \geq 0$;
- (i)$_b$ for any $t \in \mathcal{E}_\rho$ and (θ,ω), $g_{t,\tilde{\alpha}_\rho}(\theta,\omega) \geq 0$;
- (ii) for any $\alpha > \tilde{\alpha}_\rho$, there exist $t \in \mathcal{D}^{\rm cv}_\rho$ and (θ,ω) such that $g_{t,\alpha}(\theta,\omega) < 0$.

Jointly with (3.9.15), (i)$_a$–(i)$_b$ establish that $\alpha_\rho \geq \tilde{\alpha}_\rho$, whereas (ii) entails that $\alpha_\rho \leq \tilde{\alpha}_\rho$. Therefore, it is sufficient to prove (i)$_a$–(ii). To do so, fix $t \in \mathcal{D}_\rho \setminus \mathcal{F}_\rho$ and write

$$g_{t,\alpha}(\theta,\omega) = i_t(\omega) + \alpha s_t(\theta,\omega),$$

where the intercept function $i_t(\omega)$ and slope function $s_t(\theta,\omega)$ are defined as

$$i_t(\omega) := p_t + (p_tq_t - p_t - q_t)(\cos\omega)^2$$

and

$$s_t(\theta,\omega) := (p_t-1)\{(1+(q_t-1)(\cos\omega)^2)\cos\theta + \sin(2\omega)\sin\theta\}.$$

Note that $i_t(\omega) \geq \min(p_t, p_t + (p_t q_t - p_t - q_t)) = \min(p_t, (p_t - 1)q_t) \geq 0$. Actually, if $i_t(\omega) = 0$, then $q_t = 0$ and $(\cos\omega)^2 = 1$, which yields $g_{t,\alpha}(\theta,\omega) = 0$ for any $\alpha \in [0,1]$ and $(\theta,\omega) \in [-\pi,\pi] \times [0,\pi]$. We may thus ignore this case when investigating when $g_{t,\alpha}(\theta,\omega)$ is negative. Assume thus that $i_t(\omega) > 0$. Defining then

$$\alpha_t(\theta,\omega) := \begin{cases} -i_t(\omega)/s_t(\theta,\omega) & \text{if } s_t(\theta,\omega) < 0 \\ \infty & \text{otherwise,} \end{cases}$$

we have that, in the range $\alpha \in [0,\infty)$, $g_{t,\alpha}(\theta,\omega) \geq 0$ if and only if $\alpha \leq \alpha_t(\theta,\omega)$ (if $\alpha_t(\theta,\omega) = \infty$, then this is obviously to be read as $g_{t,\alpha}(\theta,\omega) \geq 0$ for any $\alpha \in [0,\infty)$).

Assume for a moment that

$$\min\left\{\alpha_t(\theta,\omega) : (\theta,\omega) \in [-\pi,\pi] \times [0,\pi]\right\} = \begin{cases} \tilde{\alpha}_{t,\rho} (< 1) & \text{for } t \in \mathcal{D}_\rho^{\text{cv}} \\ 1 & \text{for } t \in \mathcal{E}_\rho. \end{cases} \tag{3.9.16}$$

Then, for any $t \in \mathcal{D}_\rho^{\text{cv}}$ and $(\theta,\omega) \in [-\pi,\pi] \times [0,\pi]$, we have $\tilde{\alpha}_\rho \leq \tilde{\alpha}_{t,\rho} \leq \alpha_t(\theta,\omega)$, so that $g_{t,\tilde{\alpha}_\rho}(\theta,\omega) \geq 0$. This establishes (i)$_a$ above. Similarly, for any $t \in \mathcal{E}_\rho$ and $(\theta,\omega) \in [-\pi,\pi] \times [0,\pi]$, we then have $\tilde{\alpha}_\rho \leq 1 \leq \alpha_t(\theta,\omega)$, so that $g_{t,\tilde{\alpha}_\rho}(\theta,\omega) \geq 0$, which proves (i)$_b$. Finally, for any $\alpha > \tilde{\alpha}_\rho$, there exists $t \in \mathcal{D}_\rho^{\text{cv}}$ such that $\alpha > \tilde{\alpha}_{t,\rho}$, which, according to (3.9.16), implies that there exists (θ,ω) such that $\alpha > \alpha_t(\theta,\omega)$. With these t, θ and ω, we then have $g_{t,\alpha}(\theta,\omega) < 0$, which proves (ii) above. Therefore, it only remains to establish (3.9.16).

To do so, fix $t \in \mathcal{D}_\rho^{\text{cv}} \cup \mathcal{E}_\rho$ arbitrarily. For any fixed $\omega \in [0,\pi]$, the fact that $i_t(\omega)$ is positive and does not depend on θ implies that

$$\alpha_t(\omega) := \min\{\alpha_t(\theta,\omega) : \theta \in [-\pi,\pi]\} = -\frac{i_t(\omega)}{\min\{s_t(\theta,\omega) : \theta \in [-\pi,\pi]\}}.$$

Since we safely excluded the case for which $q_t = 0$ and $(\cos\omega)^2 = 1$, we have $1 + (q_t - 1)(\cos\omega)^2 > 0$, so that the Cauchy–Schwarz inequality readily yields

$$\begin{aligned} \alpha_t(\omega) &= \frac{i_t(\omega)}{(p_t - 1)\sqrt{(1 + (q_t - 1)(\cos\omega)^2)^2 + (\sin(2\omega))^2}} \\ &= \frac{p_t + (p_t q_t - p_t - q_t)(\cos\omega)^2}{(p_t - 1)\sqrt{(q_t - 3)(q_t + 1)(\cos\omega)^4 + 2(q_t + 1)(\cos\omega)^2 + 1}}. \end{aligned}$$

Obviously, $\inf\{\alpha_t(\omega) : \omega \in [0,\pi]\} = \inf\{f_t(\lambda) : \lambda \in [0,1]\}$, with

$$f_t(\lambda) = \frac{p_t + (p_t q_t - p_t - q_t)\lambda}{(p_t - 1)\sqrt{(q_t - 3)(q_t + 1)\lambda^2 + 2(q_t + 1)\lambda + 1}}.$$

A direct calculation shows that

$$f_t'(\lambda) = \frac{(2p_t - q_t)(q_t + 1)\lambda - (2p_t + q_t)}{(p_t - 1)((q_t - 3)(q_t + 1)\lambda^2 + 2(q_t + 1)\lambda + 1)^{3/2}}. \tag{3.9.17}$$

We need to consider two cases. (a) $t \in \mathcal{D}_\rho^{\text{cv}}$. Provided that $2p_t - q_t \neq 0$, f_t admits a single critical point, namely

$$\lambda_{t_*} := \frac{2p_t + q_t}{(2p_t - q_t)(q_t + 1)}.$$

It is easy to check that if $2p_t - q_t - 2 > 0$, then $\lambda_{t_*} \in (0,1)$. Writing $u(t) = t^2/\rho(t)$, a direct computation shows that

$$2p_t - q_t - 2 = \frac{t^2 u''(t)}{u(t) - tu'(t)} = \frac{(\rho(t))^2 u''(t)}{t\psi_-(t) - \rho(t)}\cdot \tag{3.9.18}$$

Since $t \in \mathcal{D}_\rho^{\text{cv}}$, this yields $2p_t - q_t - 2 > 0$. Therefore, λ_{t_*} is well-defined and $\lambda_{t_*} \in (0,1)$. Clearly, $f'(\lambda) < 0$ for $\lambda \in [0, \lambda_{t_*})$ and $f'(\lambda) > 0$ for $\lambda \in (\lambda_{t_*}, 1]$, so that

$$\min\big\{\alpha_t(\theta,\omega) : (\theta,\omega) \in [-\pi,\pi] \times [0,\pi]\big\} = \min\{f_t(\lambda) : \lambda \in [0,1]\} = f_t(\lambda_{t_*}),$$

which, after easy computations, provides

$$\min\big\{\alpha_t(\theta,\omega) : (\theta,\omega) \in [-\pi,\pi] \times [0,\pi]\big\} = \tilde{\alpha}_{t,\rho},$$

as was to be showed in (3.9.16); note that $\tilde{\alpha}_{t,\rho} = f_t(\lambda_{t_*}) < f_t(1) = 1$.

(b) $t \in \mathcal{E}_\rho$. For such a t, (3.9.18) entails that $2p_t - q_t - 2 \leq 0$. For any $\lambda \in (0,1)$, we thus have

$$\begin{aligned}\ell_t(\lambda) &:= (2p_t - q_t)(q_t+1)\lambda - (2p_t + q_t)\\ &\leq \max(\ell_t(0), \ell_t(1)) = \max(-(2p_t + q_t), (2p_t - q_t - 2)q_t) \leq 0.\end{aligned}$$

It then follows from (3.9.17) that f_t is monotone non-increasing in $[0,1]$, so that its minimal value over $[0,1]$ is $f_t(1) = 1$. Consequently,

$$\min\big\{\alpha_t(\theta,\omega) : (\theta,\omega) \in [-\pi,\pi] \times [0,\pi]\big\} = 1.$$

This ends the proof of (3.9.16), hence establishes the theorem in the case where ρ is twice continuously differentiable on $(0,\infty)$.

Extension to the general case is then straightforward. To be more specific, assume now that ρ is only piecewise twice continuously differentiable. If $\alpha \leq \tilde{\alpha}_\rho$, then we proved above that the Hessian matrix $\nabla^2 H^\rho_{\alpha,u}(x)$ is positive semi-definite for any $u \in \mathcal{S}^{d-1}$ and $x \in \mathbb{R}^d$ such that $\|x\| = t \in \mathcal{D}_\rho$. Proceeding exactly as in the second part of the proof of Theorem 3.3.2 (with $H^\rho_{\alpha,u}$ instead of $H^\rho_{1,u}$), the convexity of ρ then implies that $x \mapsto H^\rho_{\alpha,u}(x)$ is convex over $\mathbb{R}^d$ for any $u \in \mathcal{S}^{d-1}$, which shows that $\alpha_\rho \geq \tilde{\alpha}_\rho$. Finally, if $\alpha > \tilde{\alpha}_\rho$, then the first part of the proof shows that there exist $t \in \mathcal{D}_\rho$ and $u, v, y \in \mathcal{S}^{d-1}$ such that $v'\nabla^2 H^\rho_{\alpha,u}(ty)v < 0$, which of course implies that $x \mapsto H^\rho_{\alpha,u}(x)$ is not convex over $\mathbb{R}^d$. Therefore, $\alpha_\rho \leq \tilde{\alpha}_\rho$, which establishes the result. ■

Corollary 3.9.9. *Let $\rho \in \mathcal{C}$ be such that $t \mapsto \rho(t)/t$ is convex on $(0,\infty)$. Then $\alpha_\rho \geq \sqrt{2/3} \approx .8165$.*

Proof of Corollary 3.9.9. Recall from (3.9.18) that, for any $t \in \mathcal{D}_\rho^{\text{cv}}$, we have $2p_t - q_t - 2 > 0$, that is, $p_t > (q_t+2)/2$. Since, for any $q > 0$, the map

$$p \mapsto \sqrt{\frac{q(4p^2 - 4p - q)}{4(p-1)^2(q+1)}}$$

is monotone non-increasing in $((q+2)/2, \infty)$, Theorem 3.3.3 yields

$$\alpha_\rho \geq \inf_{t \in \mathcal{D}_\rho^{\mathrm{cv}}} \lim_{p \to \infty} \sqrt{\frac{q_t(4p^2 - 4p - q_t)}{4(p-1)^2(q_t+1)}} = \inf_{t \in \mathcal{D}_\rho^{\mathrm{cv}}} \sqrt{\frac{q_t}{q_t+1}}.$$

The result then follows from the fact that the convexity of $t \mapsto v(t) = \rho(t)/t$ entails that $q_t = 2 + (tv''(t)/v'(t)) \geq 2$ for any $t \in \mathcal{D}_\rho^{\mathrm{cv}}$, where v' and v'' stand for the first and second derivatives of v, respectively (these are well-defined on $\mathcal{D}_\rho$). ■

The proof of Theorem 3.3.4 requires the following lemma.

Lemma 3.9.10. *Let $\rho \in \mathcal{C}$. Fix $x \in \mathbb{R}^d \setminus \{0\}$ and $y \in \mathbb{R}^d$ such that $\rho(\|x\|) + \rho(\|y\|) - 2\rho(\|(x+y)/2\|) = 0$. Then, there exists $\lambda \geq 0$ such that $y = \lambda x$.*

Proof of Lemma 3.9.10. Since ρ is monotone strictly increasing on $[0, \infty)$ (Lemma 3.9.4) and convex, we have $2\rho(\|x+y\|/2) \leq 2\rho(\|x\|/2 + \|y\|/2) \leq \rho(\|x\|) + \rho(\|y\|)$. The assumption on x and y entails that these inequalities must be equalities. Since ρ is monotone strictly increasing, we must then have $\|x+y\| = \|x\| + \|y\|$, so that $y = \lambda x$ for some $\lambda \geq 0$. ■

Proof of Theorem 3.3.4. (i) Fix $\alpha \in [0, \alpha_\rho) \cup \{0\}$. Since the map $\mu \mapsto M^\rho_{\alpha,u}(\mu)$ is continuous (Lemma 3.9.5), it is enough to show that it is midpoint strictly convex. Assume ad absurdum that there exist $\mu_1, \mu_2 \in \mathbb{R}^d$, with $\mu_1 \neq \mu_2$, such that $M^\rho_{\alpha,u}(\mu_1) + M^\rho_{\alpha,u}(\mu_2) - 2M^\rho_{\alpha,u}((\mu_1+\mu_2)/2) \leq 0$. Since convexity of $H^\rho_{\alpha,u}$ (which holds since $\alpha \leq \alpha_\rho$) trivially implies convexity of $M^\rho_{\alpha,u}$, we must then have $M^\rho_{\alpha,u}(\mu_1) + M^\rho_{\alpha,u}(\mu_2) - 2M^\rho_{\alpha,u}((\mu_1+\mu_2)/2) = 0$, that is

$$\int_{\mathbb{R}^d} \left\{ H^\rho_{\alpha,u}(z-\mu_1) + H^\rho_{\alpha,u}(z-\mu_2) - 2H^\rho_{\alpha,u}(z-(\mu_1+\mu_2)/2) \right\} dP(z) = 0.$$

The convexity of $H^\rho_{\alpha,u}$ implies that

$$H^\rho_{\alpha,u}(z-\mu_1) + H^\rho_{\alpha,u}(z-\mu_2) - 2H^\rho_{\alpha,u}(z-(\mu_1+\mu_2)/2) = 0 \tag{3.9.19}$$

for P-almost all z. Now, fix z satisfying (3.9.19). Using the notation introduced in the proof of Lemma 3.3.1, we then have $a(z-\mu_1, z-\mu_2) + \alpha u' V(z-\mu_1, z-\mu_2) = 0$. If $\alpha = 0$, then we have $a(z-\mu_1, z-\mu_2) = 0$. If $\alpha \in (0, \alpha_\rho)$, then

$$\begin{aligned} 0 &= a(z-\mu_1, z-\mu_2) + \alpha u' V(z-\mu_1, z-\mu_2) \\ &\geq a(z-\mu_1, z-\mu_2) - \alpha \|V(z-\mu_1, z-\mu_2)\| \\ &\geq a(z-\mu_1, z-\mu_2) - \alpha_\rho \|V(z-\mu_1, z-\mu_2)\| \geq 0, \end{aligned}$$

since $\rho \in \mathcal{C}_{\alpha_\rho} = V_{\alpha_\rho}$ by definition. These inequalities must therefore be equalities, so that $V(z-\mu_1, z-\mu_2) = 0$ (since $\alpha < \alpha_\rho$), which in turn implies that $a(z-\mu_1, z-\mu_2) = 0$. For any $\alpha \in [0, \alpha_\rho) \cup \{0\}$, we thus obtained that $a(z-\mu_1, z-\mu_2) = 0$, that is,

$$\rho(\|z-\mu_1\|) + \rho(\|z-\mu_2\|) - 2\rho(\|z-(\mu_1+\mu_2)/2\|). \tag{3.9.20}$$

Since $\mu_1 \neq \mu_2$, we cannot have that both $z - \mu_1$ and $z - \mu_2$ are equal to the zero vector in $\mathbb{R}^d$. Without any loss of generality, assume that $z - \mu_1 \neq 0$. Lemma 3.9.10 then implies that $z - \mu_2 = \lambda(z - \mu_1)$ for some $\lambda \in [0, \infty) \setminus \{1\}$ (since $\mu_1 \neq \mu_2$, we cannot have $\lambda = 1$), so that, in particular, z belongs to the line containing μ_1 and μ_2.

For (3.9.19) to be satisfied for P-almost all z, we must thus have that P is concentrated on the line containing μ_1 and μ_2. Now, note that, with $f(t) = \rho(t\|z - \mu_1\|)$, it follows from (3.9.20) that

$$f(1) + f(\lambda) - 2f((1 + \lambda)/2) = 0.$$

Since f is convex on $[0, \infty)$, it follows that f is an affine function on the open interval with endpoints λ and 1, which in turn implies that ρ is an affine function on the open interval with endpoints $\lambda\|z - \mu_1\|$ and $\|z - \mu_1\|$. Consequently, there exists an open interval on which ψ_- is constant. Summing up, we showed that P is concentrated on a line and that there exists an open interval on which ψ_- is constant. Since this is a contradiction, we conclude that $M^\rho_{\alpha,u}$ is midpoint strictly convex, hence strictly convex.

Now, for $\alpha \in [0, \alpha_\rho) \cup \{0\}$, it follows from Theorem 3.2.1 that at least one ρ-quantile $\mu^\rho_{\alpha,u}$—that is, a global minimizer of $\mu \mapsto M^\rho_{\alpha,u}(\mu)$—exists. Strict convexity of $M^\rho_{\alpha,u}$ of course implies that this minimizer is unique. ■

Proofs for Section 3.4

The proof of Theorem 3.4.1 requires the following lemma.

Lemma 3.9.11. *Let $P \in \mathcal{P}^\rho_d$ be spherically symmetric about the origin of $\mathbb{R}^d$ and let $\mu \in \mathbb{R}^d \setminus \{0\}$. Then, $P[\{z \in \mathbb{R}^d \setminus \{0\} : \|z - \mu\| \notin \mathcal{D}_\rho\}] = 0$.*

Proof of Lemma 3.9.11. First note that $\{z \in \mathbb{R}^d \setminus \{0\} : \|z - \mu\| \notin \mathcal{D}_\rho\} = \{\mu\} \cup B$, where

$$B := \bigcup_{t \in (0,\infty) \setminus \mathcal{D}_\rho} B_t$$

involves $B_t := \{z \in \mathbb{R}^d \setminus \{0\} : \|z - \mu\| = t\}$. The sphericity assumption implies that $P[\{\mu\}] = 0$. Since, by assumption, $(0, \infty) \setminus \mathcal{D}_\rho$ is at most countable, it is then sufficient to prove that $P[B_t] = 0$ for any $t \in (0, \infty) \setminus \mathcal{D}_\rho$.

To do so, fix $t \in (0, \infty) \setminus \mathcal{D}_\rho$. Pick arbitrarily $v \in \mathcal{S}^{d-1}$ with $v'\mu = 0$, and partition B_t into $B_t = B_{t,\geq} \cup B_{t,<}$, with $B_{t,\geq} := \{z \in B_t : v'(z - \mu) \geq 0\}$ and $B_{t,<} := \{z \in B_t : v'(z - \mu) < 0\}$. Sphericity implies that $P[B_{t,\geq}] \geq P[B_{t,<}]$, so that $P[B_t] = P[B_{t,\geq}] + P[B_{t,<}] \leq 2P[B_{t,\geq}]$. Let then O_k, $k = 1, 2, \ldots$, be pairwise different $d \times d$ orthogonal matrices fixing the orthogonal complement to $\{\lambda\mu + \eta v : \lambda, \eta \in \mathbb{R}\}$. By construction, the sets $O_k B_{t,\geq} = \{O_k z : z \in B_{t,\geq}\}$, $k = 1, 2, \ldots$, are pairwise disjoint. Since sphericity implies that $P[O_k B_{t,\geq}] = P[B_{t,\geq}]$ for any k, we must then have $P[B_{t,\geq}] = 0$. It follows that $P[B_t] \leq 2P[B_{t,\geq}] = 0$, which establishes the result. ■

Proof of Theorem 3.4.1. (i) Fix $\alpha = 0$ and $u \in \mathcal{S}^{d-1}$, and assume ad absurdum that $\mu \neq 0$ is a minimizer of $M^\rho_{\alpha,u}$. Then, the sphericity assumption implies that $-\mu$

is a minimizer, too (Proposition 3.2.2). It thus follows from Theorem 3.3.4 that P is concentrated on a line. From sphericity, we must then have P is a Dirac at the origin of $\mathbb{R}^d$, which provides $M^{\rho}_{\alpha,u}(\mu) = \rho(\|\mu\|)$ for any $\mu \in \mathbb{R}^d$. Since $\rho(0) = 0$ and $\rho(t) > 0$ for any $t > 0$, we conclude that the only minimizer of $M^{\rho}_{\alpha,u}$ is the origin of $\mathbb{R}^d$, a contradiction.

(ii) Fix $\alpha > 0$ and $u \in \mathcal{S}^{d-1}$. We consider two cases.

(A) Fix an arbitrary μ that does not belong to the line $\{\lambda u : \lambda \in \mathbb{R}\}$. Then $t := \|\mu\| > 0$ and $w := \mu/\|\mu\| \in \mathcal{S}^{d-1} \setminus \{u\}$. Since $\mu = tw \neq 0$, the sphericity assumption implies that $P[\{\mu\}] = 0$. Letting $v = (I - ww')u = u - (u'w)w$, Proposition 3.5.1 then yields

$$\begin{aligned}\frac{\partial M^{\rho}_{\alpha,u}}{\partial v}(\mu) &= -\alpha v'\mathrm{E}\Big[\frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\xi_{Z,\mu}\Big]u + \alpha v'\mathrm{E}\Big[\frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\frac{Z_\mu Z_\mu'}{\|Z_\mu\|^2}\xi_{Z,\mu}\Big]u \\ &\quad -v'\mathrm{E}\Big[\{\psi_-(\|Z_\mu\|)\mathbb{I}[v'Z_\mu > 0] + \psi_+(\|Z_\mu\|)\mathbb{I}[v'Z_\mu < 0]\}\Big(1 + \alpha\frac{u'Z_\mu}{\|Z_\mu\|}\Big)\frac{Z_\mu}{\|Z_\mu\|}\Big] \\ &=: -\alpha\mathrm{E}\Big[\frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\xi_{Z,\mu}\Big](1 - (u'w)^2) + T_1 + T_2, \qquad (3.9.21)\end{aligned}$$

say. Since $v'\mu = 0$, we have

$$\begin{aligned}T_1 &= \alpha\mathrm{E}\Big[\frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\frac{(v'Z)Z_\mu'(v + (u'w)w)}{\|Z_\mu\|^2}\xi_{Z,\mu}\Big] \\ &= \alpha\mathrm{E}\Big[\frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\frac{(v'Z)^2}{\|Z_\mu\|^2}\xi_{Z,\mu}\Big] + \alpha(u'w)\mathrm{E}\Big[\frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\frac{(v'Z)(w'Z - t)}{\|Z_\mu\|^2}\xi_{Z,\mu}\Big].\end{aligned}$$

Pick a $d \times d$ orthogonal matrix O such that $Ow = w$ and $Ov = -v$ (such a matrix O exists since w and v are orthogonal). Since OZ and Z are equal in distribution and since $\|OZ - \mu\| = \|O(Z - \mu)\| = \|Z_\mu\|$ almost surely, we have

$$\mathrm{E}\Big[\frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\frac{(v'Z)(w'Z - t)}{\|Z_\mu\|^2}\xi_{Z,\mu}\Big] = -\mathrm{E}\Big[\frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\frac{(v'Z)(w'Z - t)}{\|Z_\mu\|^2}\xi_{Z,\mu}\Big] = 0,$$

so that

$$T_1 = \alpha\mathrm{E}\Big[\frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\frac{(v'Z)^2}{\|Z_\mu\|^2}\xi_{Z,\mu}\Big]. \qquad (3.9.22)$$

Now, turning to T_2, note that multiplying the random variable in the expectation by the indicator $\mathbb{I}[Z \in A]$, with $A = \{z \in \mathbb{R}^d : \|z - \mu\| \in \mathcal{D}_\rho\}$ will not change the value of T_2. Indeed, this only discards, in the corresponding integral in z, (a) the value $z = 0$, that makes the integrand equal to zero (recall that $v'\mu = 0$) and (b) the non-zero values of z such that $\|z - \mu\| \notin \mathcal{D}_\rho$, that form a set with P-probability zero (Lemma 3.9.11).

Consequently,

$$
\begin{aligned}
T_2 &= -v'\mathrm{E}\Big[\psi_-(\|Z_\mu\|)\Big(1+\alpha\frac{u'Z_\mu}{\|Z_\mu\|}\Big)\frac{Z_\mu}{\|Z_\mu\|}\xi_{Z,\mu}\Big] \\
&= -\mathrm{E}\Big[\psi_-(\|Z_\mu\|)\frac{v'Z}{\|Z_\mu\|}\xi_{Z,\mu}\Big] - \alpha\mathrm{E}\Big[\psi_-(\|Z_\mu\|)\frac{(v'Z)(u'Z_\mu)}{\|Z_\mu\|^2}\xi_{Z,\mu}\Big] \\
&= -\mathrm{E}\Big[\psi_-(\|Z_\mu\|)\frac{v'Z}{\|Z_\mu\|}\xi_{Z,\mu}\Big] - \alpha\mathrm{E}\Big[\psi_-(\|Z_\mu\|)\frac{(v'Z)^2}{\|Z_\mu\|^2}\xi_{Z,\mu}\Big] \\
&\qquad -\alpha(u'w)\mathrm{E}\Big[\psi_-(\|Z_\mu\|)\frac{(v'Z)(w'Z-t)}{\|Z_\mu\|^2}\xi_{Z,\mu}\Big].
\end{aligned}
$$

Using again the fact that OZ and Z are equal in distribution and that $\|OZ-\mu\| = \|Z_\mu\|$ almost surely, this provides

$$
T_2 = -\alpha\mathrm{E}\Big[\psi_-(\|Z_\mu\|)\frac{(v'Z)^2}{\|Z_\mu\|^2}\xi_{Z,\mu}\Big].
$$

Jointly with (3.9.22), this shows that

$$
T_1 + T_2 = -\alpha\mathrm{E}\Big[\Big(\psi_-(\|Z_\mu\|) - \frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\Big)\frac{(v'Z)^2}{\|Z_\mu\|^2}\xi_{Z,\mu}\Big] \leq 0,
$$

since $\psi_-(t) \geq \rho(t)/t$ for any $t > 0$. Therefore, (3.9.21) provides

$$
\frac{\partial M^\rho_{\alpha,u}}{\partial v}(\mu) \leq -\alpha\mathrm{E}\Big[\frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\xi_{Z,\mu}\Big](1-(u'w)^2) < 0,
$$

which shows that μ is not a minimizer of $M^\rho_{\alpha,u}$.

(B) Fix $\mu = -tu$, with $t > 0$. Since we still have $P[\{\mu\}] = 0$, Proposition 3.5.1 provides

$$
\begin{aligned}
\frac{\partial M^\rho_{\alpha,u}}{\partial u}(\mu) &= -\alpha\mathrm{E}\Big[\frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\xi_{Z,\mu}\Big] + \alpha\mathrm{E}\Big[\frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\frac{(u'Z_\mu)^2}{\|Z_\mu\|^2}\xi_{Z,\mu}\Big] \\
&\quad -\mathrm{E}\Big[\{\psi_-(\|Z_\mu\|)\mathbb{I}[u'Z_\mu > 0] + \psi_+(\|Z_\mu\|)\mathbb{I}[u'Z_\mu < 0]\}\Big(1+\alpha\frac{u'Z_\mu}{\|Z_\mu\|}\Big)\frac{u'Z_\mu}{\|Z_\mu\|}\Big].
\end{aligned}
$$

From Lemma 3.9.11, we obtain

$$
\begin{aligned}
\frac{\partial M^\rho_{\alpha,u}}{\partial u}(\mu) &= -\alpha\mathrm{E}\Big[\frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\xi_{Z,\mu}\Big] + \alpha\mathrm{E}\Big[\frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\frac{(u'Z_\mu)^2}{\|Z_\mu\|^2}\xi_{Z,\mu}\Big]u \\
&\quad -\mathrm{E}\Big[\psi_-(\|Z_\mu\|)\Big(1+\alpha\frac{u'Z_\mu}{\|Z_\mu\|}\Big)\frac{u'Z_\mu}{\|Z_\mu\|}\xi_{Z,\mu}\mathbb{I}[Z\neq 0]\mathbb{I}[\|Z-\mu\|\in\mathcal{D}_\rho]\Big] - \psi_-(t)(1+\alpha)P[\{0\}],
\end{aligned}
$$

which rewrites

$$
\begin{aligned}
\frac{\partial M^\rho_{\alpha,u}}{\partial u}(\mu) &= -\alpha\mathrm{E}\Big[\frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\xi_{Z,\mu}\Big] - \mathrm{E}\Big[\psi_-(\|Z_\mu\|)\frac{u'Z_\mu}{\|Z_\mu\|}\xi_{Z,\mu}\Big] \\
&\qquad -\alpha\mathrm{E}\Big[\Big(\psi_-(\|Z_\mu\|) - \frac{\rho(\|Z_\mu\|)}{\|Z_\mu\|}\Big)\frac{(u'Z_\mu)^2}{\|Z_\mu\|^2}\xi_{Z,\mu}\Big].
\end{aligned}
$$

Since $\psi(t) \geq \rho(t)/t > 0$ for any $t > 0$, this yields

$$\frac{\partial M^{\rho}_{\alpha,u}}{\partial u}(\mu) < -\mathrm{E}\bigg[\psi_-(\|Z_\mu\|)\frac{u'Z_\mu}{\|Z_\mu\|}\xi_{Z,\mu}\bigg] = -\mathrm{E}\bigg[\psi_-(\|Z_\mu\|)\frac{u'Z+t}{\|Z_\mu\|}\xi_{Z,\mu}\bigg]. \qquad (3.9.23)$$

Now, let Γ_u be an arbitrary $d \times (d-1)$ matrix whose columns form an orthonormal basis of the orthogonal complement to u in $\mathbb{R}^d$, and define the random $(d-1)$-vector Y through $Z = (u'Z)u + \Gamma_u Y$. Note that $\|Z_\mu\|^2 = \|Z + tu\|^2 = \|(u'Z + t)u + \Gamma_u Y\|^2 = (u'Z+t)^2+\|Y\|^2$ and that, for any $r \geq 0$, the distribution of $u'Z$, conditional on $\|Y\| = r$, is symmetric about zero. Therefore, the monotonicity of ψ_- implies that

$$\mathrm{E}\bigg[\frac{\psi_-(((u'Z+t)^2+\|Y\|^2)^{1/2})(u'Z+t)}{((u'Z+t)^2+\|Y\|^2)^{1/2}}\xi_{Z,\mu}\mathbb{I}[|u'Z|>t]\,\bigg|\,\|Y\|=r\bigg] \geq 0$$

for any $r \geq 0$. It follows that

$$\mathrm{E}\bigg[\psi_-(\|Z_\mu\|)\frac{u'Z+t}{\|Z_\mu\|}\xi_{Z,\mu}\bigg] \geq \mathrm{E}\bigg[\psi_-(\|Z_\mu\|)\frac{u'Z+t}{\|Z_\mu\|}\xi_{Z,\mu}\mathbb{I}[|u'Z|>t]\bigg]$$

$$= \mathrm{E}\bigg[\mathrm{E}\bigg[\frac{\psi_-(((u'Z+t)^2+\|Y\|^2)^{1/2})(u'Z+t)}{((u'Z+t)^2+\|Y\|^2)^{1/2}}\xi_{Z,\mu}\mathbb{I}[|u'Z|>t]\,\bigg|\,\|Y\|\bigg]\bigg] \geq 0.$$

We conclude that the partial derivative in (3.9.23) is strictly negative, hence that μ is not a minimizer of $M^{\rho}_{\alpha,u}$. Together with the result proved in case (A), this establishes that any minimizer of $M^{\rho}_{\alpha,u}$ belongs to the halfline $\{\lambda u : \lambda \geq 0\}$. ■

Proposition 3.9.12. *Let $\rho \in \mathcal{C}$ and $P \in \mathcal{P}^{\rho}_d$. Fix $u \in \mathcal{S}^{d-1}$. Assume that P is spherically symmetric about the origin of $\mathbb{R}^d$. Then, (i) for $\rho(t) = t$, the origin of $\mathbb{R}^d$ is the ρ-quantile of P of order α in direction u if and only if $\alpha \leq P[\{0\}]$ (in which case this quantile is unique); (ii) if $\psi_+(0)P[\{0\}] + P[\|Z\| \in (0,\infty) \setminus \mathcal{D}_\rho] = 0$, then, for $\alpha \in (0,1)$, any ρ-quantile $\mu^{\rho}_{\alpha,u}$ belongs to the halfline $\{\lambda u : \lambda > 0\}$.*

Proof of Proposition 3.9.12. (i) First note that, for $\rho(t) = t$, it readily follows from Proposition 3.5.1 that

$$\frac{\partial M^{\rho}_{\alpha,u}}{\partial u}(0) = (1-\alpha)P[\{0\}] - \alpha\mathrm{E}[\xi_{Z,0}] - \mathrm{E}\bigg[\frac{u'Z}{\|Z\|}\xi_{Z,0}\bigg] = P[\{0\}] - \alpha. \qquad (3.9.24)$$

We then consider three cases. (a) For $\alpha = 0$, the origin of $\mathbb{R}^d$ is the only ρ-quantile of order α in direction u (Theorem 3.4.1(i)). (b) For $\alpha \in (0, P[\{0\}]]$, any ρ-quantile of order α in direction u belongs to $\{\lambda u : \lambda \geq 0\}$ (Theorem 3.4.1(ii)), and the directional derivative in (3.9.24) is non-negative. Convexity of $\mu \mapsto M^{\rho}_{\alpha,u}$ implies that $(\partial M^{\rho}_{\alpha,u}/\partial u)(tu)$ is a monotone non-increasing function of t, that will thus take non-negative values for any $t > 0$. This implies that the origin of $\mathbb{R}^d$ is a ρ-quantile of order α in direction u. Ad absurdum, assume then that this ρ-quantile is not unique. Then Theorem 3.3.4 implies that P is concentrated on a line, hence that $P[\{0\}] = 1$. It results that $M^{\rho}_{\alpha,u} = \|\mu\| - au'\mu$, which is minimized at $\mu = 0$ only, a contradiction. (c) For $\alpha \in (P[\{0\}], 1)$, the directional derivative in (3.9.24) is strictly negative, so that the origin of $\mathbb{R}^d$ is not a ρ-quantile of order α in direction u.

(ii) Since $\psi_+(0)P[\{0\}] + P[\|Z\| \in (0,\infty) \setminus \mathcal{D}_\rho] = 0$, Proposition 3.5.1 provides

$$\begin{aligned}\frac{\partial M^\rho_{\alpha,u}}{\partial u}(0) &= -\alpha \mathrm{E}\Big[\frac{\rho(\|Z\|)}{\|Z\|}\xi_{Z,0}\Big] + \alpha \mathrm{E}\Big[\frac{\rho(\|Z\|)}{\|Z\|}\frac{(u'Z)^2}{\|Z\|^2}\xi_{Z,0}\Big] \\ &\quad -\mathrm{E}\Big[\psi_-(\|Z\|)\Big(1+\alpha\frac{u'Z}{\|Z\|}\Big)\frac{u'Z}{\|Z\|}\xi_{Z,0}\Big] \\ &= -\alpha \mathrm{E}\Big[\frac{\rho(\|Z\|)}{\|Z\|}\xi_{Z,0}\Big] - \alpha \mathrm{E}\Big[\Big(\psi_-(\|Z\|) - \frac{\rho(\|Z\|)}{\|Z\|}\Big)\frac{(u'Z)^2}{\|Z\|^2}\xi_{Z,0}\Big] < 0,\end{aligned}$$

which shows that $\mu = 0$ cannot be a minimizer of $M^\rho_{\alpha,u}$. The result thus follows from Theorem 3.4.1(ii). ■

Proof of Theorem 3.4.2. By definition, $\mathcal{C}^{\mathrm{sph}}_0 \subseteq \mathcal{C}$, whereas Lemma 3.3.1 provides $\mathcal{C} = \mathcal{C}_0 \subseteq \mathcal{C}^{\mathrm{sph}}_0$. Therefore, $\mathcal{C}^{\mathrm{sph}}_0 = \mathcal{C}$. For any $\alpha \in [0,1]$, let

$$V^{\mathrm{sph}}_\alpha := \big\{\rho \in \mathcal{C} : g^\rho_{\alpha,u}(z - su, z - tu) \geq 0 \ \forall z \in \mathbb{R}^d \ \forall s,t \in \mathbb{R} \ \forall u \in \mathcal{S}^{d-1}\big\},$$

where $g^\rho_{\alpha,u}(x,y) := a(x,y) - \alpha|u'V(x,y)|$ is based on the same quantities $a(x,y)$ and $V(x,y)$ as in the proof of Lemma 3.3.1. Note that

$$\begin{aligned}&H^\rho_{\alpha,u}(z - su) + H^\rho_{\alpha,u}(z - tu) - 2H^\rho_{\alpha,u}(z - (s+t)u/2) \\ &\qquad = a(z - su, z - tu) + \alpha u'V(z - su, z - tu). \qquad (3.9.25)\end{aligned}$$

Obviously, if $\rho \in V^{\mathrm{sph}}_\alpha$, then (3.9.25) implies that $t \mapsto H^\rho_{\alpha,u}(z - tu)$ is midpoint convex for any $u \in \mathcal{S}^{d-1}$, that is, $\rho \in \mathcal{C}^{\mathrm{sph}}_\alpha$. Conversely, if $\rho \in \mathcal{C}^{\mathrm{sph}}_\alpha$, then, writing $c_0 = -\mathrm{Sign}(u'V(z - su, z - tu))c$ for any quantity c, we have

$$\begin{aligned}&a(z - su, z - tu) - \alpha|u'V(z - su, z - tu)| \\ &\qquad = a(z - s_0u_0, z - t_0u_0) + \alpha u_0'V(z - s_0u_0, z - t_0u_0) \geq 0,\end{aligned}$$

so that $\rho \in V^{\mathrm{sph}}_\alpha$. Therefore, $\mathcal{C}^{\mathrm{sph}}_\alpha = V^{\mathrm{sph}}_\alpha$, which implies that $\mathcal{C}^{\mathrm{sph}}_{\alpha_2} = V^{\mathrm{sph}}_{\alpha_2} \subseteq V^{\mathrm{sph}}_{\alpha_1} = \mathcal{C}^{\mathrm{sph}}_{\alpha_1}$ for any $\alpha_1 < \alpha_2$.

Note that if $\alpha^{\mathrm{sph}}_\rho = 0$, then we need to have $\alpha = 0$ and the result follows from Theorem 3.3.4. We may thus assume that $\alpha^{\mathrm{sph}}_\rho > 0$. Fix then $\alpha \in [0, \alpha^{\mathrm{sph}}_\rho)$ and $u \in \mathcal{S}^{d-1}$. We proceed as in the proof of Theorem 3.3.4. From continuity of $t \mapsto M^\rho_{\alpha,u}(tu)$ (Lemma 3.9.5), it is sufficient to prove that this map is midpoint strictly convex. Assume ad absurdum that there exist $s,t > 0$, with $s \neq t$, such that $M^\rho_{\alpha,u}(su) + M^\rho_{\alpha,u}(tu) - 2M^\rho_{\alpha,u}((s+t)u/2) \leq 0$, that is,

$$\int_{\mathbb{R}^d} \big\{H^\rho_{\alpha,u}(z - su) + H^\rho_{\alpha,u}(z - tu) - 2H^\rho_{\alpha,u}(z - (s+t)u/2)\big\}\, dP(z) \leq 0.$$

Since $\alpha \leq \alpha^{\mathrm{sph}}_\rho$, the integrand is non-negative for any z, so that the integral must be equal to zero, which (using again the fact that the integrand is non-negative for any z) entails that

$$H^\rho_{\alpha,u}(z - su) + H^\rho_{\alpha,u}(z - tu) - 2H^\rho_{\alpha,u}(z - (s+t)u/2) = 0 \qquad (3.9.26)$$

for P-almost all z. Any z satisfying (3.9.26) satisfies

$$\begin{aligned} 0 &= a(z-su, z-tu) + \alpha u'V(z-su, z-tu) \\ &\geq a(z-su, z-tu) - \alpha|u'V(z-su, z-tu)| \\ &\geq a(z-su, z-tu) - \alpha_\rho^{\mathrm{sph}}|u'V(z-su, z-tu)| \geq 0, \end{aligned}$$

since $\rho \in \mathcal{C}_{\alpha_\rho^{\mathrm{sph}}} = V_{\alpha_\rho^{\mathrm{sph}}}$ by definition. Since $\alpha < \alpha_\rho^{\mathrm{sph}}$, we must have $u'V(z-su, z-tu) = 0$, hence also $a(z-su, z-tu) = 0$, for P-almost all z. Thus,

$$\rho(\|z-su\|) + \rho(\|z-tu\|) - 2\rho(\|z-(s+t)u/2\|)$$

for P-almost all z, which, in view of Lemma 3.9.10, entails that P is concentrated on the line $\{\lambda u : \lambda \in \mathbb{R}\}$. The sphericity assumption then implies that $P[\{0\}] = 1$, which yields

$$M_{\alpha,u}^{\rho}(tu) = \rho(|t|)\left(1 - \alpha\frac{t}{|t|}\right)\xi_{t,0} = \rho(|t|)\{(1+\alpha)\mathbb{I}[t<0] + (1-\alpha)\mathbb{I}[t \geq 0]\}.$$

Thus, $t \mapsto M_{\alpha,u}^{\rho}(tu)$ is strictly convex on $\mathbb{R}$, so that $M_{\alpha,u}^{\rho}(su) + M_{\alpha,u}^{\rho}(tu) - 2M_{\alpha,u}^{\rho}((s+t)u/2) > 0$, a contradiction. We thus proved that $t \mapsto M_{\alpha,u}^{\rho}(tu)$ is always strictly convex on $(0,\infty)$ under sphericity, which, in view of Theorem 3.4.1, implies uniqueness of $\mu_{\alpha,u}^{\rho}$. ■

Proof of Theorem 3.4.3. We prove the result only in the case where ρ is twice continuously differentiable on $(0,\infty)$ (extension to the general case where ρ is piecewise twice continuously differentiable on $(0,\infty)$ can indeed be done as in the proof of Theorem 3.3.3). Letting $\tilde{\alpha}_{t,\rho}^{\mathrm{sph}} := \sqrt{\beta_{p_t,q_t}}$ for any $t \in \mathcal{D}_\rho^{\mathrm{sph}}$, we need to prove that

$$\alpha_\rho^{\mathrm{sph}} = \tilde{\alpha}_\rho^{\mathrm{sph}} := \begin{cases} \inf_{t \in \mathcal{D}_\rho^{\mathrm{sph}}} \tilde{\alpha}_{t,\rho}^{\mathrm{sph}} & \text{if } \mathcal{D}_\rho^{\mathrm{sph}} \neq \emptyset \\ 1 & \text{otherwise} \end{cases} \tag{3.9.27}$$

and that $\tilde{\alpha}_\rho^{\mathrm{sph}} < 1$ if $\mathcal{D}_\rho^{\mathrm{sph}} \neq \emptyset$ (we will actually show below that $\tilde{\alpha}_{t,\rho}^{\mathrm{sph}} < 1$ for any $t \in \mathcal{D}_\rho^{\mathrm{sph}}$). To prove that $\alpha_\rho^{\mathrm{sph}} = \tilde{\alpha}_\rho^{\mathrm{sph}}$, we need to show that (i) for any $\alpha \leq \tilde{\alpha}_\rho^{\mathrm{sph}}$,

$$\left(\frac{d^2}{dr^2} H_{\alpha,u}^{\rho}(z-ru)\Big|_{r=s} = \right) u'\nabla^2 H_{\alpha,u}^{\rho}(z-su)u \geq 0$$

for any $u \in \mathcal{S}^{d-1}$, $r > 0$ and $z \in \mathbb{R}^d$ such that $z - su \neq 0$ and that (ii) if $\tilde{\alpha}_\rho^{\mathrm{sph}} < 1$, then, for any $\alpha > \tilde{\alpha}_\rho^{\mathrm{sph}}$, there exist such values of z, s, u providing $u'\nabla^2 H_{\alpha,u}^{\rho}(z-su)u < 0$. Clearly, letting $x = z - su$ and writing $x = ty$, with $t = \|x\|$ and $y = x/\|x\|$, we have to show that (i) $u'\nabla^2 H_{\alpha,u}^{\rho}(ty)u \geq 0$ for any $t > 0$, $u, y \in \mathcal{S}^{d-1}$ and $\alpha \leq \tilde{\alpha}_\rho^{\mathrm{sph}}$ and that (ii) if $\tilde{\alpha}_\rho^{\mathrm{sph}} < 1$, then, for any $\alpha > \tilde{\alpha}_\rho^{\mathrm{sph}}$, there exist such values of t, u, y for which $u'\nabla^2 H_{\alpha,u}^{\rho}(ty)u < 0$.

To prove (i)–(ii), partition $\mathcal{D}_\rho = (0,\infty)$ into $\mathcal{D}_\rho^{\mathrm{cv}} \cup \mathcal{E}_\rho \cup \mathcal{F}_\rho$ as in the proof of Theorem 3.3.3. It follows from (3.9.15) that if $t \in \mathcal{F}_\rho$, then $u'\nabla^2 H_{\alpha,u}^{\rho}(ty)u$ for any $\alpha \in [0,1]$

and $u, y \in \mathcal{S}^{d-1}$. Now, recall from the proof of Theorem 3.3.3 that for $t \in (\mathcal{D}_\rho \setminus \mathcal{F}_\rho) = \mathcal{D}_\rho^{\rm cv} \cup \mathcal{E}_\rho$,

$$\frac{t^2}{\rho(t)} v' \nabla^2 H^\rho_{\alpha,u}(ty) v = g_{t,\alpha}(\theta, \omega),$$

with

$$g_{t,\alpha}(\theta,\omega) := (p_t - 1)(q_t - 2)(1 + \alpha\cos\theta)(\cos\omega)^2 + 1 - (\cos\omega)^2$$
$$+(p_t - 1)\big\{(1 + \alpha\cos\theta)(1 - (\cos\omega)^2) + 2(\cos\omega)^2 + 2\alpha(\cos\omega)\cos(\theta - \omega)\big\},$$

where $\omega = \arccos(v'y) \in [0, \pi]$ is the angle between v and y and $\theta = s_{u,v,y} \arccos(u'y) \in [-\pi, \pi]$ is the signed angle between u and y. A close inspection of the proof of Theorem 3.3.3 reveals that we established there that if $t \in \mathcal{E}_\rho$, then $g_{t,\alpha}(\theta, \omega) \geq 0$ for any $\alpha \in [0, 1]$ and (θ, ω), which ensures that $u'\nabla^2 H^\rho_{\alpha,u}(ty)u$ for any $t \in \mathcal{E}_\rho$, $\alpha \in [0, 1]$ and $u, y \in \mathcal{S}^{d-1}$.

Therefore, it remains to prove that (i) $u'\nabla^2 H^\rho_{\alpha,u}(ty)u \geq 0$ for any $t \in \mathcal{D}_\rho^{\rm cv}$, $u, y \in \mathcal{S}^{d-1}$ and $\alpha \leq \tilde{\alpha}_\rho^{\rm sph}$, and that (ii) if $\tilde{\alpha}_\rho^{\rm sph} < 1$, then, for any $\alpha > \tilde{\alpha}_\rho^{\rm sph}$, there exist $t \in \mathcal{D}_\rho^{\rm cv}$ and $u, y \in \mathcal{S}^{d-1}$ such that $u'\nabla^2 H^\rho_{\alpha,u}(ty)u < 0$. To do so, put, for any $\omega \in [0, \pi]$,

$$h_{t,\alpha}(\omega) := g_{t,\alpha}(\omega, \omega) = \frac{t^2}{\rho(t)} u'\nabla^2 H^\rho_{\alpha,u}(ty)u.$$

Following the same approach as in the proof of Theorem 3.3.3, write then

$$h_{t,\alpha}(\omega) = i_t(\omega) + \alpha s_t^{\rm sph}(\omega),$$

where the intercept function $i_t(\omega)$ is still given by

$$i_t(\omega) = p_t + (p_t q_t - p_t - q_t)(\cos\omega)^2$$

and where the slope function $s_t^{\rm sph}(\omega)$ is now defined as

$$s_t^{\rm sph}(\omega) := (p_t - 1)(\cos\omega)\{(q_t - 3)(\cos\omega)^2 + 3\}.$$

Since the intercept is the same as in the proof of Theorem 3.3.3, we still have that $i_t(\omega) \geq 0$ and that if $i_t(\omega) = 0$, then $q_t = 0$ and $(\cos\omega)^2 = 1$, so that $h_{t,\alpha}(\omega) = 0$ for any $\alpha \in [0, 1]$ and $\omega \in [0, \pi]$. From now one, we thus safely restrict the case where $i_t(\omega) > 0$ when investigating when $h_{t,\alpha}(\omega)$ is negative. Putting then

$$\alpha_t^{\rm sph}(\omega) := \begin{cases} -i_t(\omega)/s_t^{\rm sph}(\omega) & \text{if } s_t^{\rm sph}(\omega) < 0 \\ \infty & \text{otherwise,} \end{cases}$$

we have that, in the range $\alpha \in [0, \infty)$, $h_{t,\alpha}(\omega) \geq 0$ if and only if $\alpha \leq \alpha_t^{\rm sph}(\omega)$ (if $\alpha_t^{\rm sph}(\omega) = \infty$, then this is to be read as $h_{t,\alpha}(\omega) \geq 0$ for any $\alpha \in [0, \infty)$).

In the beginning of the proof, we defined $\tilde{\alpha}_{t,\rho}^{\rm sph} := \sqrt{\beta_{p_t,q_t}}$ for any $t \in \mathcal{D}_\rho^{\rm sph}$. Let further $\tilde{\alpha}_{t,\rho}^{\rm sph} := 1$ for any $t \in \mathcal{D}_\rho^{\rm cv} \setminus \mathcal{D}_\rho^{\rm sph}$. To establish the theorem, it is then sufficient to prove that

$$\min\big\{\alpha_t^{\rm sph}(\omega) : \omega \in [0, \pi]\big\} = \tilde{\alpha}_{t,\rho}^{\rm sph} \qquad \text{for any } t \in \mathcal{D}_\rho^{\rm cv}, \tag{3.9.28}$$

$$\tilde{\alpha}_{t,\rho}^{\mathrm{sph}} < 1 \text{ for any } t \in \mathcal{D}_\rho^{\mathrm{sph}} \tag{3.9.29}$$

and

$$\mathcal{D}_\rho^{\mathrm{sph}} \subseteq \mathcal{D}_\rho^{\mathrm{cv}}. \tag{3.9.30}$$

Indeed, for any $\alpha \leq \tilde{\alpha}_\rho^{\mathrm{sph}}$, $t \in \mathcal{D}_\rho^{\mathrm{cv}}$ and $\omega \in [0, \pi]$, we then have $\alpha \leq \tilde{\alpha}_{t,\rho}^{\mathrm{sph}} \leq \alpha_t^{\mathrm{sph}}(\omega)$, hence $h_{t,\alpha}(\omega) \geq 0$. This proves (i) (note that if $\mathcal{D}_\rho^{\mathrm{cv}}$ is empty, then Theorem 3.3.3 yields $\alpha_\rho = 1$, which, since $\alpha_\rho^{\mathrm{sph}} \geq \alpha_\rho$, implies that $\alpha_\rho^{\mathrm{sph}} = 1$, as claimed by the theorem since (3.9.30) entails that $\mathcal{D}_\rho^{\mathrm{sph}}$ is then empty, too). Now, assume that $\tilde{\alpha}_\rho^{\mathrm{sph}} < 1$. Then, for any $\alpha > \tilde{\alpha}_\rho^{\mathrm{sph}}$, there exists $t \in \mathcal{D}_\rho^{\mathrm{cv}}$ such that $\alpha > \tilde{\alpha}_{t,\rho}^{\mathrm{sph}}$, which, in view of (3.9.28), ensures that $h_{t,\alpha}(\omega) < 0$ for some $\omega \in [0, \pi]$. This establishes (ii), hence the result.

It thus remains to prove (3.9.28)–(3.9.30). Since $\alpha_t^{\mathrm{sph}}(\omega)$ depends on ω only through $\cos\omega$ and since nonnegative values of $\cos\omega$ provide $\alpha_t^{\mathrm{sph}}(\omega) = \infty$, one has

$$\min\left\{\alpha_t^{\mathrm{sph}}(\omega) : \omega \in [0, \pi]\right\} = \min\left\{\ell_t(s) : s \in (0, 1]\right\} =: m_t,$$

where we let

$$\ell_t(s) := \frac{p_t + (p_t q_t - p_t - q_t)s}{(p_t - 1)\sqrt{s}((q_t - 3)s + 3)}.$$

Since $\ell_t(s)$ diverges to infinity as $s \to 0$ from above, the minimum of ℓ_t in $s \in (0, 1]$ indeed exists, and it is equal to the minimum between $\ell_t(1) = 1$ and the infimum of ℓ_t over $(0, 1)$. The derivative of ℓ_t at $s \in (0, 1)$ has the same sign as

$$d_t(s) := a_t s^2 + 3(2p_t - q_t)s - 3p_t, \quad \text{with } a_t := (3 - q_t)(p_t q_t - p_t - q_t).$$

Below, we use the term "root" for a value of s such that $d_t(s) = 0$. Obviously, there are at most two roots in $(0, 1)$ and $d_t(0) = -3p_t < 0$. In the rest of the proof, we organize the discussion in two levels, the first one ((A)–(B)) involving the sign of

$$d_t(1) = q_t(3(p_t - 2) - (p_t q_t - p_t - q_t))$$

and the second one ((1)–(3)) associated with the sign of a_t.

(A) Assume that $d_t(1) > 0$. Since $q_t \geq 0$, we then have $3(p_t - 2) - (p_t q_t - p_t - q_t) > 0$, which rewrites $4p_t + q_t - p_t q_t - 6 > 0$, i.e., $t \in \mathcal{D}_\rho^{\mathrm{sph}}$. Note that $3(p_t - 2) - (p_t q_t - p_t - q_t) > 0$ entails that $q_t < (2(2p_t - 3))/(p_t - 1) \leq 2(p_t - 1)$, so that $2p_t - q_t - 2 > 0$. In view of (3.9.18), this proves (3.9.30).

Since $d_t(0) < 0$ and d_t is convex/concave, there is exactly one root in $(0, 1)$, that is the minimizer of ℓ_t over $(0, 1]$ (since $d_t(s) < 0$ below this root and $d_t(s) > 0$ above it). If (A1) $a_t > 0$, then $s \mapsto d_t(s)$ is convex, so that the root in $(0, 1)$ is the larger root, namely

$$r_t = \frac{-3(2p_t - q_t) + \sqrt{\Delta_t}}{2a_t}, \tag{3.9.31}$$

where

$$\begin{aligned}
\Delta_t &= 9(2p_t - q_t)^2 + 12p_t a_t \\
&= 12q_t\Big(p_t - \frac{3}{2}\Big)\Big(\frac{3}{2}(2p_t - q_t) - (p_t q_t - p_t - q_t)\Big).
\end{aligned}$$

In this case, the minimum is thus $m_t = \ell_t(r_t)(< \ell_t(1) = 1)$, which, after some computations, is shown to be equal to $\tilde{\alpha}^{\rm sph}_{t,\rho}$ (recall that $t \in \mathcal{D}^{\rm sph}_\rho$). If (A2) $a_t < 0$, then $s \mapsto d_t(s)$ is concave, so that the root in $(0,1)$ is now the smaller root, which is still r_t. The minimum is thus $m_t = \ell_t(r_t) = \tilde{\alpha}^{\rm sph}_{t,\rho}$, too. If (A3) $a_t = 0$, then either $q_t = 3$ or $p_tq_t - p_t - q_t = 0$. If $q_t = 3$, then $p_t > 3$ (since $d_t(1) > 0$) and the root (in $(0,1)$) is $p_t/(2p_t - 3)$, which is the limit of r_t as $q_t \to 3$. Thus, the minimum in this case is still $m_t = \ell_t(r_t) = \tilde{\alpha}^{\rm sph}_{t,\rho}$ (recall from the statement of the theorem that the value $\tilde{\alpha}^{\rm sph}_{t,\rho} = \sqrt{\beta_{p_t,q_t}}$ is defined as a limit when q_t makes its value undefined). If $p_tq_t - p_t - q_t = 0$, then $p_t > 2$ (since $d_t(1) > 0$) and the root in $(0,1)$ is $(p_t - 1)/(2p_t - 3)$, which is the limit of r_t as $q_t \to p_t/(p_t - 1)$. Thus, the minimum in this case is once more $m_t = \ell_t(r_t) = \tilde{\alpha}^{\rm sph}_{t,\rho}$.

(B) Assume that $d_t(1) \leq 0$, so that $t \in \mathcal{D}^{\rm cv}_\rho \setminus \mathcal{D}^{\rm sph}_\rho$. If (B1) $a_t > 0$, then convexity of $s \mapsto d_t(s)$ implies that $d_t(s) \leq 0$ for any $s \in [0,1]$ (recall that $d_t(0) < 0$), so that the minimum is $m_t = \ell_t(1) = 1 = \tilde{\alpha}^{\rm sph}_{t,\rho}$. If (B2) $a_t < 0$, then $s \mapsto d_t(s)$ is concave, and there might have zero, one or two roots in $(0,1]$. If there is zero, one root, or two roots with the smaller root—namely, r_t in (3.9.31)—larger than or equal to one, then $d_t(s) \leq 0$ for any $s \in (0,1]$, so that the minimum is $m_t = \ell_t(1) = 1 = \tilde{\alpha}^{\rm sph}_{t,\rho}$. Assume then that there are two roots and that $m_t < 1$. Both roots are positive (since their sum and products are positive), and they must both belong to $(0,1]$ (since $r_t < 1$ and $d_t(1) \leq 0$). Thus we must have $d_t'(1) < 0$, which yields

$$p_tq_t - p_t - q_t > 3(p_t - \tfrac{3}{2}). \qquad (3.9.32)$$

We need to consider three cases: (a) $p_t > \frac{3}{2}$. Then, $p_tq_t - p_t - q_t > 0$, so that $q_t > 3$ (recall that $a_t < 0$). Thus, (3.9.32) yields $\frac{3}{2}(2p_t - q_t) - (p_tq_t - p_t - q_t) \geq 3(3-q)/2 < 0$, which implies that $\Delta_t \leq 0$, a contradiction since we assumed that there are two roots. (b) $p_t \leq \frac{3}{2}$ and $p_tq_t - p_t - q_t \leq 0$. Since $2p_t - q_t \geq 2$ (recall that $t \in \mathcal{D}^{\rm cv}_\rho$), we then have $\Delta_t \leq 0$, which provides the same contradiction as above. (c) $p_t \leq \frac{3}{2}$ and $p_tq_t - p_t - q_t > 0$. Since $a_t < 0$, we have $q_t > 3 \geq 2p_t$. Therefore, $2p_t - q_t - 2 < 0$, which contradicts the fact that $t \in \mathcal{D}^{\rm cv}_\rho$. Finally, (B3) if $a_t = 0$, then $s \mapsto d_t(s)$ is linear and both $d_t(0)$ and $d_t(1)$ are nonpositive. Thus, $d_t(s) \leq 0$ for any $s \in (0,1]$, so that the minimum is still $m_t = \ell_t(1) = 1 = \tilde{\alpha}^{\rm sph}_{t,\rho}$.

Let us summarize the findings of this discussion. Any $t \in \mathcal{D}^{\rm sph}_\rho$ corresponds to case (A), where the minimum $\min\{\alpha^{\rm sph}_t(\omega) : \omega \in [0,\pi]\} = \tilde{\alpha}^{\rm sph}_{t,\rho}(< 1)$, whereas any $t \in \mathcal{D}^{\rm cv}_\rho \setminus \mathcal{D}^{\rm sph}_\rho$ corresponds to case (B), where the minimum $\min\{\alpha^{\rm sph}_t(\omega) : \omega \in [0,\pi]\} = \tilde{\alpha}^{\rm sph}_{t,\rho} = 1$. This proves both (3.9.28) and (3.9.29). Since (3.9.30) was established when discussing case (A), the result is proved. ■

Proofs for Section 3.5

In this section, for a map $g : \mathbb{R}^k \to \mathbb{R}$ admitting directional derivatives at $x(\in \mathbb{R}^k)$ in any direction $v(\in \mathbb{R}^d \setminus \{0\})$, we write $\nabla g(x)$ for the vector stacking the k partial derivatives $\partial_{x_\ell} g(x)$, $\ell = 1, \ldots, k$, on top of each other, irrespective of whether or not g is differentiable at x.

Lemma 3.9.13. *Let $\rho \in \mathcal{C}$, $\alpha \in [0,1]$ and $u \in \mathcal{S}^{d-1}$. Then, (i) for any $x \in \mathbb{R}^d$ and $v \in \mathbb{R}^d \setminus \{0\}$, $H^\rho_{\alpha,u}$ admits a directional derivative at x in direction v, given by*

$$\frac{\partial H^\rho_{\alpha,u}}{\partial v}(x) = \Big(\psi_-(\|x\|)\mathbb{I}[v'x < 0] + \psi_+(\|x\|)\mathbb{I}[v'x > 0]\Big)\left(1 + \alpha\frac{u'x}{\|x\|}\right)\frac{v'x}{\|x\|}\xi_{x,0}$$

$$+ \alpha\frac{\rho(\|x\|)}{\|x\|}v'\left(u - \frac{(u'x)x}{\|x\|^2}\right)\xi_{x,0} + \psi_+(0)(\|v\| + \alpha u'v)\mathbb{I}[x = 0].$$

(ii) If $x_0 \in \mathbb{R}^d$ is such that $\|x_0\| \in \mathcal{D}_\rho$, then $H^\rho_{\alpha,u}$ is twice continuously differentiable in a neighbourhood $\mathcal{N}$ of x_0, with gradient

$$\nabla H^\rho_{\alpha,u}(x) = \psi_-(\|x\|)\left(1 + \alpha\frac{u'x}{\|x\|}\right)\frac{x}{\|x\|} + \alpha\frac{\rho(\|x\|)}{\|x\|}\left(u - \frac{(u'x)x}{\|x\|^2}\right)$$

and Hessian matrix

$$\nabla^2 H^\rho_{\alpha,u}(x) = \frac{\|x\|^2\psi'_-(\|x\|) - 2\|x\|\psi_-(\|x\|) + 2\rho(\|x\|)}{\|x\|^2}\left(1 + \alpha\frac{u'x}{\|x\|}\right)\frac{xx'}{\|x\|^2}$$

$$+\frac{\rho(\|x\|)}{\|x\|^2}\left(I_d - \frac{xx'}{\|x\|^2}\right)$$

$$+\frac{\|x\|\psi_-(\|x\|) - \rho(\|x\|)}{\|x\|^2}\left\{\left(1 + \alpha\frac{u'x}{\|x\|}\right)\left(I_d - \frac{xx'}{\|x\|^2}\right) + 2\frac{xx'}{\|x\|^2} + \alpha\frac{xu' + ux'}{\|x\|}\right\}$$

for any $x \in \mathcal{N}$.

Proof of Lemma 3.9.13. (i) The proof, which is a direct computation of the directional derivative, is left to the reader. (ii) Fix $x_0 \in \mathbb{R}^d$ such that $\|x_0\| \in \mathcal{D}_\rho$, and write $\varphi(s) = \rho(s)/s$ for $s \in (0,\infty)$. Since ρ (hence, also φ) is twice differentiable over a neighbourhood of $\|x_0\|$, the map $x \mapsto H^\rho_{\alpha,u}(x) = \varphi(\|x\|)(\|x\| + \alpha u'x)$ is twice differentiable over a neighbourhood $\mathcal{N}$ of x_0 and we have

$$\frac{\partial H^\rho_{\alpha,u}}{\partial x_i}(x) = \varphi'(\|x\|)\frac{x_i}{\|x\|}(\|x\| + \alpha u'x) + \varphi(\|x\|)\left(\frac{x_i}{\|x\|} + \alpha u_i\right) \tag{3.9.33}$$

for any $x \in \mathcal{N}$. Leibniz's rule then yields

$$\frac{\partial}{\partial x_j}\left(\frac{\partial H^\rho_{\alpha,u}}{\partial x_i}\right)(x)$$

$$= \left(\varphi''(\|x\|)\frac{x_jx_i}{\|x\|^2} + \frac{\varphi'(\|x\|)}{\|x\|}\delta_{ij} - \varphi'(\|x\|)\frac{x_ix_j}{\|x\|^3}\right)(\|x\| + \alpha u'x)$$

$$+\varphi'(\|x\|)\frac{x_i}{\|x\|}\left(\frac{x_j}{\|x\|} + \alpha u_j\right) + \varphi'(\|x\|)\frac{x_j}{\|x\|}\left(\frac{x_i}{\|x\|} + \alpha u_i\right)$$

$$+\varphi(\|x\|)\left(\delta_{ij}\frac{1}{\|x\|} - \frac{x_ix_j}{\|x\|^3}\right).$$

With $y = x/\|x\|$ and $t = \|x\|$, this rewrites

$$\frac{\partial}{\partial x_j}\left(\frac{\partial H^\rho_{\alpha,u}}{\partial x_i}\right)(x) = t\varphi''(t)(1 + \alpha u'y)y_iy_j + \frac{\varphi(t)}{t}(\delta_{ij} - y_iy_j)$$

$$+\varphi'(t)\big\{\delta_{ij}(1 + \alpha u'y) - y_iy_j(1 + \alpha u'y) + 2y_iy_j + \alpha y_iu_j + \alpha u_iy_j\big\}. \tag{3.9.34}$$

Clearly, (3.9.33) and (3.9.34) yield

$$\begin{aligned}\nabla H^{\rho}_{\alpha,u}(x) &= t\varphi'(t)(1+\alpha u'y)y+\varphi(t)(y+\alpha u)\\ &= (t\varphi'(t)+\varphi(t))(1+\alpha u'y)y+\alpha\varphi(t)(u-(u'y)y)\end{aligned}$$

and

$$\begin{aligned}\nabla^2 H^{\rho}_{\alpha,u}(x) = t\varphi''(t)(1+\alpha u'y)yy' + \frac{\varphi(t)}{t}(I_d-yy')\\ +\varphi'(t)\big\{(1+\alpha u'y)(I_d-yy')+2yy'+\alpha yu'+\alpha uy'\big\},\end{aligned}$$

respectively. Expressing φ, φ' and φ'' in terms of ρ provides the result. ■

Lemma 3.9.14. *Let $\rho\in\mathcal{C}$, $\alpha\in[0,1]$ and $u\in\mathcal{S}^{d-1}$. Then,*

$$\begin{aligned}&\big|H^{\rho}_{\alpha,u}(z-\mu_2)-H^{\rho}_{\alpha,u}(z-\mu_1)\big|\xi_{z,\mu_1}\xi_{z,\mu_2}\\ &\qquad\le(1+3\alpha)\psi_-(\|z-\mu_1\|+\|\mu_2-\mu_1\|)\|\mu_2-\mu_1\|\end{aligned}$$

for any $z,\mu_1,\mu_2\in\mathbb{R}^d$ with $\mu_1\neq\mu_2$.

Proof of Lemma 3.9.14. Since the inequality to be proved trivially follows from Lemma 3.9.4(iii) for $z\in\{\mu_1,\mu_2\}$, we restrict to $z\notin\{\mu_1,\mu_2\}$. Write then

$$\big(H^{\rho}_{\alpha,u}(z-\mu_2)-H^{\rho}_{\alpha,u}(z-\mu_1)\big)\xi_{z,\mu_1}\xi_{z,\mu_2}=f_{\alpha,u}(z,\mu_1,\mu_2)+g_{\alpha,u}(z,\mu_1,\mu_2),$$

with

$$f_{\alpha,u}(z,\mu_1,\mu_2):=\big(\rho(\|z-\mu_2\|)-\rho(\|z-\mu_1\|)\big)\left(1+\alpha u'\frac{z-\mu_2}{\|z-\mu_2\|}\right)$$

and

$$g_{\alpha,u}(z,\mu_1,\mu_2):=\alpha\rho(\|z-\mu_1\|)\left(u'\frac{z-\mu_2}{\|z-\mu_2\|}-u'\frac{z-\mu_1}{\|z-\mu_1\|}\right).$$

Lemma 3.9.3 and the monotonicity of ψ_- ensure that, for some c between $\|z-\mu_1\|$ and $\|z-\mu_2\|$,

$$\begin{aligned}|\rho(\|z-\mu_2\|)-\rho(\|z-\mu_1\|)| &\le\psi_-(c)\big|\|z-\mu_2\|-\|z-\mu_1\|\big|\\ &\le\psi_-(\|z-\mu_1\|+\|\mu_2-\mu_1\|)\|\mu_2-\mu_1\|,\end{aligned}$$

so that

$$|f_{\alpha,u}(z,\mu_1,\mu_2)|\le(1+\alpha)\psi_-(\|z-\mu_1\|+\|\mu_2-\mu_1\|)\|\mu_2-\mu_1\|.$$

Now, using Lemma 4.8.5, Lemma 3.9.4(i), and the monotonicity of ψ_- again, provides

$$\begin{aligned}&|g_{\alpha,u}(z,\mu_1,\mu_2)|\le\alpha\rho(\|z-\mu_1\|)\left\|\frac{z-\mu_2}{\|z-\mu_2\|}-\frac{z-\mu_1}{\|z-\mu_1\|}\right\|\\ &\le 2\alpha\frac{\rho(\|z-\mu_1\|)\|\mu_2-\mu_1\|}{\|z-\mu_1\|}\le(2\alpha\psi_-(\|z-\mu_1\|+\|\mu_2-\mu_1\|)\|\mu_2-\mu_1\|.\end{aligned}$$

The result follows. ■

Proof of Proposition 3.5.1. (i) For any $t > 0$, write

$$\frac{M^{\rho}_{\alpha,u}(\mu+tv) - M^{\rho}_{\alpha,u}(\mu)}{t} = \frac{1}{t}\int_{\mathbb{R}^d} \Big\{ H^{\rho}_{\alpha,u}(z-\mu-tv) - H^{\rho}_{\alpha,u}(z-\mu) \Big\}\, dP(z)$$

$$= \frac{\rho(t\|v\|)}{t}\Big(1-\alpha u'\frac{v}{\|v\|}\Big)P[\{\mu\}] - \frac{\rho(t\|v\|)}{t}\Big(1+\alpha u'\frac{v}{\|v\|}\Big)P[\{\mu+tv\}]$$

$$+ \int_{\mathbb{R}^d} \frac{H^{\rho}_{\alpha,u}(z-\mu-tv) - H^{\rho}_{\alpha,u}(z-\mu)}{t}\xi_{z,\mu+tv}\xi_{z,\mu}\, dP(z).$$

Since Lemma 3.9.4(ii) implies that $\rho(t\|v\|)/t = \|v\|\rho(t\|v\|)/(t\|v\|) \leq \|v\|\rho(\|v\|)/\|v\| = \rho(\|v\|)$ for $t \in (0,1]$, Lemma 3.9.1 yields

$$\lim_{t \overset{>}{\to} 0} \frac{\rho(t\|v\|)}{t}\Big(1+\alpha u'\frac{v}{\|v\|}\Big)P[\{\mu+tv\}] = 0.$$

Now, with $\delta = \delta_\mu$ as in (3.2.4), Lemma 3.9.14 implies that

$$\left|\frac{H^{\rho}_{\alpha,u}(z-\mu-tv) - H^{\rho}_{\alpha,u}(z-\mu)}{t}\right|\xi_{z,\mu}\xi_{z,\mu+tv} \leq \|v\|(1+3\alpha)\psi_-(\|z-\mu\|+\delta_\mu)$$

as soon as $t < \delta_\mu/\|v\|$. Applying Lebesgue's DCT, Lemma 3.9.13 then provides

$$\lim_{t \overset{>}{\to} 0} \frac{M^{\rho}_{\alpha,u}(\mu+tv) - M^{\rho}_{\alpha,u}(\mu)}{t}$$

$$= \psi_+(0)\big(\|v\| - \alpha u'v\big)P[\{\mu\}] + \int_{\mathbb{R}^d} \frac{\partial H^{\rho}_{\alpha,u}}{\partial(-v)}(z-\mu)\xi_{z,\mu}\, dP(z), \qquad (3.9.35)$$

which establishes the result. ■

Proof of Theorem 3.5.2. (i) We start with necessity. Assume thus that the map $M^{\rho}_{\alpha,u}$ is differentiable at μ_0. Then, directional derivatives at μ_0 in direction v are linear in v, which imposes that

$$\frac{\partial M^{\rho}_{\alpha,u}}{\partial v}(\mu_0) + \frac{\partial M^{\rho}_{\alpha,u}}{\partial(-v)}(\mu_0) = 0,$$

that is,

$$\int_{\mathbb{R}^d} \big(\psi_+(\|z-\mu_0\|) - \psi_-(\|z-\mu_0\|)\big)\frac{|v'(z-\mu_0)|}{\|z-\mu_0\|}\Big(1+\alpha\frac{u'(z-\mu_0)}{\|z-\mu_0\|}\Big)\xi_{z,\mu_0} dP(z)$$

$$+2\psi_+(0)P[\{\mu_0\}] = 0$$

for any $v \in \mathcal{S}^{d-1}$. Since convexity of ρ implies that $\psi_+(t) \geq \psi_-(t)$ for any $t > 0$, we must then have that $\psi_+(0)P[\{\mu_0\}] = 0$ and that, for any $v \in \mathcal{S}^{d-1}$,

$$\big(\psi_+(\|z-\mu_0\|) - \psi_-(\|z-\mu_0\|)\big)\frac{|v'(z-\mu_0)|}{\|z-\mu_0\|}\xi_{z,\mu_0} = 0$$

for P-almost every $z \in \mathbb{R}^d$. This implies

$$0 = \int_{\mathbb{R}^d} \big(\psi_+(\|z-\mu_0\|) - \psi_-(\|z-\mu_0\|)\big) \frac{|v'(z-\mu_0)|}{\|z-\mu_0\|} \xi_{z,\mu_0} \, dP(z)$$

$$= \sum_{t \in (0,\infty)\setminus \mathcal{D}_\rho} \frac{1}{t}\big(\psi_+(t) - \psi_-(t)\big) \int_{\mathbb{R}^d} |v'(z-\mu_0)| \mathbb{I}[\|z-\mu_0\| = t] \, dP(z).$$

This implies that for any $t \in (0,\infty) \setminus \mathcal{D}_\rho$ and $v \in \mathcal{S}^{d-1}$, we have $|v'(z-\mu_0)|\mathbb{I}[\|z-\mu_0\| = t] = 0$ for P-almost every $z \in \mathbb{R}^d$. Fix then $t \in (0,\infty) \setminus \mathcal{D}_\rho$ and let $\mathcal{V} \subset \mathcal{S}^{d-1}$ be dense in $\mathcal{S}^{d-1}$ and at most countable. For any $v \in \mathcal{V}$, there exists a subset $E_v \subseteq \mathbb{R}^d$ with P-probability one such that $|v'(z-\mu_0)|\mathbb{I}[\|z-\mu_0\| = t] = 0$ for any $z \in E_v$. Since $\mathcal{V}$ is countable, $E := \cap_{v\in\mathcal{V}} E_v$ has probability one, too. Thus, $|v'(z-\mu_0)|\mathbb{I}[\|z-\mu_0\| = t] = 0$ for any $v \in \mathcal{V}$ and $z \in E$. Since $\mathcal{V}$ is dense in $\mathcal{S}^{d-1}$ and $v \mapsto |v'(z-\mu_0)|$ is continuous, we have $|v'(z-\mu_0)|\mathbb{I}[\|z-\mu_0\| = t] = 0$ for any $z \in E$ and any $v \in \mathcal{S}^{d-1}$. Taking the supremum over $v \in \mathcal{S}^{d-1}$ then yields $0 = \|z-\mu_0\|\mathbb{I}[\|z-\mu_0\| = t] = t\mathbb{I}[\|z-\mu_0\| = t]$ for any $z \in E$, hence $P[\|Z-\mu_0\| = t] = 0$. Since this holds for any t in the at most countable set $(0,\infty) \setminus \mathcal{D}_\rho$, we conclude that $P[\|Z-\mu_0\| \in (0,\infty) \setminus \mathcal{D}_\rho] = 0$.

Turning to sufficiency, assume now that $\psi_+(0)P[\{\mu_0\}] + P[\|Z-\mu_0\| \in (0,\infty)\setminus\mathcal{D}_\rho] = 0$. It then directly follows from Proposition 3.5.1 that $\nabla M^\rho_{\alpha,u}(\mu_0)$ takes the form given in the statement of the theorem (recall that we define the gradient as the vector stacking partial derivatives on top of each other, irrespective of whether or not the function is actually differentiable). Further observe that, since the assumption $\psi_+(0)P[\{\mu_0\}] + P[\|Z-\mu_0\| \in (0,\infty) \setminus \mathcal{D}_\rho] = 0$ implies that, for any $v \in \mathbb{R}^d \setminus \{0\}$,

$$\frac{\partial H^\rho_{\alpha,u}}{\partial(-v)}(z-\mu_0) = -\frac{\partial H^\rho_{\alpha,u}}{\partial v}(z-\mu_0)$$

for P-almost any z, (3.9.35) entails that

$$\nabla M^\rho_{\alpha,u}(\mu_0) = -\int_{\mathbb{R}^d\setminus\{\mu_0\}} \nabla H^\rho_{\alpha,u}(z-\mu_0)\, dP(z). \tag{3.9.36}$$

We want to show that

$$\lim_{\mu\to\mu_0} \frac{M^\rho_{\alpha,u}(\mu) - M^\rho_{\alpha,u}(\mu_0) - (\mu-\mu_0)'\nabla M^\rho_{\alpha,u}(\mu_0)}{\|\mu-\mu_0\|} = 0.$$

Let us then write

$$\frac{M^\rho_{\alpha,u}(\mu) - M^\rho_{\alpha,u}(\mu_0) - (\mu-\mu_0)'\nabla M^\rho_{\alpha,u}(\mu_0)}{\|\mu-\mu_0\|}$$

$$= \int_{\mathbb{R}^d} \frac{H^\rho_{\alpha,u}(z-\mu) - H^\rho_{\alpha,u}(z-\mu_0) + (\mu-\mu_0)'\nabla H^\rho_{\alpha,u}(z-\mu_0)}{\|\mu-\mu_0\|} \xi_{z,\mu}\xi_{z,\mu_0}\, dP(z)$$

$$+\frac{H^\rho_{\alpha,u}(\mu_0-\mu)}{\|\mu-\mu_0\|}P[\{\mu_0\}] - \frac{H_{\alpha,u}(\mu-\mu_0)}{\|\mu-\mu_0\|}P[\{\mu\}] + \frac{(\mu-\mu_0)'}{\|\mu-\mu_0\|}\nabla H^\rho_{\alpha,u}(\mu-\mu_0)P[\{\mu\}]. \tag{3.9.37}$$

Note that, as $\mu \to \mu_0$,

$$0 \le \frac{H^\rho_{\alpha,u}(\mu_0-\mu)}{\|\mu-\mu_0\|}P[\{\mu_0\}] \le (1+\alpha)\frac{\rho(\|\mu-\mu_0\|)}{\|\mu-\mu_0\|}P[\{\mu_0\}]$$

$$\le (1+\alpha)\psi_-(\|\mu-\mu_0\|)P[\{\mu_0\}] \to (1+\alpha)\psi_+(0)P[\{\mu_0\}] = 0,$$

so that the second term in (3.9.37) is $o(1)$ as $\mu \to \mu_0$. Proceeding in the same way and using Lemma 3.9.1 shows that the third term is also $o(1)$ as $\mu \to \mu_0$. Now, the same holds for the fourth term, since

$$\big\|\nabla H^{\rho}_{\alpha,u}(\mu - \mu_0)\big\| P[\{\mu\}] \leq (1+3\alpha)\psi_+(\|\mu - \mu_0\|) P[\{\mu\}]$$

(see Lemma 3.9.13(i)) converges to zero as $\mu \to \mu_0$ (this follows from Lemma 3.9.1 and from the monotonicity of ψ_+). Therefore it only remains to show that the first term in (3.9.37) is also $o(1)$ as $\mu \to \mu_0$.

To do so, let $z \in \mathbb{R}^d$ be such that $\|z - \mu_0\| \in \mathcal{D}_\rho$. With $\delta = \delta_{\mu_0}$ as in (3.2.4), Lemma 3.9.14 ensures that, as soon as $\|\mu - \mu_0\| < \delta$, we have that $|H^{\rho}_{\alpha,u}(z - \mu_0) - H^{\rho}_{\alpha,u}(z - \mu)|\xi_{z,\mu_0}\xi_{z,\mu}/\|\mu - \mu_0\|$ is upper-bounded by the P-integrable function $z \mapsto (1+3\alpha)\psi_-(\|z - \mu_0\| + \delta)$, that does not depend on μ. Moreover, Lemma 3.9.13(ii) yields

$$\left|\frac{(\mu - \mu_0)'\nabla H^{\rho}_{\alpha,u}(z - \mu_0)}{\|\mu - \mu_0\|}\right| \leq \|\nabla H^{\rho}_{\alpha,u}(z - \mu_0)\| \leq (1+3\alpha)\psi_-(\|z - \mu_0\|),$$

which is P-integrable and does not depend on μ. Since $P[\|Z - \mu_0\| \in \mathcal{D}_\rho] = 1$ by assumption, we may thus apply Lebesgue's DCT, which, by differentiability of $H^{\rho}_{\alpha,u}$ at $z - \mu_0 (\neq 0)$ (see Lemma 3.9.13(ii) again), entails that

$$\lim_{\mu \to \mu_0} \frac{M^{\rho}_{\alpha,u}(\mu_0) - M^{\rho}_{\alpha,u}(\mu) - \nabla M^{\rho}_{\alpha,u}(\mu_0)'(\mu - \mu_0)}{\|\mu - \mu_0\|} = 0.$$

We conclude that $M^{\rho}_{\alpha,u}$ is differentiable at μ_0.

(ii) Let $\mathcal{N}$ be an open subset of $\mathbb{R}^d$ such that $\psi_+(0)P[\{\mu\}] + P[\|Z - \mu\| \in (0,\infty)\backslash\mathcal{D}_\rho] = 0$ for any $\mu \in \mathcal{N}$. Part (i) of the result guarantees that $\mu \mapsto M^{\rho}_{\alpha,u}(\mu)$ is differentiable at any $\mu \in \mathcal{N}$, with corresponding gradient $\nabla M^{\rho}_{\alpha,u}(\mu) = v(\mu) - \alpha T(\mu)u$. Using again the inequality $\rho(t)/t \leq \psi_-(t)$ for any $t > 0$, a direct application of Lebesgue's DCT shows that the maps $\mu \mapsto v(\mu)$ and $\mu \mapsto T(\mu)$ are continuous on $\mathcal{N}$, so that the gradient map $\mu \mapsto \nabla M^{\rho}_{\alpha,u}(\mu)$ is also continuous on $\mathcal{N}$. This establishes continuous differentiability on $\mathcal{N}$. ■

The proof of Theorem 3.5.3 requires the following lemma.

Lemma 3.9.15. *Let the assumptions of Theorem 3.5.3 hold. If Assumption (A) holds, then, for any* $\mu \in \mathbb{R}^d$, $v \in \mathbb{R}^d \setminus \{0\}$, $t > 0$, *and* $z \in \mathbb{R}^d \setminus \{\mu, \mu + tv\}$,

(i) $$\left|\frac{\rho(\|z - \mu - tv\|)}{\|z - \mu - tv\|} - \frac{\rho(\|z - \mu\|)}{\|z - \mu\|}\right| \leq \frac{\psi_-(\|z - \mu\|)}{\|z - \mu\|} t\|v\|$$

and

(ii) $$\big|\psi_-(\|z - \mu - tv\|) - \psi_-(\|z - \mu\|)\big| \leq \frac{\psi_-(\|z - \mu\|)}{\|z - \mu\|} t\|v\|,$$

whereas if Assumption (A′) holds, then

(iii) $$\left|\frac{\rho(\|z - \mu - tv\|)}{\|z - \mu - tv\|} - \frac{\rho(\|z - \mu\|)}{\|z - \mu\|}\right| \leq \frac{\psi_-(\|z - \mu\| + r)}{\|z - \mu\| + r} t\|v\|$$

for any $\mu \in \mathbb{R}^d$, $v \in \mathbb{R}^d \setminus \{0\}$, $t \in (0, r/\|v\|]$, $z \in \mathbb{R}^d \setminus \{\mu, \mu + tv\}$, *and*

(iv) $\big|\psi_-(\|z-\mu-tv\|) - \psi_-(\|z-\mu\|)\big| \le \psi_-'(\|z-\mu\|+r)t\|v\|$

for any $\mu \in \mathbb{R}^d$, $v \in \mathbb{R}^d \setminus \{0\}$, $t \in (0, r/\|v\|]$, *and* $z \in \mathbb{R}^d$ *(in (iii)–(iv), r is as in Assumption (A′)).*

Proof of Lemma 3.9.15. (i) Let $\mu \in \mathbb{R}^d$, $v \in \mathbb{R}^d \setminus \{0\}$, $t > 0$, and $z \in \mathbb{R}^d \setminus \{\mu, \mu + tv\}$. Assume first that $\|z-\mu-tv\| < \|z-\mu\|$. Since $s \mapsto \rho(s)/s$ is monotone non-decreasing on $(0,\infty)$ (Lemma 3.9.4), we have

$$\begin{aligned}\left|\frac{\rho(\|z-\mu-tv\|)}{\|z-\mu-tv\|} - \frac{\rho(\|z-\mu\|)}{\|z-\mu\|}\right| &= \frac{\rho(\|z-\mu\|)}{\|z-\mu\|} - \frac{\rho(\|z-\mu-tv\|)}{\|z-\mu-tv\|} \\ &\le \frac{\rho(\|z-\mu\|)}{\|z-\mu\|} - \frac{\|z-\mu-tv\|}{\|z-\mu\|}\frac{\rho(\|z-\mu-tv\|)}{\|z-\mu-tv\|} \\ &= \frac{\rho(\|z-\mu\|) - \rho(\|z-\mu-tv\|)}{\|z-\mu\|}.\end{aligned}$$

Using Lemma 3.9.3 and the fact that ψ_- is monotone non-decreasing, we thus obtain (here, $c \in (\|z-\mu-tv\|, \|z-\mu\|)$)

$$\begin{aligned}\left|\frac{\rho(\|z-\mu-tv\|)}{\|z-\mu-tv\|} - \frac{\rho(\|z-\mu\|)}{\|z-\mu\|}\right| &\le \frac{\psi_-(c)\big(\|z-\mu\| - \|z-\mu-tv\|\big)}{\|z-\mu\|} \\ &\le \frac{\psi_-(\|z-\mu\|)}{\|z-\mu\|}t\|v\|.\end{aligned}$$

Assume then that $\|z-\mu-tv\| > \|z-\mu\|$. The same reasoning leads to

$$\begin{aligned}\left|\frac{\rho(\|z-\mu-tv\|)}{\|z-\mu-tv\|} - \frac{\rho(\|z-\mu\|)}{\|z-\mu\|}\right| &= \frac{\rho(\|z-\mu-tv\|)}{\|z-\mu-tv\|} - \frac{\rho(\|z-\mu\|)}{\|z-\mu\|} \\ &\le \frac{\rho(\|z-\mu-tv\|)}{\|z-\mu-tv\|} - \frac{\|z-\mu\|}{\|z-\mu-tv\|}\frac{\rho(\|z-\mu\|)}{\|z-\mu\|} \\ &= \frac{\rho(\|z-\mu-tv\|) - \rho(\|z-\mu\|)}{\|z-\mu-tv\|},\end{aligned}$$

which yields (for some $c \in (\|z-\mu\|, \|z-\mu-tv\|)$)

$$\begin{aligned}\left|\frac{\rho(\|z-\mu-tv\|)}{\|z-\mu-tv\|} - \frac{\rho(\|z-\mu\|)}{\|z-\mu\|}\right| &\le \frac{\psi_-(c)\big(\|z-\mu-tv\| - \|z-\mu\|\big)}{\|z-\mu-tv\|} \\ &\le \frac{\psi_-(\|z-\mu-tv\|)}{\|z-\mu-tv\|}t\|v\| \le \frac{\psi_-(\|z-\mu\|)}{\|z-\mu\|}t\|v\|,\end{aligned}$$

where the last inequality follows from the fact that

$$s \mapsto \frac{\psi_-(s)}{s} = \frac{\psi_-(s) - \psi_+(0)}{s-0} + \frac{\psi_+(0)}{s}$$

is, as the sum of two monotone non-increasing functions on $(0,\infty)$ (recall that ψ_- is concave under Assumption (A)), itself monotone non-increasing on $(0,\infty)$. Since the inequality in (i) trivially holds when $\|z-\mu-tv\| = \|z-\mu\|$, the result is proved.

(ii) Let $\mu \in \mathbb{R}^d$, $v \in \mathbb{R}^d \setminus \{0\}$, $t > 0$, and $z \in \mathbb{R}^d \setminus \{\mu, \mu + tv\}$. If $\|z - \mu - tv\| < \|z - \mu\|$, then

$$\begin{aligned}|\psi_-(\|z-\mu-tv\|)-\psi_-(\|z-\mu\|)| &= \psi_-(\|z-\mu\|)-\psi_-(\|z-\mu-tv\|)\\ &= \frac{\psi_-(\|z-\mu\|)-\psi_-(\|z-\mu-tv\|)}{\|z-\mu\|-\|z-\mu-tv\|}\big(\|z-\mu\|-\|z-\mu-tv\|\big)\\ &\le \frac{\psi_-(\|z-\mu\|)}{\|z-\mu\|}\,t\|v\|,\end{aligned}$$

where we used the fact that, since $s \mapsto \psi_-(s)/s$ is monotone non-increasing,

$$\frac{\psi_-(u)-\psi_-(v)}{u-v} \le \frac{\psi_-(u)}{u}$$

for any $u > v > 0$. If $\|z - \mu - tv\| > \|z - \mu\|$, then the same argument yields

$$\begin{aligned}|\psi_-(\|z-\mu-tv\|)-\psi_-(\|z-\mu\|)| &= \psi_-(\|z-\mu-tv\|)-\psi_-(\|z-\mu\|)\\ &= \frac{\psi_-(\|z-\mu-tv\|)-\psi_-(\|z-\mu\|)}{\|z-\mu-tv\|-\|z-\mu\|}\big(\|z-\mu-tv\|-\|z-\mu\|\big)\\ &\le \frac{\psi_-(\|z-\mu-tv\|)}{\|z-\mu-tv\|}\,t\|v\| \le \frac{\psi_-(\|z-\mu\|)}{\|z-\mu\|}\,t\|v\|,\end{aligned}$$

where the last inequality results from the fact that $s \mapsto \psi_-(s)/s$ is monotone non-increasing. Since the inequality in (ii) also trivially holds when $\|z - \mu - tv\| = \|z - \mu\|$, the result is proved.

(iii) Let $\mu \in \mathbb{R}^d$, $v \in \mathbb{R}^d \setminus \{0\}$, $t \in (0, r/\|v\|]$, and $z \in \mathbb{R}^d \setminus \{\mu, \mu + tv\}$. Proceeding as in Part (i) of the proof yields

$$\begin{aligned}&\left|\frac{\rho(\|z-\mu-tv\|)}{\|z-\mu-tv\|}-\frac{\rho(\|z-\mu\|)}{\|z-\mu\|}\right|\\ &\qquad\le \max\left(\frac{\psi_-(\|z-\mu\|)}{\|z-\mu\|},\frac{\psi_-(\|z-\mu-tv\|)}{\|z-\mu-tv\|}\right)t\|v\|.\end{aligned}$$

Since convexity of ψ_- implies that

$$t \mapsto \frac{\psi_-(t)}{t} = \frac{\psi_-(t)-\psi_+(0)}{t-0}$$

is monotone non-decreasing on $(0, \infty)$, we then obtain

$$\left|\frac{\rho(\|z-\mu-tv\|)}{\|z-\mu-tv\|}-\frac{\rho(\|z-\mu\|)}{\|z-\mu\|}\right| \le \frac{\psi_-(\|z-\mu\|+r)}{\|z-\mu\|+r}\,t\|v\|.$$

(iv) Let $\mu \in \mathbb{R}^d$, $v \in \mathbb{R}^d \setminus \{0\}$, $t \in (0, r/\|v\|]$, and $z \in \mathbb{R}^d$. Lemma 3.9.3 guarantees the existence of a c between $\|z - \mu\|$ and $\|z - \mu - tv\|$ such that

$$\begin{aligned}\big|\psi_-(\|z-\mu-tv\|)-\psi_-(\|z-\mu\|)\big| &\le \psi_-'(c)\big|\|z-\mu-tv\|-\|z-\mu\|\big|\\ &\le \psi_-'(c)t\|v\|.\end{aligned}$$

Since convexity of ψ_- implies that $t \mapsto \psi'_-(t)$ is monotone non-decreasing on $(0, \infty)$, we have $\psi'_-(c) \le \psi'_-(\max(\|z-\mu\|, \|z-\mu-tv\|)) \le \psi'_-(\|z-\mu\|+r)$, which yields the desired inequality. ■

Proof of Theorem 3.5.3. We first prove the result under Assumption (A). Let $\mathcal{N}$ be an open neighborhood of $\{\mu_0\}$ such that $P[\{\mu\}] + P[\|Z-\mu\| \in (0,\infty) \setminus \mathcal{D}_\rho] = 0$ for any $\mu \in \mathcal{N}$. Theorem 3.5.2 ensures that $M^\rho_{\alpha,u}$ is continuously differentiable on $\mathcal{N}$, with corresponding gradient (below, we may define $\nabla H^\rho_{\alpha,u}(z-\mu)$ arbitrarily at the z values where the gradient is undefined, since the collection of such z values has P-probability zero by assumption)

$$
\begin{aligned}
\nabla M^\rho_{\alpha,u}(\mu) &= -\int_{\mathbb{R}^d \setminus \{\mu\}} \nabla H^\rho_{\alpha,u}(z-\mu)\, dP(z) \\
&= \int_{\mathbb{R}^d} \bigg\{ -\psi_-(\|z-\mu\|)\Big(1+\alpha \frac{u'(z-\mu)}{\|z-\mu\|}\Big) \frac{z-\mu}{\|z-\mu\|}\xi_{z,\mu} \\
&\qquad -\alpha u \frac{\rho(\|z-\mu\|)}{\|z-\mu\|}\xi_{z,\mu} + \alpha \frac{\rho(\|z-\mu\|)}{\|z-\mu\|}\frac{u'(z-\mu)}{\|z-\mu\|}\frac{z-\mu}{\|z-\mu\|}\xi_{z,\mu} \bigg\}\, dP(z)
\end{aligned}
$$

for any $\mu \in \mathcal{N}$: see (3.9.36) and Lemma 3.9.13(ii). Fix $v \in \mathbb{R}^d \setminus \{0\}$ and $t \in (0, t_0)$, where t_0 is such that $B(\mu_0, t_0\|v\|) \subset \mathcal{N}$. Let then

$$
\begin{aligned}
f_1(z) &= \psi_-(\|z-\mu_0\|),\ f_2(z) = 1 + \alpha u' \frac{z-\mu_0}{\|z-\mu_0\|},\ f_3(z) = \frac{z-\mu_0}{\|z-\mu_0\|}, \\
g_1(z) &= \frac{\rho(\|z-\mu_0\|)}{\|z-\mu_0\|},\ g_2(z) = \frac{u'(z-\mu_0)}{\|z-\mu_0\|},\ g_3(z) = \frac{z-\mu_0}{\|z-\mu_0\|}.
\end{aligned}
$$

Since P is non-atomic in $\mathcal{N}$, we have

$$
\begin{aligned}
&\frac{\nabla M^\rho_{\alpha,u}(\mu_0+tv) - \nabla M^\rho_{\alpha,u}(\mu_0)}{t} \\
&\quad = -\int_{\mathbb{R}^d \setminus \{\mu_0\}} \frac{\nabla H^\rho_{\alpha,u}(z-\mu_0-tv) - \nabla H^\rho_{\alpha,u}(z-\mu_0)}{t} \xi_{z,\mu_0+tv}\, dP(z) \\
&\quad = \int_{\mathbb{R}^d \setminus \{\mu_0\}} \{I_{t,1}(z) + I_{t,2}(z) + I_{t,3}(z)\} \xi_{z,\mu_0+tv}\, dP(z),
\end{aligned}
$$

where we let

$$
\begin{aligned}
I_{t,1}(z) &= \frac{f_1(z)f_2(z)f_3(z) - f_1(z-tv)f_2(z-tv)f_3(z-tv)}{t} \\
I_{t,2}(z) &= \alpha u\, \frac{g_1(z) - g_1(z-tv)}{t},
\end{aligned}
$$

and

$$
I_{t,3}(z) = \alpha\, \frac{g_1(z-tv)g_2(z-tv)g_3(z-tv) - g_1(z)g_2(z)g_3(z)}{t}.
$$

Decompose $I_{t,1}(z)$ into

$$I_{t,1}(z) = \frac{f_1(z) - f_1(z-tv)}{t} f_2(z-tv) f_3(z-tv)$$
$$+ f_1(z)\frac{f_2(z) - f_2(z-tv)}{t} f_3(z-tv) + f_1(z) f_2(z) \frac{f_3(z) - f_3(z-tv)}{t}.$$

Applying Lemma 3.9.15(ii) and Lemma 4.8.5, we obtain

$$\|I_{t,1}(z)\| \le 2\|v\| \frac{\psi_-(\|z-\mu_0\|)}{\|z-\mu_0\|} + 2\|v\| \frac{\psi_-(\|z-\mu_0\|)}{\|z-\mu_0\|} + 4\|v\| \frac{\psi_-(\|z-\mu_0\|)}{\|z-\mu_0\|}$$

for any $z \in \mathbb{R}^d \setminus \{\mu_0, \mu_0 + tv\}$. Lemma 3.9.15(i) directly yields

$$\|I_{t,2}(z)\| \le \|v\| \frac{\psi_-(\|z-\mu_0\|)}{\|z-\mu_0\|}$$

for any $z \in \mathbb{R}^d \setminus \{\mu_0, \mu_0 + tv\}$. Finally, considering a decomposition similar as the one used for $I_{t,1}(z)$, Lemma 3.9.15(i) and Lemma 4.8.5 provide

$$\|I_{t,3}(z)\| \le \|v\| \frac{\psi_-(\|z-\mu_0\|)}{\|z-\mu_0\|} + 2\|v\| \frac{\rho(\|z-\mu_0\|)}{\|z-\mu_0\|^2} + 2\|v\| \frac{\rho(\|z-\mu_0\|)}{\|z-\mu_0\|^2}$$

for any $z \in \mathbb{R}^d \setminus \{\mu_0, \mu_0 + tv\}$. Since $\rho(t)/t \le \psi_-(t)$ for any $t > 0$, we conclude that $z \mapsto \|I_{t,1}(z) + I_{t,2}(z) + I_{t,3}(z)\|$ is upper-bounded P-almost everywhere by $z \mapsto C\|v\|\psi_-(\|z-\mu_0\|)/\|z-\mu_0\|$ (here, C is a positive real constant), which is a P-integrable function by assumption. For any $i = 1, \ldots, d$, Lebesgue's DCT therefore provides

$$\lim_{t \overset{>}{\to} 0} \frac{\partial_i M^\rho_{\alpha,u}(\mu_0 + tv) - \partial_i M^\rho_{\alpha,u}(\mu_0)}{t}$$
$$= -\int_{\mathbb{R}^d \setminus \{\mu_0\}} \lim_{t \overset{>}{\to} 0} \frac{\partial_i H^\rho_{\alpha,u}(z - \mu_0 - tv) - \partial_i H^\rho_{\alpha,u}(z-\mu_0)}{t} \xi_{z,\mu_0+tv} \, dP(z)$$
$$= \sum_{j=1}^d v_j \int_{\mathbb{R}^d \setminus \{\mu_0\}} \partial_j \partial_i H^\rho_{\alpha,u}(z-\mu_0) \, dP(z) = (\nabla^2 M^\rho_{\alpha,u}(\mu_0) v)_i, \tag{3.9.38}$$

where the limit of the integrand exists for P-almost any $z \in \mathbb{R}^d$ (for the other values of z, the integrand in the last integral may be defined arbitrarily). The form of the Hessian matrix provided in the statement of the theorem then follows from Lemma 3.9.13(ii).

It remains to show the result also holds when Assumption (A′) holds. The proof proceeds along the same lines, but the upper-bounding of $z \mapsto \|I_{t,1}(z) + I_{t,2}(z) + I_{t,3}(z)\|$ is now based on Lemma 3.9.15(iii)–(iv). More precisely, Lemma 3.9.15(iv) and Lemma 4.8.5 entail that

$$\|I_{t,1}(z)\| \le 2\|v\|\psi'_-(\|z-\mu_0\| + r) + 2\|v\| \frac{\psi_-(\|z-\mu_0\|)}{\|z-\mu_0\|} + 4\|v\| \frac{\psi_-(\|z-\mu_0\|)}{\|z-\mu_0\|}$$

for any $z \in \mathbb{R}^d \setminus \{\mu_0, \mu_0 + tv\}$ and $t \in (0, r/\|v\|]$. Lemma 3.9.15(iii) directly yields

$$\|I_{t,2}(z)\| \le \|v\| \frac{\psi_-(\|z-\mu_0\| + r)}{\|z-\mu_0\| + r}$$

for any $z \in \mathbb{R}^d \setminus \{\mu_0, \mu_0 + tv\}$ and $t \in (0, r/\|v\|]$. Finally, considering a decomposition similar as the one used for $I_{t,1}(z)$, Lemma 3.9.15(iii) and Lemma 4.8.5 provide

$$\|I_{t,3}(z)\| \leq \|v\| \frac{\psi_-(\|z-\mu_0\|+r)}{\|z-\mu_0\|+r} + 2\|v\| \frac{\rho(\|z-\mu_0\|)}{\|z-\mu_0\|^2} + 2\|v\| \frac{\rho(\|z-\mu_0\|)}{\|z-\mu_0\|^2}$$

for any $z \in \mathbb{R}^d \setminus \{\mu_0, \mu_0 + tv\}$ and $t \in (0, r/\|v\|]$. Since $\rho(t)/t^2 \leq \psi_-(t)/t \leq \psi'_-(t)$ for any $t > 0$ and $t \mapsto \psi'_-(t)$ is monotone non-decreasing on $(0, \infty)$, we conclude that $z \mapsto \|I_{t,1}(z) + I_{t,2}(z) + I_{t,3}(z)\|$ is upper-bounded P-almost everywhere by $z \mapsto C\|v\|\psi'_-(\|z-\mu_0\|+r)$ (here, C is a positive real constant), which is a P-integrable function by assumption. Consequently, we can still apply Lebesgue's DCT, which shows that (3.9.38) holds under Assumption (A′), too. ■

Proofs for Section 3.6

Proof of Proposition 3.6.2. Let $(\alpha_n u_n)$ be a sequence in $\mathcal{B}^d_{\alpha_\rho}$ converging to $\alpha u \in \mathcal{B}^d_{\alpha_\rho}$. We want to show that $Q(\alpha_n u_n) \to Q(\alpha u)$, that is, $\mu^\rho_{\alpha_n,u_n} \to \mu^\rho_{\alpha,u}$. Since $M^\rho_{\alpha_n,u_n}(\mu^\rho_{\alpha,u}) \to M^\rho_{\alpha,u}(\mu^\rho_{\alpha,u})$ (Lemma 3.9.5), $(M^\rho_{\alpha_n,u_n}(\mu^\rho_{\alpha,u}))$ is a bounded sequence. Since $M^\rho_{\alpha_n,u_n}(\mu^\rho_{\alpha_n,u_n}) \leq M^\rho_{\alpha_n,u_n}(\mu^\rho_{\alpha,u})$ for any n by definition of ρ-quantiles, the sequence $(M^\rho_{\alpha_n,u_n}(\mu^\rho_{\alpha_n,u_n}))$ is upper-bounded. Since (α_n) is a sequence in $[0, \alpha_\rho)$, we have that

$$\limsup_{n\to\infty} \alpha_n u_n' \mu^\rho_{\alpha_n,u_n} / \|\mu^\rho_{\alpha_n,u_n}\| < 1\cdot$$

Lemma 3.9.6 and the fact that $(M^\rho_{\alpha_n,u_n}(\mu^\rho_{\alpha_n,u_n}))$ is upper-bounded then entail that the sequence $(\mu^\rho_{\alpha_n,u_n})$ is bounded. Now, assume ad absurdum that $(\mu^\rho_{\alpha_n,u_n})$ does not converge to $\mu^\rho_{\alpha,u}$. Upon extraction of a subsequence, we may assume that there exists $\varepsilon > 0$ such that $\|\mu^\rho_{\alpha_n,u_n} - \mu^\rho_{\alpha,u}\| \geq \varepsilon$ for any n. This implies that no subsequence of $(\mu^\rho_{\alpha_n,u_n})$ converges to $\mu^\rho_{\alpha,u}$. However, since $(\mu^\rho_{\alpha_n,u_n})$ is a bounded sequence, it has a subsequence $(\mu^\rho_{\alpha_{n_k},u_{n_k}})$ that converges in $\mathbb{R}^d$, to v say. By taking limits as $k \to \infty$ in both sides of

$$M^\rho_{\alpha_{n_k},u_{n_k}}(\mu^\rho_{\alpha_{n_k},u_{n_k}}) \leq M^\rho_{\alpha_{n_k},u_{n_k}}(\mu^\rho_{\alpha,u}),$$

Lemma 3.9.5 then yields $M^\rho_{\alpha,u}(v) \leq M^\rho_{\alpha,u}(\mu_{\alpha,u})$. Since $\mu^\rho_{\alpha,u}$ is the unique minimizer of $M^\rho_{\alpha,u}$ (Theorem 3.3.4), we have $v = \mu^\rho_{\alpha,u}$, so that we identified a subsequence of $(\mu^\rho_{\alpha_n,u_n})$ that converges to $\mu^\rho_{\alpha,u}$. Since this is a contradiction, we conclude that $(\mu^\rho_{\alpha_n,u_n})$ converges to $\mu^\rho_{\alpha,u}$. ■

Lemma 3.9.16. *Let $\rho \in \mathcal{C}$ and assume that $P \in \mathcal{P}^d_\rho$ is not a Dirac probability measure. Then, for any $\mu \in \mathbb{R}^d$, the matrix $T(\mu)$ defined in Theorem 3.5.2 is positive definite, hence invertible.*

Proof of Lemma 3.9.16. First note that, since $\rho(t)/t \leq \psi_-(t)$ for any $t > 0$, we have

$$\begin{aligned} v'T(\mu)v &= \mathrm{E}\bigg[\bigg\{\frac{\rho(\|Z-\mu\|)}{\|Z-\mu\|} \\ &\qquad + \bigg(\psi_-(\|Z-\mu\|) - \frac{\rho(\|Z-\mu\|)}{\|Z-\mu\|}\bigg)\frac{\{v'(Z-\mu)\}^2}{\|Z-\mu\|^2}\bigg\}\xi_{Z,\mu}\bigg] \\ &\geq \mathrm{E}\bigg[\frac{\rho(\|Z-\mu\|)}{\|Z-\mu\|}\xi_{Z,\mu}\bigg] \geq 0 \end{aligned}$$

for any $\mu \in \mathbb{R}^d$ and $v \in \mathcal{S}^{d-1}$. Now, assume ad absurdum that there exist $\mu \in \mathbb{R}^d$ and $v \in \mathcal{S}^{d-1}$ such that $v'T(\mu)v = 0$. We must then have

$$\int_{\mathbb{R}^d \setminus \{\mu\}} \frac{\rho(\|z-\mu\|)}{\|z-\mu\|}\, dP(z) = 0,$$

which, since $\rho(t) > 0$ for any $t > 0$, implies that $P[\{\mu\}] = 1$. This contradicts the assumption that P is not a Dirac probability measure. ■

Proof of Theorem 3.6.4. Under the assumptions of the theorem, $\mu = Q(\alpha u)$, with $\alpha \in [0, \alpha_\rho)$ and $u \in \mathcal{S}^{d-1}$, implies that $R(\mu) = \alpha u$; see the discussion above the statement of the theorem. It directly follows that $R \circ Q$ is the identity map on $\mathcal{B}^d_{\alpha_\rho}$, which in turn entails that $Q : \mathcal{B}^d_{\alpha_\rho} \to \mathcal{Z}_\rho$ and $R : \mathcal{Z}_\rho \to \mathcal{B}^d_{\alpha_\rho}$ are one-to-one maps. Now, since $R(\mu) = (T(u))^{-1}v(\mu)$ for any $\mu \in \mathbb{R}^d$, continuity of R is a direct consequence of the fact that, under the assumptions considered, the maps $\mu \mapsto T(\mu)$ and $\mu \mapsto v(\mu)$ are continuous on $\mathbb{R}^d$; see the proof of Theorem 3.5.2. Since continuity of Q was already established in Proposition 3.6.2, the result is proved. ■

The proof of Theorem 3.6.5 requires Lemma 3.9.18 below, which itself relies on the following preliminary result.

Lemma 3.9.17. *Let $\rho \in \mathcal{C}$ be such that $t \mapsto t^2/\rho(t)$ is concave on $(0, \infty)$. Then, $t \mapsto \rho(t)/t^2$ is convex and monotone non-increasing on $(0, \infty)$.*

Proof of Lemma 3.9.17. Since $t \mapsto g(t) = t^2/\rho(t)$ is concave on $(0, \infty)$, it is left-differentiable on $(0, \infty)$ and the corresponding left-derivative, g'_- say, is monotone non-increasing. Therefore, for any $t_0 > 0$, Lemma 3.9.3(i) yields

$$g(t) \leq g(t_0) + g'_-(t_0)(t - t_0)$$

for any $t > t_0$. If $g'_-(t_0) < 0$ for some $t_0 > 0$, then it follows that $g(t)$ is strictly negative for t large, which is a contradiction. Thus, $g'_-(t) \geq 0$ for any $t > 0$. Lemma 3.9.3(i) then implies that g is monotone non-decreasing on $(0, \infty)$, hence that $1/g$ is monotone non-increasing on $(0, \infty)$. It remains to show that $1/g$ is convex on $(0, \infty)$. To that end, fix $0 < s < t$ and $\lambda \in (0, 1)$. Using concavity of g, we obtain by Jensen's inequality (for the convex function $z \mapsto 1/z$ and for the measure attributing weight $1 - \lambda$ and λ to $g(s)$ and $g(t)$, respectively)

$$\frac{1}{g((1-\lambda)s + \lambda t)} \leq \frac{1}{(1-\lambda)g(s) + \lambda g(t)} \leq (1-\lambda)\frac{1}{g(s)} + \lambda\frac{1}{g(t)},$$

which establishes the result. ∎

Lemma 3.9.18. *Let $\rho \in \mathcal{C}$ be such that $t \mapsto t^2/\rho(t)$ is concave on $(0,\infty)$. Assume that P is not concentrated on a line. Assume further that $\psi_+(0)P[\{\mu\}] + P[\|Z-\mu\| \in (0,\infty) \setminus \mathcal{D}_\rho] = 0$ for any $\mu \in \mathbb{R}^d$. Then, there is no $u \in \mathcal{S}^{d-1}$ for which the equation $\nabla M^{\rho}_{1,u}(\mu) = 0$ has a solution.*

Proof of Lemma 3.9.18. Ad absurdum, let $u \in \mathcal{S}^{d-1}$ and $\mu \in \mathbb{R}^d$ be such that $\nabla M^{\rho}_{1,u}(\mu) = 0$. Thus, we must have $u'\nabla M^{\rho}_{1,u}(\mu) = 0$, which, in view of Theorem 3.5.2, rewrites

$$\mathrm{E}\bigg[\bigg\{\frac{\rho(\|Z-\mu\|)}{\|Z-\mu\|} + \bigg(\psi_-(\|Z-\mu\|) - \frac{\rho(\|Z-\mu\|)}{\|Z-\mu\|}\bigg)\frac{\{u'(Z-\mu)\}^2}{\|Z-\mu\|^2}\bigg\}\xi_{Z,\mu}\bigg] + \mathrm{E}\bigg[\psi_-(\|Z-\mu\|)\frac{u'(Z-\mu)}{\|Z-\mu\|}\xi_{Z,\mu}\bigg] = 0.$$

Straightforward computations allow us to rewrite this as

$$\mathrm{E}\Big[g_1(\|Z-\mu\|)\zeta^2_{Z,\mu} + g_2(\|Z-\mu\|)\zeta_{Z,\mu}\Big] = 0, \tag{3.9.39}$$

where we let

$$\zeta_{Z,\mu} := \bigg(1 + \frac{u'(Z-\mu)}{\|Z-\mu\|}\bigg)\xi_{Z,\mu},$$

$$g_1(t) := \psi_-(t) - \frac{\rho(t)}{t} \quad \text{and} \quad g_2(t) := \frac{2\rho(t)}{t} - \psi_-(t)$$

for any $t > 0$. Lemma 3.9.4 entails that $g_1(t) \geq 0$ for any $t > 0$. Since Lemma 3.9.17 implies that $t \mapsto \rho(t)/t^2$, hence also $t \mapsto \ln(\rho(t)/t^2)$, is non-increasing over $(0,\infty)$, we have (here, we consider left-differentiation)

$$0 \geq \bigg(\ln\Big(\frac{\rho(t)}{t^2}\Big)\bigg)' = \frac{\psi_-(t)}{\rho(t)} - \frac{2}{t} = \frac{1}{\rho(t)}\bigg(\psi_-(t) - \frac{2\rho(t)}{t}\bigg)$$

for any $t > 0$, so that we also have $g_2(t) \geq 0$ for any $t > 0$. Moreover, since $g_1(t) + g_2(t) = \rho(t)/t > 0$ for any $t > 0$, we have $\max(g_1(t), g_2(t)) > 0$ for any $t > 0$. Since the assumption that P is not concentrated on a line ensures that $A := \{z \in \mathbb{R}^d : \zeta_{z,\mu} = 0\}$ satisfies

$$P[A] = P[Z - \mu \in \{cu : c \leq 0\}] \leq P[Z \in \{\mu + cu : c \in \mathbb{R}\}] < 1,$$

we thus have

$$\mathrm{E}\Big[g_1(\|Z-\mu\|)\zeta^2_{Z,\mu} + g_2(\|Z-\mu\|)\zeta_{Z,\mu}\Big] = \int_{\mathbb{R}^d \setminus A} \big\{g_1(\|z-\mu\|)\zeta^2_{z,\mu} + g_2(\|z-\mu\|)\zeta_{z,\mu}\big\}\, dP(z) > 0,$$

since $g_1(\|z-\mu\|)\zeta^2_{z,\mu} + g_2(\|z-\mu\|)\zeta_{z,\mu} > 0$ for any $z \notin A$ (the discussion above implies that the nonnegative quantities $g_1(\|z-\mu\|)$ and $g_2(\|z-\mu\|)$ cannot be both zero at the same z). Since this contradicts (3.9.39), the result is proved. ■

Proof of Theorem 3.6.5. We first show that $R(\mathbb{R}^d) \subseteq \mathcal{B}^d$. To do so, assume, ad absurdum, that there exists $\mu \in \mathbb{R}^d$ such that $\|R(\mu)\| \geq 1$. We consider two cases. (a) $\|R(\mu)\| = 1$, so that $R(\mu) = u$ for some $u \in \mathcal{S}^{d-1}$. Then, $\nabla M^\rho_{1,u}(\mu) = T(\mu)(R(\mu) - u) = 0$, which, in view of Lemma 3.9.18, is a contradiction. (b) $\|R(\mu)\| > 1$. Fix then arbitrarily $\alpha_0 u_0 \in \mathcal{B}^d$, and observe that $\|R(\mu^\rho_{\alpha_0,u_0})\| = \|\alpha_0 u_0\| = \alpha_0 < 1$. Recalling that R is continuous (Theorem 3.6.4), the intermediate value theorem then guarantees that there exists $\lambda \in (0,1)$ such that $\|R((1-\lambda)\mu^\rho_{\alpha_0,u_0} + \lambda\mu)\| = 1$, which, proceeding as in (a), provides a contradiction. Therefore, $R(\mathbb{R}^d) \subseteq \mathcal{B}^d$.

Now, fix $\mu \in \mathbb{R}^d$. Since $R(\mathbb{R}^d) \subseteq \mathcal{B}^d$, there exist $\alpha \in [0,1)$ and $u \in \mathcal{S}^{d-1}$ such that $R(\mu) = \alpha u$, which implies that $\mu = \mu^\rho_{\alpha,u} = Q(\alpha u)$. Therefore, $\mathcal{Z}_\rho = Q(\mathcal{B}^d) = \mathbb{R}^d$, and the result follows from Theorem 3.6.4. ■

Proofs for Section 3.7

Proof of Theorem 3.7.1. (i) Assume, ad absurdum, that $\|\mu^\rho_{\alpha_n,u_n}\|$ does not diverge to ∞ as $n \to \infty$. The sequence $(\mu^\rho_{\alpha_n,u_n})$ then admits a subsequence that is bounded, hence a further subsequence, $(\mu^\rho_{\alpha_{n_\ell},u_{n_\ell}})$, that converges in $\mathbb{R}^d$. Let us denote as μ_* the corresponding limit. Since R is continuous, taking limits as $\ell \to \infty$ in both sides of $\|R(\mu^\rho_{\alpha_{n_\ell},u_{n_\ell}})\| = \|\alpha_{n_\ell} u_{n_\ell}\| = \alpha_{n_\ell}$ then provides $\|R(\mu_*)\| = 1$. Thus, $R(\mu_*) = v$ for some $v \in \mathcal{S}^{d-1}$, which implies that $\nabla M^\rho_{1,v}(\mu_*) = T(\mu_*)(R(\mu_*) - v) = 0$. Since this contradicts Lemma 3.9.18, the result is proved. (ii) We now assume that u_n converges to $u (\in \mathcal{S}^{d-1})$. Consider an arbitrary subsequence $(\omega_k = \mu^\rho_{\alpha_{n_k},u_{n_k}})$ of $(\mu^\rho_{\alpha_n,u_n})$ and fix $\mu_0 \in \mathbb{R}^d$ arbitrarily. The continuity of $(\alpha,u) \mapsto M^\rho_{\alpha,u}(\mu_0)$ (Lemma 3.9.5) implies that the sequence $(M^\rho_{\alpha_{n_k},u_{n_k}}(\mu_0))$ is bounded. Since, by definition,

$$M^\rho_{\alpha_{n_k},u_{n_k}}(\omega_k) \leq M^\rho_{\alpha_{n_k},u_{n_k}}(\mu_0)$$

for any k, the sequence $(M^\rho_{\alpha_{n_k},u_{n_k}}(\omega_k))$ is then upper-bounded, too. Assume now that

$$\limsup_{k\to\infty} u'_{n_k}\omega_k/\|\omega_k\| < 1.$$

Since $\|\omega_k\| \to \infty$, Lemma 3.9.6 then implies that $M_{\alpha_{n_k},u_{n_k}}(\omega_k) \to \infty$ as $k \to \infty$, which contradicts the fact that the sequence $(M^\rho_{\alpha_{n_k},u_{n_k}}(\omega_k))$ is bounded. Therefore, we must have

$$\limsup_{k\to\infty} u'_{n_k}\omega_k/\|\omega_k\| = 1.$$

Since u_{n_k} converges to u, this entails that $\limsup_{k\to\infty} u'\omega_k/\|\omega_k\| = 1$, so that (ω_k) admits a subsequence (ω_{k_ℓ}) for which $u'_{n_{k_\ell}}\omega_{k_\ell}/\|\omega_{k_\ell}\| \to 1$ as $\ell \to \infty$. We thus proved that any subsequence of $(\mu^\rho_{\alpha_n,u_n}/\|\mu^\rho_{\alpha_n,u_n}\|)$ admits a further subsequence converging to u. This implies that $\mu^\rho_{\alpha_n,u_n}/\|\mu^\rho_{\alpha_n,u_n}\| \to u$ as $n \to \infty$. ■

The proof of Theorem 3.7.2 requires the following result.

Lemma 3.9.19. *Let the assumptions of Theorem 3.7.2 hold. Then, for any sequences* $(\alpha_n) \in [0,1)$, $(u_n) \in \mathcal{S}^{d-1}$, *and* $(\mu_n) \in \mathbb{R}^d$ *such that* $\|\mu_n\| \to \infty$, *we have that* $M^{\rho}_{\alpha_n,u_n}(\mu_n) \to \infty$ *as* $n \to \infty$.

Proof of Lemma 3.9.19. We prove the result by showing that any subsequence of $(M^{\rho}_{\alpha_n,u_n}(\mu_n))$ has a further subsequence converging to ∞. To do so, fix a subsequence (μ_{n_k}) of (μ_n). If $\limsup_{k\to\infty} \alpha_{n_k} u'_{n_k}\mu_{n_k}/\|\mu_{n_k}\| < 1$, then Lemma 3.9.6 yields $M^{\rho}_{\alpha_{n_k},u_{n_k}}(\mu_{n_k}) \to \infty$ as $k \to \infty$. We may thus assume that $\limsup_{k\to\infty} \alpha_{n_k} u'_{n_k}\mu_{n_k}/\|\mu_{n_k}\| = 1$. Take then a subsequence (k_ℓ) such that $\alpha_{n_{k_\ell}} u'_{n_{k_\ell}}\mu_{n_{k_\ell}}/\|\mu_{n_{k_\ell}}\| \to 1$ as $\ell \to \infty$. By abuse of notation, we write α_ℓ, u_ℓ and μ_ℓ instead of $\alpha_{n_{k_\ell}}$, $u_{n_{k_\ell}}$ and $\mu_{n_{k_\ell}}$, respectively. By compactness of $[0,1]$ and $\mathcal{S}^{d-1}$, we can assume, up to further extraction of a subsequence, that $\alpha_\ell \to 1$, $u_\ell \to u \in \mathcal{S}^{d-1}$ and $u'_\ell\mu_\ell/\|\mu_\ell\| \to 1$ as $\ell \to \infty$.

For any $z \in \mathbb{R}^d$ and $v \in \mathcal{S}^{d-1}$, denote as $D_v(z) = \{z + \lambda v : \lambda \in \mathbb{R}\}$ the line through z with direction v, and note that the distance between $y (\in \mathbb{R}^d)$ and $D_v(z)$ is given by

$$d_v(y,z) = \|(I_d - vv')(z-y)\|. \tag{3.9.40}$$

It will then play a key role in the proof that the P-probability of

$$E := \Big\{ z \in \mathbb{R}^d : \liminf_{\ell\to\infty} d_{u_\ell}(\mu_\ell, z) > 0 \Big\}$$

is positive. To address this point, take $z_0 \in \mathbb{R}^d \setminus E$ (if no such z_0 exists, then $E = \mathbb{R}^d$ has P-probability one). Up to extraction of yet another subsequence (which we do again without changing the notation), we may assume that $\lim_{\ell\to\infty} d_{u_\ell}(\mu_\ell, z_0) = 0$. For any $z \notin D_u(z_0)$, the inequality

$$d_{u_\ell}(\mu_\ell, z) \geq d_{u_\ell}(z, z_0) - d_{u_\ell}(\mu_\ell, z_0)$$

(which readily follows from using the triangle inequality in (3.9.40)) yields

$$\liminf_{\ell\to\infty} d_{u_\ell}(\mu_\ell, z) \geq \lim_{\ell\to\infty} d_{u_\ell}(z, z_0) - \lim_{\ell\to\infty} d_{u_\ell}(\mu_\ell, z_0) = d_u(z, z_0) > 0.$$

This shows that $\mathbb{R}^d / D_u(z_0) \subset E$, so that the assumption that P is not concentrated on a line yields $P[E] \geq P[\mathbb{R}^d / D_u(z_0)] > 0$.

To proceed with the proof, we need to consider two cases, according to the assumptions considered.

(i) We assume first that $\rho(t)/t^2 \to \infty$ as $t \to \infty$ and that Condition (b) holds. Let us then write

$$M^{\rho}_{\alpha_\ell,u_\ell}(\mu_\ell) = \mathcal{J}_1(\alpha_\ell, u_\ell, \mu_\ell) + \mathcal{J}_2(\alpha_\ell, u_\ell, \mu_\ell) + \mathcal{J}_3(\alpha_\ell, u_\ell, \mu_\ell)$$

with

$$\mathcal{J}_1(\alpha, u, \mu) = \int_{\mathbb{R}^d} \big(\rho(\|z-\mu\|) - \rho(\|z\|)\big)\bigg(1 + \alpha \frac{u'(z-\mu)}{\|z-\mu\|}\bigg)\xi_{z,\mu}\, dP(z),$$

$$\begin{aligned}
\mathcal{J}_2(\alpha, u, \mu) &= \int_{\mathbb{R}^d\setminus\{0,\mu\}} \rho(\|z\|)\bigg\{\bigg(1 + \alpha\frac{u'(z-\mu)}{\|z-\mu\|}\bigg)\xi_{z,\mu} - \bigg(1 + \alpha\frac{u'z}{\|z\|}\bigg)\xi_{z,0}\bigg\}\, dP(z) \\
&= \int_{\mathbb{R}^d\setminus\{0\}} \alpha\rho(\|z\|)\bigg(\frac{u'(z-\mu)}{\|z-\mu\|} - \frac{u'z}{\|z\|}\bigg)\xi_{z,\mu}\, dP(z),
\end{aligned}$$

and

$$\begin{aligned}\mathcal{J}_3(\alpha,u,\mu) &= \int_{\{0,\mu\}} \rho(\|z\|)\bigg\{\bigg(1+\alpha\frac{u'(z-\mu)}{\|z-\mu\|}\bigg)\xi_{z,\mu} - \bigg(1+\alpha\frac{u'z}{\|z\|}\bigg)\xi_{z,0}\bigg\}\,dP(z) \\ &= -\rho(\|\mu\|)\bigg(1+\alpha\frac{u'\mu}{\|\mu\|}\bigg)P[\{\mu\}].\end{aligned}$$

Let us observe that for any $z \in \mathbb{R}^d$ such that $\|z-\mu_\ell\| > (u'_\ell(z-\mu_\ell))^2$,

$$\begin{aligned}1+\alpha_\ell\frac{u'_\ell(z-\mu_\ell)}{\|z-\mu_\ell\|} &= \frac{1-\alpha_\ell^2\Big(\frac{u'_\ell(z-\mu_\ell)}{\|z-\mu_\ell\|}\Big)^2}{1-\alpha_\ell\frac{u'(z-\mu_\ell)}{\|z-\mu_\ell\|}} \\ &= \frac{1-\Big(\frac{u'_\ell(z-\mu_\ell)}{\|z-\mu_\ell\|}\Big)^2+\Big(\frac{u'_\ell(z-\mu_\ell)}{\|z-\mu_\ell\|}\Big)^2(1-\alpha_\ell^2)}{1-\alpha_\ell\frac{u'_\ell(z-\mu_\ell)}{\|z-\mu_\ell\|}} \\ &= \frac{d^2_{u_\ell}(\mu_\ell,z)}{\|z-\mu_\ell\|^2\Big(1-\alpha_\ell\frac{u'_\ell(z-\mu_\ell)}{\|z-\mu_\ell\|}\Big)} + \bigg(\frac{u'_\ell(z-\mu_\ell)}{\|z-\mu_\ell\|}\bigg)^2\frac{1-\alpha_\ell^2}{1-\alpha_\ell\frac{u'_\ell(z-\mu_\ell)}{\|z-\mu_\ell\|}} \\ &\geq \frac{d^2_{u_\ell}(\mu_\ell,z)}{\|z-\mu_\ell\|^2\Big(1-\alpha_\ell\frac{u'_\ell(z-\mu_\ell)}{\|z-\mu_\ell\|}\Big)}.\end{aligned}$$

For any $z \in E$, we then have

$$\begin{aligned}&\liminf_{\ell\to\infty}\rho(\|z-\mu_\ell\|)\Big(1+\alpha_\ell\frac{u'_\ell(z-\mu_\ell)}{\|z-\mu_\ell\|}\Big) \\ &\qquad\geq \liminf_{\ell\to\infty}\frac{\rho(\|z-\mu_\ell\|)d^2_{u_\ell}(\mu_\ell,z)}{\|z-\mu_\ell\|^2\Big(1-\alpha_\ell\frac{u'_\ell(z-\mu_\ell)}{\|z-\mu_\ell\|}\Big)} \\ &\qquad\geq \liminf_{\ell\to\infty}\frac{\rho(\|z-\mu_\ell\|)}{\|z-\mu_\ell\|^2}\times\liminf_{\ell\to\infty}\frac{d^2_{u_\ell}(\mu_\ell,z)}{1-\alpha_\ell\frac{u'_\ell(z-\mu_\ell)}{\|z-\mu_\ell\|}} \\ &\qquad= \infty,\end{aligned}$$

as $\rho(t)/t^2 \to \infty$ as $t \to \infty$. Since the function

$$z \mapsto \big(\rho(\|z-\mu\|)-\rho(\|z\|)\big)\bigg(1+\alpha\frac{u'(z-\mu)}{\|z-\mu\|}\bigg)\xi_{z,\mu}$$

is lower-bounded by the P-integrable function $z \mapsto -2\rho(\|z\|)$ that does not depend on μ, Fatou's Lemma entails that

$$
\begin{aligned}
&\liminf_{\ell\to\infty} \mathcal{J}_1(\alpha_\ell, u_\ell, \mu_\ell)\\
&\quad\geq \int_{\mathbb{R}^d} \liminf_{\ell\to\infty} \big(\rho(\|z-\mu_\ell\|)-\rho(\|z\|)\big)\left(1+\alpha_\ell \frac{u'_\ell(z-\mu_\ell)}{\|z-\mu_\ell\|}\right)\xi_{z,\mu_\ell}\, dP(z)\\
&\quad= \int_{\mathbb{R}^d} \liminf_{\ell\to\infty} \rho(\|z-\mu_\ell\|)\left(1+\alpha_\ell \frac{u'_\ell(z-\mu_\ell)}{\|z-\mu_\ell\|}\right)\xi_{z,\mu_\ell}\, dP(z)\\
&\quad\geq \int_{E} \liminf_{\ell\to\infty} \rho(\|z-\mu_\ell\|)\left(1+\alpha_\ell \frac{u'_\ell(z-\mu_\ell)}{\|z-\mu_\ell\|}\right)\xi_{z,\mu_\ell}\, dP(z)\\
&\quad= \infty
\end{aligned}
$$

(recall that $P[E] > 0$). Observe next that, for any ℓ,

$$
\big|\mathcal{J}_2(\alpha_\ell, u_\ell, \mu_\ell)\big| \leq 2\int_{\mathbb{R}^d} \rho(\|z\|)\, dP(z) < \infty
$$

and

$$
\big|\mathcal{J}_3(\alpha_\ell, u_\ell, \mu_\ell)\big| \leq 2\int_{\mathbb{R}^d} \rho(\|z\|)\, dP(z) < \infty.
$$

Therefore, $\liminf_{\ell\to\infty} M^\rho_{\alpha_\ell,u_\ell}(\mu_\ell) = \infty$, so that $M^\rho_{\alpha_\ell,u_\ell}(\mu_\ell) \to \infty$ as $\ell \to \infty$, which proves the result.

(ii) We assume now that $\rho(t)/t^3$ is bounded away from 0 as $t \to \infty$ (but we do not assume anymore that Condition (b) holds). Write $M^\rho_{\alpha_\ell,u_\ell}(\mu_\ell)/\|\mu_\ell\| = \mathcal{I}_1(\alpha_\ell, u_\ell, \mu_\ell) + \mathcal{I}_2(\alpha_\ell, u_\ell, \mu_\ell) + \mathcal{I}_3(\alpha_\ell, u_\ell, \mu_\ell)$ as in (3.9.12). For any $z \in E$, we have

$$
\begin{aligned}
&\liminf_{\ell\to\infty} \frac{\rho(\|z-\mu_\ell\|)}{\|\mu_\ell\|}\left(1+\alpha_\ell \frac{u'_\ell(z-\mu_\ell)}{\|z-\mu_\ell\|}\right)\\
&\quad\geq \liminf_{\ell\to\infty} \frac{\rho(\|z-\mu_\ell\|)d^2_{u_\ell}(\mu_\ell,z)}{\|\mu_\ell\|\|z-\mu_\ell\|^2\left(1-\alpha_\ell\frac{u'_\ell(z-\mu_\ell)}{\|z-\mu_\ell\|}\right)}\\
&\quad\geq \liminf_{\ell\to\infty} \frac{\rho(\|z-\mu_\ell\|)}{\|\mu_\ell\|\|z-\mu_\ell\|^2} \times \liminf_{\ell\to\infty} \frac{d^2_{u_\ell}(\mu_\ell,z)}{1-\alpha_\ell\frac{u'_\ell(z-\mu_\ell)}{\|z-\mu_\ell\|}}\\
&\quad> 0,
\end{aligned}
$$

since it is now assumed that $\liminf_{t\to\infty} \rho(t)/t^3 > 0$. For the same reason as in the proof of Lemma 3.9.6, we may apply Fatou's Lemma for $\mathcal{I}_1$, which yields

$$
\begin{aligned}
&\liminf_{\ell\to\infty} \mathcal{I}_1(\alpha_\ell, u_\ell, \mu_\ell) \\
&\quad \geq \int_{\mathbb{R}^d} \liminf_{\ell\to\infty} \frac{\rho(\|z-\mu_\ell\|)-\rho(\|z\|)}{\|\mu_\ell\|}\left(1+\alpha_\ell \frac{u_\ell'(z-\mu_\ell)}{\|z-\mu_\ell\|}\right)\xi_{z,\mu_\ell}\, dP(z) \\
&\quad = \int_{\mathbb{R}^d} \liminf_{\ell\to\infty} \frac{\rho(\|z-\mu_\ell\|)}{\|\mu_\ell\|}\left(1+\alpha_\ell \frac{u_\ell'(z-\mu_\ell)}{\|z-\mu_\ell\|}\right)\xi_{z,\mu_\ell}\, dP(z) \\
&\quad \geq \int_{E} \liminf_{\ell\to\infty} \frac{\rho(\|z-\mu_\ell\|)}{\|\mu_\ell\|}\left(1+\alpha_\ell \frac{u_\ell'(z-\mu_\ell)}{\|z-\mu_\ell\|}\right)\xi_{z,\mu_\ell}\, dP(z) \\
&\quad > 0.
\end{aligned}
$$

Since the same arguments as in the proof of Lemma 3.9.6 show that

$$\lim_{\ell\to\infty} \mathcal{I}_2(\alpha_\ell, u_\ell, \mu_\ell) = 0 \quad \text{and} \quad \liminf_{\ell\to\infty} \mathcal{I}_3(\alpha_\ell, u_\ell, \mu_\ell) \geq 0,$$

we obtain that $\liminf_{\ell\to\infty} M^\rho_{\alpha_\ell,u_\ell}(\mu_\ell)/\|\mu_\ell\| > 0$. Therefore, $M^\rho_{\alpha_\ell,u_\ell}(\mu_\ell) \to \infty$ as $\ell \to \infty$, which establishes the result. ■

Proof of Theorem 3.7.2. Let (α_n) be a sequence in $[0,1)$ and (u_n) be a sequence in $\mathcal{S}^{d-1}$. Ad absurdum, assume that there exists a sequence $(\mu^\rho_{\alpha_n,u_n})$ of ρ-quantiles that exits any bounded set. In particular, $(\mu^\rho_{\alpha_n,u_n})$ is unbounded. With $\mu_0 \in \mathbb{R}^d$ fixed arbitrarily, we have

$$M^\rho_{\alpha_n,u_n}(\mu^\rho_{\alpha_n,u_n}) \leq M^\rho_{\alpha_n,u_n}(\mu_0)$$

for any n. From continuity of $(\alpha,u) \mapsto M^\rho_{\alpha,u}(\mu_0)$ and compactness of $[0,1]$ and $\mathcal{S}^{d-1}$, there exists $M > 0$ such that $M^\rho_{\alpha_n,u_n}(\mu_0) \leq M$ for any n, so that

$$M^\rho_{\alpha_n,u_n}(\mu^\rho_{\alpha_n,u_n}) \leq M \tag{3.9.41}$$

for any n. Since the sequence $(\mu^\rho_{\alpha_n,u_n})$ is unbounded, it admits a subsequence $(\mu^\rho_{\alpha_{n_k},u_{n_k}})$ such that $\|\mu^\rho_{\alpha_{n_k},u_{n_k}}\| \to \infty$ as $k \to \infty$. Lemma 3.9.19 then implies that $M^\rho_{\alpha_n,u_n}(\mu^\rho_{\alpha_n,u_n}) \to \infty$ as $k \to \infty$, which contradicts (3.9.41). ■

We now prove Proposition 3.7.3.

Proof of Proposition 3.7.3. Fix $u \in \mathcal{S}^{d-1}$ and let (α_n) be a sequence in $[0,1)$ that converges to one. Let $\mu^\rho_{\alpha_n,u}$ be a sequence of ρ-quantiles and let $\mu_0 \in \mathbb{R}^d$ be fixed. We have

$$M^\rho_{\alpha_n,u}(\mu^\rho_{\alpha_n,u}) \leq M^\rho_{\alpha_n,u}(\mu_0)$$

for any n. Since $(\mu^\rho_{\alpha_n,u})$ is bounded (Theorem 3.7.2), it admits a subsequence $(\mu^\rho_{\alpha_{n_k},u})$ that converges in $\mathbb{R}^d$, to v say. Continuity of $(\alpha,u,\mu) \mapsto M^\rho_{\alpha,u}(\mu)$ on $[0,1] \times \mathcal{S}^{d-1} \times \mathbb{R}^d$ implies that

$$M^\rho_{1,u}(v) \leq M^\rho_{1,u}(\mu_0)$$

for any n. Since this holds for any $\mu_0 \in \mathbb{R}^d$, we conclude that v minimizes $\mu \mapsto M^\rho_{1,u}(\mu)$ over $\mathbb{R}^d$, which establishes the result. ■

Proof of Corollary 3.7.4. Let (α_n) be a sequence in $[0,1)$ that converges to $\alpha \in [0,1]$ and (u_n) be a sequence in $\mathcal{S}^{d-1}$ that converges to u. Let $(\mu^\rho_{\alpha_n,u_n})$ be a sequence of ρ-quantiles and $(\mu^\rho_{\alpha_{n_k},u_{n_k}})$ be a convergent subsequence. Denote as v its limit. For an arbitrarily fixed $\mu_0 \in \mathbb{R}^d$, we then have

$$M^\rho_{\alpha_{n_k},u_{n_k}}(\mu^\rho_{\alpha_{n_k},u_{n_k}}) \leq M^\rho_{\alpha_{n_k},u_{n_k}}(\mu_0)$$

for any k. Continuity of $(\alpha, u, \mu) \mapsto M^\rho_{\alpha,u}(\mu)$ on $[0,1] \times \mathcal{S}^{d-1} \times \mathbb{R}^d$ implies that

$$M^\rho_{\alpha,u}(v) \leq M^\rho_{\alpha,u}(\mu_0)$$

for any n. Since μ_0 is arbitrary, v minimizes $\mu \mapsto M^\rho_{\alpha,u}$, hence is a ρ-quantile of order α in direction u. (ii) Assume that $\mu \mapsto M^\rho_{\alpha,u}$ has a unique minimizer $\mu^\rho_{\alpha,u}$ and let now $(\mu^\rho_{\alpha_{n_k},u_{n_k}})$ be an arbitrary subsequence of $(\mu^\rho_{\alpha_n,u_n})$. This subsequence is bounded (Theorem 3.7.2), hence admits a converging subsequence. Part (i) of the result implies that the corresponding limit must be the unique minimizer $\mu^\rho_{\alpha,u}$ of $\mu \mapsto M^\rho_{\alpha,u}(\mu)$. We thus proved than any subsequence of $(\mu^\rho_{\alpha_n,u_n})$ admits a further subsequence converging to $\mu^\rho_{\alpha,u}$. This entails that $\mu^\rho_{\alpha_n,u_n} \to \mu^\rho_{\alpha,u}$ as $n \to \infty$. ■

Proofs for Section 3.8

Proof of Theorem 3.8.1. The result follows by applying Theorem 1 in [79], as Theorem 3.2.1 and Theorem 3.3.4 show that the conditions (i)–(iii) in page 1515 of that paper are satisfied. ■

We will need Lemmas 3.9.21–3.9.23 below to prove Theorem 3.8.2. The first of these lemmas in turn requires the following preliminary result.

Lemma 3.9.20. *Let $\rho \in \mathcal{C}$. Fix $\alpha \in [0, \alpha_\rho) \cup \{0\}$ and $u \in \mathcal{S}^{d-1}$. Fix $v \in \mathcal{S}^{d-1}$ and $x \in \mathbb{R}^d$ with $\|x\| \in \mathcal{D}_\rho$ such that $v'\nabla^2 H^\rho_{\alpha,u}(x)v = 0$. Then, $|v'x| = \|x\|$ and $\psi'_-(\|x\|) = 0$.*

Proof of Lemma 3.9.20. As seen in the proof of Theorem 3.3.3,

$$\frac{t^2}{\rho(t)} v'\nabla^2 H^\rho_{\alpha,u}(x)v = g_{t,\alpha}(\theta, \omega) = i_t(\omega) + \alpha s_t(\theta, \omega),$$

where $t = \|x\|$, $\cos\theta = u'x/\|x\|$, $\cos\omega = v'x/\|x\|$, and

$$\begin{aligned} i_t(\omega) &= \frac{t\psi_-(t)}{\rho(t)} + \frac{t^2\psi'_-(t) - t\psi_-(t)}{\rho(t)}(\cos\omega)^2 \\ &= \frac{t^2}{\rho(t)}\left(\frac{\psi_-(t)}{t}(1 - (\cos\omega)^2) + \psi'_-(t)(\cos\omega)^2\right) \geq 0 \end{aligned}$$

(we will not need the expression of $s_t(\theta, \omega)$ here). We consider two cases. (a) $\alpha = 0$. Since $v'\nabla^2 H^\rho_{\alpha,u}(x)v = 0$, we have $i_t(\omega) = 0$, which yields

$$\frac{\psi_-(t)}{t}(1 - (\cos\omega)^2) = 0 \quad \text{and} \quad \psi'_-(t)(\cos\omega)^2 = 0.$$

Since $\psi_-(t) \geq \rho(t)/t > 0$ for any $t > 0$ (Lemma 3.9.4), we must then have $(\cos\omega)^2 = 1$ and $\psi'_-(t) = 0$, which shows that the result holds in this case. (b) $\alpha > 0$ (so that $0 < \alpha < \alpha_\rho$). Since $v'\nabla^2 H^\rho_{\alpha,u}(x)v = 0$, we have $g_{t,\alpha}(\theta, \omega) = 0$. Ad absurdum, assume then that $(\cos\omega)^2 \neq 1$ or $\psi'_-(t) > 0$. Then, $i_t(\omega) > 0$, so that we must have $s_t(\theta, \omega) < 0$. Therefore, any $\alpha_0 > \alpha$ will provide $g_{t,\alpha_0}(\theta, \omega) < 0$, which implies $\alpha_0 > \alpha_\rho$. We thus proved that any $\alpha_0 > \alpha$ satisfies $\alpha_0 > \alpha_\rho$, which contradicts the assumption that $\alpha < \alpha_\rho$. Therefore, $(\cos\omega)^2 = 1$ and $\psi'_-(t) = 0$ in case (b), too. ■

Lemma 3.9.21. *Let the assumptions of Theorem 3.5.3 hold and assume further that* $P[E_{\mu_0,v}] < 1$ *for any* $v \in \mathcal{S}^{d-1}$, *with*

$$E_{\mu_0,v} := \{z \in \mathbb{R}^d : \psi'_-(\|z - \mu_0\|) = 0 \text{ and } z \in L_{\mu_0}(v)\},$$

where $L_{\mu_0}(v) = \{\mu_0 + \lambda v : \lambda \in \mathbb{R}\}$ *is the line through* μ_0 *with direction* v. *Then, for any* $\alpha \in [0, \alpha_\rho) \cup \{0\}$ *and* $u \in \mathcal{S}^{d-1}$, *the Hessian matrix* $\nabla^2 M^\rho_{\alpha,u}(\mu_0)$ *is positive definite.*

Proof of Lemma 3.9.21. Fix $\alpha \in [0, \alpha_\rho) \cup \{0\}$ and $u \in \mathcal{S}^{d-1}$. Under the assumptions considered, it follows from (3.9.38) that

$$\nabla^2 M^\rho_{\alpha,u}(\mu_0) = \int_G \nabla^2 H^\rho_{\alpha,u}(z - \mu_0)\, dP(z), \tag{3.9.42}$$

where $G := \{z \in \mathbb{R}^d : \|z - \mu_0\| \notin \mathcal{D}_\rho\}$ has probability one (so that $\nabla^2 H^\rho_{\alpha,u}(z - \mu_0)$ may be defined arbitrarily for $z \notin G$). Assume then, ad absurdum, that $v'\nabla^2 M^\rho_{\alpha,u}(\mu_0)v = 0$ for some $v \in \mathcal{S}^{d-1}$. It directly follows from (3.9.42) and convexity of $H^\rho_{\alpha,u}$ that $v'\nabla^2 H^\rho_{\alpha,u}(z - \mu_0)v = 0$ for any $z \in G$. Lemma 3.9.20 thus implies that $\psi'_-(\|z - \mu_0\|) = 0$ and $z \in L_{\mu_0}$ for P-almost all $z \in \mathbb{R}^d$, a contradiction. ■

Lemma 3.9.22. *Let* $\rho \in \mathcal{C}$ *and* $P \in \mathcal{P}^\rho_d$. *Fix* $\alpha \in [0, 1)$, $u \in \mathcal{S}^{d-1}$, *and* $\mu \in \mathbb{R}^d$. *Assume that*

$$\int_{\mathbb{R}^d} \psi^2_-(\|z - \mu\|)\, dP(z) < \infty.$$

Let Z *be a random* d*-vector with distribution* P. *Then, the* $d \times d$ *matrix*

$$\mathrm{E}[(\nabla H^\rho_{\alpha,u}(Z - \mu))(\nabla H^\rho_{\alpha,u}(Z - \mu))'\mathbb{I}[\|Z - \mu\| \in \mathcal{D}_\rho]]$$

exists and is finite.

Proof of Lemma 3.9.22. Fix $r \in \{1, \ldots, d\}$. Lemma 3.9.13(ii) readily entails that, if $\|z - \mu\| \in \mathcal{D}_\rho$, then

$$(\partial_r H^\rho_{\alpha,u}(z - \mu))^2 \leq C\bigg(\psi^2_-(\|z - \mu\|) + \frac{\rho^2(\|z - \mu\|)}{\|z - \mu\|^2}\bigg)$$

for some positive constant C. Since $\rho(t)/t \leq \psi_-(t)$ for any $t > 0$ (Lemma 3.9.4), this implies that

$$\mathrm{E}[(\partial_r H^\rho_{\alpha,u}(Z-\mu))^2 \mathbb{I}[\|Z-\mu\| \in \mathcal{D}_\rho]] < \infty.$$

Since this hols for any $r \in \{1,\ldots,d\}$, the result follows from the Cauchy–Schwarz inequality. ■

Lemma 3.9.23. *Let the assumptions of Theorem 3.8.2 hold and write $F = \{z \in \mathbb{R}^d : \|z - \mu^\rho_{\alpha,\mu}\| \in \mathcal{D}_\rho\}$. Then, for any $h \in \mathbb{R}^d$,*

$$\begin{aligned} R_n := \sum_{i=1}^n \big\{H^\rho_{\alpha,u}\big(Z_i - \mu^\rho_{\alpha,u} - \tfrac{h}{\sqrt{n}}\big) - H^\rho_{\alpha,u}(Z_i - \mu^\rho_{\alpha,u})\big\}\mathbb{I}[Z_i \in F] \\ + \frac{1}{\sqrt{n}} \sum_{i=1}^n h' \nabla H^\rho_{\alpha,u}(Z_i - \mu^\rho_{\alpha,u}) \mathbb{I}[Z_i \in F] \\ - \frac{1}{2n} \sum_{i=1}^n h' \nabla^2 H^\rho_{\alpha,u}(Z_i - \mu^\rho_{\alpha,u}) h \mathbb{I}[Z_i \in F] \end{aligned}$$

converges to 0 *in the $L_1(P)$ sense as $n \to \infty$.*

Proof of Lemma 3.9.23. For $z \in \mathbb{R}^d$, let

$$J_n(z) = n\big\{H^\rho_{\alpha,u}\big(z - \mu^\rho_{\alpha,u} - \tfrac{h}{\sqrt{n}}\big) - H^\rho_{\alpha,u}(z - \mu^\rho_{\alpha,u}) + \tfrac{h'}{\sqrt{n}} \nabla H^\rho_{\alpha,u}(z - \mu^\rho_{\alpha,u})\big\}\mathbb{I}[z \in F]$$

and $J(z) = \frac{1}{2} h' \nabla^2 H^\rho_{\alpha,u}(z - \mu^\rho_{\alpha,u}) h \mathbb{I}[z \in F]$. First note that $J_n(z) \to J(z)$ as $n \to \infty$ for any $z \in \mathbb{R}^d$ (this is trivial for $z \notin F$ and results from Lemma 3.9.13(ii) for $z \in F$). Since $P[F] = 1$ by assumption, observe that (3.9.36) and (3.9.38) yield

$$\begin{aligned} E[J_n(Z_1)] - E[J(Z_1)] = n\Big\{ M^\rho_{\alpha,u}\big(\mu^\rho_{\alpha,u} + \tfrac{h}{\sqrt{n}}\big) - M^\rho_{\alpha,u}(\mu^\rho_{\alpha,u}) \\ - \tfrac{1}{\sqrt{n}} h' \nabla M^\rho_{\alpha,u}(\mu^\rho_{\alpha,u}) - \tfrac{1}{2n} h' \nabla^2 M^\rho_{\alpha,u}(\mu^\rho_{\alpha,u}) h \Big\}. \end{aligned}$$

Since $M^\rho_{\alpha,u}$ is twice differentiable at $\mu^\rho_{\alpha,u}$ (Theorem 3.5.3), it follows that $E[J_n(Z_1)] \to E[J(Z_1)]$ as $n \to \infty$. Now, recalling that $\alpha \in [0, \alpha_\rho) \cup \{0\}$, the map $z \mapsto H^\rho_{\alpha,u}(z)$ is convex, which implies that $J_n(z)$ and $J(z)$ are nonnegative for any $z \in \mathbb{R}^d$. Therefore, Scheffé's lemma entails that $J_n(Z_1) \to J(Z_1)$ in the $L_1(P)$ sense as $n \to \infty$, so that $E[|R_n|] \leq E[|J_n(Z_1) - J(Z_1)|] = o(1)$ as $n \to \infty$. ■

Proof of Theorem 3.8.2. Throughout the proof, $M^{\rho,P_n}_{\alpha,u}(\mu)$ stands for the objective function $M^\rho_{\alpha,u}$ associated to the empirical probability measure P_n, that is

$$M^{\rho,P_n}_{\alpha,u}(\mu) := \frac{1}{n} \sum_{i=1}^n \Big(H^\rho_{\alpha,u}(Z_i - \mu) - H^\rho_{\alpha,u}(Z_i) \Big),$$

for all $\mu \in \mathbb{R}^d$. We prove the result by applying Theorem 3 from [2] with the sequence of stochastic processes $\{G_n(\theta) = n M^{\rho,P_n}_{\alpha,u}(\mu) : \theta = \mu \in \Theta = \mathbb{R}^d\}$, the fixed parameter value $\theta_0 = \mu^\rho_{\alpha,u}$, and the sequence of estimators $\hat\theta_n = \hat\mu^\rho_{\alpha,u}$. We now check that under the assumption of Theorem 3.8.2, Conditions (i)–(v) from Arcones' Theorem hold with

$$\eta_n := -\frac{1}{\sqrt{n}} \sum_{i=1}^n \nabla H^\rho_{\alpha,u}(Z_i - \mu^\rho_{\alpha,u}) \mathbb{I}[Z_i \in F]$$

(as in Lemma 3.9.23, $F = \{z \in \mathbb{R}^d : \|z - \mu^\rho_{\alpha,\mu}\| \in \mathcal{D}_\rho\}$), $V_n := \frac{1}{2}A = \frac{1}{2}\nabla^2 M^\rho_{\alpha,u}(\mu^\rho_{\alpha,u})$, and $M_n = \sqrt{n} I_d$. The restriction to $\alpha \in [0, \alpha_\rho) \cup \{0\}$ ensures that the convexity requirement in Condition (i) holds, whereas Condition (ii) directly follows from the fact that, by definition, $\hat\mu^\rho_{\alpha,u}$ minimizes $\mu \mapsto nM^{\rho,P_n}_{\alpha,u}(\mu)$. We have $P[F] = 1$ by assumption, so that (3.9.36) yields

$$\mathrm{E}[\nabla H^\rho_{\alpha,u}(Z_i - \mu^\rho_{\alpha,u}) \mathbb{I}[Z_i \in F]] = -\nabla M^\rho_{\alpha,u}(\mu^\rho_{\alpha,u}) = 0$$

for any $i = 1, \ldots, n$. In view of Lemma 3.9.22, the multivariate central limit theorem thus entails that, under the assumptions considered, η_n is asymptotically d-variate normal with mean vector zero and covariance matrix B. Consequently, Condition (iv) holds. As for Condition (v), it trivially follows from the fact that $V_n = \frac{1}{2}\nabla^2 M^\rho_{\alpha,u}(\mu^\rho_{\alpha,u})$ is positive definite (Lemma 3.9.21) and does not depend on n.

Therefore, it only remains to show that Condition (iii) holds, that is, that, for each $h \in \mathbb{R}^d$,

$$nM^{\rho,P_n}_{\alpha,u}(\mu^\rho_{\alpha,u} + \tfrac{h}{\sqrt{n}}) - nM^{\rho,P_n}_{\alpha,u}(\mu^\rho_{\alpha,u}) - h'\eta_n - h'V_n h = o_{\mathrm{P}}(1) \qquad (3.9.43)$$

as $n \to \infty$. In order to do so, write

$$\begin{aligned} &nM^{\rho,P_n}_{\alpha,u}(\mu^\rho_{\alpha,u} + \tfrac{h}{\sqrt{n}}) - nM^{\rho,P_n}_{\alpha,u}(\mu^\rho_{\alpha,u}) \\ &\quad = \sum_{i=1}^n \big\{ H^\rho_{\alpha,u}\big(Z_i - \mu^\rho_{\alpha,u} - \tfrac{h}{\sqrt{n}}\big) - H^\rho_{\alpha,u}(Z_i - \mu^\rho_{\alpha,u}) \big\} \mathbb{I}[Z_i \notin F] \\ &\qquad + \sum_{i=1}^n \big\{ H^\rho_{\alpha,u}\big(Z_i - \mu^\rho_{\alpha,u} - \tfrac{h}{\sqrt{n}}\big) - H^\rho_{\alpha,u}(Z_i - \mu^\rho_{\alpha,u}) \big\} \mathbb{I}[Z_i \in F]. \end{aligned}$$

Note that

$$\sum_{i=1}^n \big\{ H^\rho_{\alpha,u}\big(Z_i - \mu^\rho_{\alpha,u} - \tfrac{h}{\sqrt{n}}\big) - H^\rho_{\alpha,u}(Z_i - \mu^\rho_{\alpha,u}) \big\} \mathbb{I}[Z_i \notin F] = o_{\mathrm{P}}(1)$$

as $n \to \infty$, since $P[F] = 1$ implies that, for any $\varepsilon > 0$,

$$\begin{aligned} &P\Big[\Big| \sum_{i=1}^n \big\{ H^\rho_{\alpha,u}\big(Z_i - \mu^\rho_{\alpha,u} - \tfrac{h}{\sqrt{n}}\big) - H^\rho_{\alpha,u}(Z_i - \mu^\rho_{\alpha,u}) \big\} \mathbb{I}[Z_i \notin F] \Big| > \varepsilon \Big] \\ &\qquad \le P\Big[\bigcup_{i=1}^n [Z_i \notin F] \Big] \le \sum_{i=1}^n P[\mathbb{R}^d \setminus F] = 0. \end{aligned}$$

Therefore,

$$
\begin{aligned}
& nM_{\alpha,u}^{\rho,P_n}(\mu_{\alpha,u}^{\rho}+\tfrac{h}{\sqrt{n}}) - nM_{\alpha,u}^{\rho,P_n}(\mu_{\alpha,u}^{\rho}) - h'\eta_n - h'V_n h \\
&\quad = \sum_{i=1}^{n}\big\{H_{\alpha,u}^{\rho}\big(Z_i-\mu_{\alpha,u}^{\rho}-\tfrac{h}{\sqrt{n}}\big) - H_{\alpha,u}^{\rho}(Z_i-\mu_{\alpha,u}^{\rho})\big\}\mathbb{I}[Z_i\in F] \\
&\quad\quad -h'\eta_n - \frac{1}{2n}\sum_{i=1}^{n} h'(\nabla^2 H_{\alpha,u}^{\rho}(Z_i-\mu_{\alpha,u}^{\rho})h\mathbb{I}[Z_i\in F] \\
&\quad\quad -\frac{1}{2}h'\left(A-\frac{1}{n}\sum_{i=1}^{n}\nabla^2 H_{\alpha,u}^{\rho}(Z_i-\mu_{\alpha,u}^{\rho})\mathbb{I}[Z_i\in F]\right)h + o_{\mathrm{P}}(1).
\end{aligned}
$$

Applying Lemma 3.9.23 and the law of large numbers thus establishes (3.9.43), which shows that Arcones' Condition (iii) holds, too.

Theorem 3 from [2] therefore applies and yields that, as $n\to\infty$,

$$
\sqrt{n}(\hat{\mu}_{\alpha,u}^{\rho}-\mu_{\alpha,u}^{\rho}) = -\frac{1}{2}V_n^{-1}\eta_n + o_{\mathrm{P}}(1) = -A^{-1}\eta_n + o_{\mathrm{P}}(1),
$$

which is the desired result. The asymptotic normal distribution of $\sqrt{n}(\hat{\mu}_{\alpha,u}^{\rho}-\mu_{\alpha,u}^{\rho})$ then readily follows from the fact that, as already mentioned, η_n is asymptotically normal with mean vector zero and covariance matrix B. ■

Chapter 4

Geometric quantiles on the hypersphere

4.1 Introduction

For several decades, an intense research activity has been dedicated to the definition of a suitable *multivariate quantile* concept, that is, of a quantile concept for probability measures over $\mathbb{R}^d$, with $d > 1$. It is of course the lack of a canonical ordering in multivariate Euclidean spaces that makes this a challenging problem. We refer, e.g., to the review paper [88] and to more recent approaches based on quantile regression ([46]) or optimal transport ([21], [44]). Clearly, the problem of defining multivariate quantiles is closely linked to the problems of defining *multivariate depths* (i.e., centrality measures) or *multivariate ranks*. The various proposals in the literature have found key applications in the context of multidimensional growth charts (see, e.g., [104] or [69]), or, more generally, in situations where multiple-output quantile regression methods are relevant.

Despite some recent contributions to the field, one of the most successful multivariate quantile concepts remains the concept of *geometric quantiles* from [17], that is defined as follows. For a probability measure P over $\mathbb{R}^d$, the geometric quantile of order α in direction u for P is defined as an arbitrary minimizer of

$$O^P_{\alpha,u}(\mu) := \int_{\mathbb{R}^d} \left\{ \|z-\mu\| - \|z\| - \alpha u'\mu \right\} dP(z); \tag{4.1.1}$$

here, $\alpha \in [0,1)$, u is a unit d-vector, and $\|z\| = \sqrt{z'z}$ is the Euclidean norm of $z \in \mathbb{R}^d$. For $\alpha = 0$ (and an arbitrary u), this provides the celebrated *geometric median* (see, e.g., [9]). The other geometric quantiles are of a directional, center-outward, nature: the larger α is, the further away the corresponding (α, u)-quantile is, essentially in direction u, from the geometric median. We will use the terminology *geometric quantiles* in the rest of this chapter, but these quantiles are sometimes rather referred to as *spatial quantiles*; see, among others, [20], [13], [41], or [12]. Like classical univariate quantiles (to which they reduce in dimension $d = 1$), geometric quantiles characterize the underlying probability measure P; see [50]. While several approaches also satisfy this characterization property, geometric quantiles enjoy a number of distinctive advantages:

(i) through convex optimization, sample geometric quantiles are easy to compute even in high dimensions. (ii) They allow for detailed asymptotic results, including Bahadur representation and asymptotic normality results (see, e.g., [50] and [17]), whereas some recent competing approaches offer at best consistency results only. (iii) Similarly, *geometric ranks* and *geometric depth*, namely the concepts of multivariate ranks and depth associated with geometric quantiles, are available in explicit forms, which, unlike for most (if not all) competing concepts, leads to trivial evaluation in the sample case. (iv) Finally, geometric quantiles allow for direct extensions in infinite-dimensional Hilbert spaces, also in the regression framework; see, e.g., [15] and [22].

Now, more and more frequently, statistical applications involve data on manifolds. Historically, this has led researchers to extend to manifolds classical Euclidean functionals, the prototypical example being the extension of the Euclidean concept of mean into the concept of *Fréchet mean* ([35]). Nowadays, more involved statistical techniques, such as functional data analysis, are also considered on manifolds, with the primary focus being often on the unit hypersphere $\mathcal{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$ of $\mathbb{R}^d$; see, e.g., [25]. The present work aims at defining a concept of quantiles on $\mathcal{S}^{d-1}$. As such, it therefore belongs to *directional statistics*[1]; we refer to the monographs [67] and [56]. Several concepts of depth have been proposed on the unit sphere ([63], [1], [84]), and, quite interestingly, also on general metric spaces ([24])! We further refer to [48] for a recent work where transformations related to *distribution functions*, which are of course linked to quantile functions, are considered on Riemannian manifolds; see also [103]. Yet the only quantile concept that was explicitly proposed on the hypersphere and investigated as such is the quantile concept from [55], that, however, lacks flexibility and also does not characterize the underlying distribution (as we will explain in the sequel).

The concept of spherical quantiles we propose requires the choice of a reference point, that is expected to play the role of the innermost quantile. Since, in Euclidean cases, the innermost quantile is a multivariate median, we choose this reference point as a Fréchet median, that is, as the L_1-analog of a Fréchet mean. Since these medians may be non-unique, we will actually define a quantile field around each such median m. Like Euclidean geometric quantiles, the resulting quantiles $\mu^m_{\alpha,u}$ will be directional in nature: they are indexed by a scalar order $\alpha \in [0, 1)$ and a unit vector u that here belongs to the tangent space $T_m\mathcal{S}^{d-1}$ to $\mathcal{S}^{d-1}$ at m. Motivated by the nice properties of geometric quantiles in the Euclidean case, we essentially define our quantiles through a stereographic projection of $\mathcal{S}^{d-1}$ from $-m$ (the antipodal point to m) onto $T_m\mathcal{S}^{d-1}$. Since this stereographic projection sends $-m$ "isotropically at infinity" in $T_m\mathcal{S}^{d-1}$, studying the proposed spherical quantiles requires understanding the nature of the [17] geometric quantiles in a version of the projective space where all points at infinity are identified. We thoroughly investigate the structural properties of our quantiles and we further study the asymptotic behaviour of their sample versions, which requires controlling the impact of estimating m. Our spherical quantile concept also allows for companion concepts of ranks and depth on the hypersphere. We show the relevance of our construction by basing tests for rotational symmetry on the proposed quantiles.

The outline of this chapter is as follows. In Section 4.2, we first discuss the "univari-

[1] In order to avoid any confusion, we will rather speak of *spherical statistics* in the sequel and will use the term *"directional"* only to refer to the directional (in u) nature of the various quantiles we consider.

ate" case $\mathcal{S}^1$ and explain what makes the stereographic construction natural to define quantiles on $\mathcal{S}^{d-1}$ for $d > 2$. Then, we carefully define our spherical quantiles and present a result that justifies how this definition treats $-m$, the point that is left aside in the stereographic projection we consider. In Section 4.3.1, (i) we show that our innermost geometric quantile actually agrees with the Fréchet median chosen as the reference point, (ii) we discuss existence and uniqueness of the proposed quantiles, and (iii) we provide a result that characterizes the behaviour of extreme quantiles (i.e., the quantiles obtained as $\alpha \to 1$) in each direction u. This allows us to formally define the spherical quantile function in Section 4.3.2. In Section 4.4, we introduce a spherical rank function that, under mild conditions, is the inverse map of the quantile function. This rank function, that is intimately related to the gradient condition defining our quantiles, allows us to show in particular that the quantile function characterizes the underlying probability measure. In Section 4.5, we define a companion concept of depth and discuss its main properties. In Section 4.6, we focus on the asymptotic properties of the sample version of our quantiles: we first establish strong consistency and asymptotic normality of the Fréchet median, then prove asymptotic normality results of all other sample quantiles (in both cases, explicit Bahadur representation results are actually provided). In Section 4.7, we introduce tests of rotational symmetry based on the proposed spherical concepts.

For the sake of convenience, we introduce here some notation that will be used throughout this chapter. First, $X =_d Y$ will mean that X and Y are equal in distribution. For $\mu \in \mathcal{S}^{d-1}$, we will denote as $\mathcal{S}^{d-1}_\mu := \mathcal{S}^{d-1} \setminus \{\mu\}$ the unit sphere deprived of μ and as $T_\mu\mathcal{S}^{d-1}$ the $(d-1)$-dimensional vector subspace of $\mathbb{R}^d$ that is parallel to the tangent hyperplane to $\mathcal{S}^{d-1}$ at μ, that is, $T_\mu\mathcal{S}^{d-1} = \{z \in \mathbb{R}^d : \mu' z = 0\}$. We will write $\mathcal{P}_{d-1}$ for the collection of all probability measures on $\mathcal{S}^{d-1}$. We will denote as $\mathbb{I}[A]$ the indicator function of the set or condition A. Throughout, $\mathrm{E}[\cdot]$ will refer to the usual expectation rather to the Fréchet mean. The d-dimensional identity matrix will be denoted as I_d. By default, all vectors will be column vectors; yet, to keep the notation light, we will often skip transpose signs when writing vectors in components—for instance, we will write $(\cos t, \sin t)$ and $(0, 0, 1)$ instead of $(\cos t, \sin t)'$ and $(0, 0, 1)'$, respectively.

4.2 Spherical geometric quantiles

In this section, we will define our concept of quantiles on the unit hypersphere $\mathcal{S}^{d-1}$ and justify the choices made in this definition. We start by discussing the circular case $d = 2$, that is, the case of the unit circle $\mathcal{S}^1 = \{(\cos t, \sin t) : t \in [0, 2\pi)\}$, then turn to the general case $d \geq 2$.

4.2.1 Circular quantiles

Fix $P \in \mathcal{P}_1$ and let the random variable T, with values in $[0, 2\pi)$, be such that $X := (\cos T, \sin T)$ has distribution P. Since the circle is a one-dimensional object, quantiles on the circle can in principle be defined from quantiles on the real line, that is, quantiles of X can in principle be defined from quantiles of T. Yet, interestingly, the circle already presents several key issues we will need to address in higher dimensions. An important issue is the lack of a canonical reference point $m = (\cos t_m, \sin t_m)$ on the circle. For any such reference point, one could, e.g., accumulate probability mass above t_m, leading to the circular quantiles $\mu_\tau^m = (\cos q_\tau^m, \sin q_\tau^m)$, $\tau \in [0,1]$, with q_τ^m the usual τ-quantile of T_m, where T_m is the random variable with values in $[t_m, t_m + 2\pi)$ such that $X =_d (\cos T_m, \sin T_m)$. This, however, cannot be generalized to higher dimensions where it is unclear what it means to accumulate probability mass "above" some reference point on $\mathcal{S}^{d-1}$ with $d > 2$. With this future extension to higher dimensions in mind, it is therefore better to choose a reference point m that will play the role of the innermost quantile, namely the median. In this spirit, a natural candidate for a reference point of this type is a *Fréchet median* $m = (\cos t_m, \sin t_m)$, that minimizes the expected arc length between X and m; see Definition 4.2.1 below. Parallel as above, one may then define the resulting circular quantiles as $\mu_\tau^m = (\cos q_\tau^m, \sin q_\tau^m)$, $\tau \in [0,1]$, now with q_τ^m the τ-quantile of the random variable $\tilde{T}_m$, with values in $[t_m - \pi, t_m + \pi)$, such that $X =_d (\cos \tilde{T}_m, \sin \tilde{T}_m)$. Provided that $P[\{-m\}] = 0$, the resulting *circular median* $\mu_{1/2}^m$ then very naturally coincides with the Fréchet median m that was used as a reference point.

These circular quantiles using a Fréchet median as a reference point are clearly satisfactory when (a) P admits a unique Fréchet median m and when (b) $P[\{-m\}] = 0$, that is, when $\tilde{T}_m$ does not charge $t_m - \pi$. The issue (a) is a structural one on the circle: unlike for distributions on the real line, where the medians (i.e., the minimizers of expected absolute deviations) always form an interval, so that a unique median can always be identified (e.g., as the centre of this interval), the topology of the circle allows for distributions with sets of Fréchet medians that are disconnected (for instance, when $X = (\cos T, \sin T)$, where T is uniform over $\cup_{k=1}^3 [(2k-1)\pi/3 - \pi/6, (2k-1)\pi/3 + \pi/6]$, the set of Fréchet medians is $\cup_{k=1}^3 \{(2k-1)\pi/3\}$). We argue that any Fréchet median then provides a perfectly valid reference point (hence, a perfectly valid innermost quantile), which in turn will provide its own corresponding collection of circular quantiles. To address issue (b), it is natural to define $\mu_\tau^m = (\cos q_\tau^m, \sin q_\tau^m)$, where q_τ^m is the τ-quantile of $\tilde{T}_m$, where $\tilde{T}_m$ is still such that $X =_d (\cos \tilde{T}_m, \sin \tilde{T}_m)$ but now takes values in $[t_m - \pi, t_m + \pi]$ and satisfies $P[\tilde{T}_m = t_m - \pi] = P[\tilde{T}_m = t_m + \pi] = P[\{-m\}]/2$. Not only does this choice respect symmetry of the circle with respect to m but it also guarantees that the resulting circular median $\mu_{1/2}^m$ coincides with the Fréchet median m even when $P[\{-m\}] > 0$. The issue of such an atom at $-m$ will also need to be carefully dealt with in higher dimensions.

4.2.2 Hyperspherical quantiles

Parallel to the multivariate Euclidean case described in the introduction, our hyperspherical quantiles will be points of $\mathcal{S}^{d-1}$ indexed by a scalar magnitude α and a direction u. In the sequel, this direction u will be relative to *a reference point* m, that is expected to play the role of the median (the innermost quantile). This is in line with the Euclidean case, where geometric quantiles are thought to be in direction u *from the geometric median* (although such localization is actually superfluous in $\mathbb{R}^d$ as, for fixed u, the halflines $\{\mu + ru : r \geq 0\}$ there "reach" the same point at infinity irrespective of their origin μ). As a reference point on the sphere, we will use a *Fréchet median.*

Definition 4.2.1. *A* Fréchet median *of $P(\in \mathcal{P}_{d-1})$ is any point $m \in \mathcal{S}^{d-1}$ that minimizes the objective function*

$$\mu \mapsto g_P(\mu) := \int_{\mathcal{S}^{d-1}} d(\mu, x)\, dP(x) \tag{4.2.2}$$

over $\mathcal{S}^{d-1}$, where $d(x,y) = \arccos(x'y)$ is the geodesic distance between x and y. The collection of all Fréchet medians of P will be called the Fréchet median set *of P.*

Lebesgue's dominated convergence theorem ensures that g_P is continuous over $\mathcal{S}^{d-1}$, so that, from compactness of $\mathcal{S}^{d-1}$, any $P \in \mathcal{P}_{d-1}$ admits at least one Fréchet median. As already mentioned when discussing the case $d = 2$, however, uniqueness is not guaranteed in general; see also [105]. We will show later that we must have $P[\{m\}] \geq P[\{-m\}]$ for any Fréchet median m (this actually follows from the gradient condition associated with the minimization problem defining m; see Lemma 4.8.6).

Fix then a Fréchet median m. The discussion at the beginning of this section suggests that spherical quantiles with respect to m should be defined in direction u from m, which motivates taking u as a unit vector in the "tangent" vector space $T_m\mathcal{S}^{d-1}$ to $\mathcal{S}^{d-1}$ at m. Consider then the stereographic projection of $\mathcal{S}^{d-1}$ from $-m$ onto $T_m\mathcal{S}^{d-1}$, namely the diffeomorphic transformation

$$\pi_m : \mathcal{S}^{d-1}_{-m} \to T_m\mathcal{S}^{d-1} : x \mapsto \pi_m(x) := \frac{x - (m'x)m}{1 + m'x} \tag{4.2.3}$$

and let P_{-m} be the probability measure induced by P on $\mathcal{S}^{d-1}_{-m}$, that is, the probability measure defined by $P_{-m}[B] = P[B]/P[\mathcal{S}^{d-1}_{-m}]$ for any Borel set B of $\mathcal{S}^{d-1}_{-m}$ (note that $P[\mathcal{S}^{d-1}_{-m}] = 0$ is excluded, as it would imply that $-m$ is the only Fréchet median of P). In a nutshell, our spherical quantiles are defined by first considering the (Euclidean) geometric quantiles in $\mathbb{R}^d$ with respect to the pushforward image $\pi_m\#P_{-m}$ of P_{-m} by the projection π_m, and then by pulling the resulting quantiles back onto $\mathcal{S}^{d-1}_{-m}$ through π_m^{-1}. More precisely, we adopt the following definition, that also takes into account a possible atom at $-m$.

Definition 4.2.2. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 2$, and let m be a Fréchet median of P. Fix $\alpha \in [0,1]$ and a unit vector u in $T_m\mathcal{S}^{d-1}$. (i) For $\alpha \in [0, p_m)$, with $p_m := 1 - P[\{-m\}]$, we say that $\mu^m_{\alpha,u} = \mu^m_{\alpha,u}(P)$ is an* m-geometric quantile of order α in direction u for P *if and only if it minimizes the objective function*

$$\begin{aligned}\mu \mapsto M^{m,P}_{\alpha,u}(\mu) &:= O^{\pi_m\#P_{-m}}_{\alpha/p_m,u}(\pi_m(\mu)) \\ &= \int_{\mathcal{S}^{d-1}_{-m}} \big\{\|\pi_m(x) - \pi_m(\mu)\| - \|\pi_m(x)\| - \alpha u'\pi_m(\mu)/p_m\big\}\, dP_{-m}(x)\end{aligned}$$

over $\mathcal{S}^{d-1}_{-m}$; the m-geometric quantiles of P associated to an order $\alpha = 0$ are called m-geometric medians *of P. (ii) For $\alpha \in [p_m, 1]$, we let $\mu^m_{\alpha,u} = \mu^m_{\alpha,u}(P) = -m$.*

Some comments are in order. Assume first that P does not charge $-m$, so that $p_m = 1$. Then Definition 4.2.2 implies that $\mu^m_{\alpha,u}(\in \mathcal{S}^{d-1}_{-m})$ is an m-geometric quantile of order $\alpha(< 1)$ in direction u for P if and only if $\pi_m(\mu^m_{\alpha,u})$ is a geometric quantile of order α

in direction u for the push-forward probability measure $\pi_m\#P$ in $\mathbb{R}^d$ (formally, this will be a corollary of Lemma 4.8.1). Since π_m is a one-to-one map from $\mathcal{S}^{d-1}_{-m}$ to $T_m\mathcal{S}^{d-1}$, the spherical quantile $\mu^m_{\alpha,u}$ is then the inverse image by π_m of the corresponding quantile for $\pi_m\#P$. While this makes Definition 4.2.2 natural when P does not charge $-m$, the definition may seem more opaque when $-m$ is an atom of P. This motivates the following result, that will explain why our concept is as natural in the latter case as in the former one. To state the result, we recall that a probability measure P over $\mathcal{S}^{d-1}$ is said to be rotationally symmetric about $\mu(\in \mathcal{S}^{d-1})$ if and only if $O\#P = O$ for any $d \times d$ orthogonal matrix such that $O\mu = \mu$.

Theorem 4.2.3. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and let m be a Fréchet median of P with $P[\{-m\}] > 0$. Fix $\alpha \in [0,1]$ and a unit vector u in $T_m\mathcal{S}^{d-1}$. Assume that P is not concentrated on a great circle containing m (which ensures existence and uniqueness of $\mu^m_{\alpha,u}(P)$; see Theorem 4.3.1 below). Let (Q_ℓ) be a sequence in $\mathcal{P}_{d-1}$ such that*

(i) Q_ℓ is rotationally symmetric about m,

(ii) there exists $c > 0$ such that $Q_\ell[\{x : d(m,x) < c\}] = 0$ for any ℓ,

(iii) $Q_\ell(\{-m\}) = 0$, and

(iv) Q_ℓ converges weakly to the Dirac probability measure at $-m$.

Then, letting $P_\ell := p_m P_{-m} + (1-p_m)Q_\ell$, still with $p_m = 1 - P[\{-m\}](< 1)$,

$$\mu^m_{\alpha,u}(P_\ell) \to \mu^m_{\alpha,u}(P) \tag{4.2.4}$$

as ℓ diverges to infinity.

As already mentioned, Definition 4.2.2 above is most natural for probability measures that do not charge $-m$, so that the quantiles $\mu^m_{\alpha,u}(P_\ell)$ in this result are the natural ones. Since the sequence of probability measures (P_ℓ) is defined in such a way that it converges weakly to P (which on the contrary charges $-m$), the "continuity" result in (4.2.4) actually justifies Definition 4.2.2 when $P[\{-m\}] > 0$ (equivalently, when $p_m < 1$). Note that this validates this definition both for $\alpha \in [0, p_m)$ and $\alpha \in [p_m, 1]$, as all values of α are covered in Theorem 4.2.3.

No stereographic projection was used in the definition of the circular quantiles in Section 4.2.1. Yet it is easy to check that the quantiles $\mu^m_{\alpha,u}$ from Definition 4.2.2 reduce for $d = 2$ to the circular quantiles defined at the end of Section 4.2.1, provided of course that the latter are reparametrized according to the center-outward indexing adopted in the present subsection. To be more precise, fix a Fréchet median $m = (\cos t_m, \sin t_m)$ on the unit circle and consider the random variable $\tilde{T}_m$ with values in $[t_m - \pi, t_m + \pi]$ such that $(\cos \tilde{T}_m, \sin \tilde{T}_m)$ has distribution P and such that $P[\tilde{T}_m = t_m - \pi] = P[\tilde{T}_m = t_m + \pi] = P[\{-m\}]/2$. Then, for any $\alpha \in [0,1]$ and any unit vector u in $T_m\mathcal{S}^1$, the quantile $\mu^m_{\alpha,u}$ from Definition 4.2.2 coincides with $(\cos q^m_\tau, \sin q^m_\tau)$, where q^m_τ is the $\tau = (\alpha s_u + 1)/2$-quantile of $\tilde{T}_m$; here, $s_u = 1$ (resp., $s_u = -1$) if u indicates the counterclockwise (resp., clockwise) direction on $\mathcal{S}^1$. In particular, both for Definition 4.2.2 and for the definition of circular quantiles at the end of Section 4.2.1, the quantiles associated with $\alpha \in [p_m, 1]$ in both directions u — equivalently, the quantiles associated with $\tau \in [0,(1-p_m)/2] \cup [(1+p_m)/2,1]$ — are equal to $-m$. As a consequence, when there is an atom in $-m$, Definition 4.2.2 is natural in dimension $d=2$, too (note that this case was not covered by Theorem 4.2.3). Since the case $d=2$ only involves univariate Euclidean quantiles, hence is well understood, we will mainly restrict to the case $d \geq 3$ when studying the proposed spherical quantiles below.

4.3 Basic properties of spherical quantiles and the spherical quantile function

We now provide some basic properties for the spherical quantiles introduced in the previous section, which will allow us to define and study the corresponding *spherical quantile function.*

4.3.1 Basic properties of spherical quantiles

We first answer the important questions of existence and uniqueness. We have the following result.

Theorem 4.3.1. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and let m be a Fréchet median of P. Fix $\alpha \in [0,1]$ and a unit vector u in $T_m\mathcal{S}^{d-1}$. Then, (i) P admits an m-geometric quantile $\mu^m_{\alpha,u}$ of order α in direction u. (ii) If P is not concentrated on a great circle containing m, then $\mu^m_{\alpha,u}$ is unique. (iii) If P is concentrated on a great circle $\mathcal{C}$ containing m, then uniqueness of $\mu^m_{\alpha,u}$ may fail only if u is one of the two directions belonging to the plane containing $\mathcal{C}$.*

This result shows that existence always holds and that uniqueness may only fail when P is not concentrated on a great circle containing m. But if P is concentrated on a great circle of $\mathcal{S}^{d-1}$ with $d \geq 3$, then P, after a suitable rotation, is actually a probability measure over the unit circle $\mathcal{S}^1 \times \{0\} \subset \mathbb{R}^2 \times \mathbb{R}^{d-2}$, which only requires circular quantiles. Thus, as soon as the problem genuinely requires (hyper)spherical quantiles (because P is not concentrated on a unit circle), we are allowed to speak of *the* m-quantile $\mu^m_{\alpha,u}(P)$ for any $\alpha \in [0,1]$ and any unit vector u in $T_m\mathcal{S}^{d-1}$.

As explained in the previous section, our spherical quantile concept uses a Fréchet median m as a reference point, that is *expected* to be the innermost quantile. From Definition 4.2.2, however, it is unclear that the resulting innermost quantile, namely the m-geometric median, coincides with the Fréchet median m. The following result shows that this is indeed the case.

Theorem 4.3.2. *Fix $P \in \mathcal{P}_{d-1}$ and a Fréchet median m of P. Then, m is an m-geometric median of P.*

This result is very general in the sense that it also covers the case where there may be several m-geometric medians, which, as explained below Theorem 4.3.1, is exceptional. In situations where all quantiles are unique, this result naturally states that the Fréchet median m that is used as a reference point is the unique m-median, which confirms that this reference point is then *the* innermost quantile for the proposed concept.

We now turn our attention to the high-order quantiles obtained as $\alpha \to p_m$ from below (it is superflous to consider quantiles with even higher orders since, for any $\alpha \geq p_m$, one has $\mu^m_{\alpha,u} = -m$ irrespective of u). In the Euclidean case, where there cannot be mass at infinity (so that high-order quantiles are obtained as $\alpha \to 1$), (i) high-order geometric quantiles exit any compact subset of $\mathbb{R}^d$ and (ii) they eventually do so in direction u, in the sense that the inner product between u and the unit vector proportional to these quantiles converges to one; see Theorem 2.1 in [42] for non-atomic measures and Theorems 2–3 in [83] for general ones. In the spherical case, we have the following result.

Theorem 4.3.3. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and let m be a Fréchet median of P. Assume that P is not concentrated on a great circle containing m. Let (α_n) be a sequence in $[0, p_m)$ that converges to p_m and (u_n) be a sequence of unit vectors in $T_m\mathcal{S}^{d-1}$. Then, (i) $\mu^m_{\alpha_n,u_n} \to -m$; (ii) if $(u_n) \to u$ for some u, then the direction v_n in which $\mu^m_{\alpha_n,u_n}$ is to be found from m (defined as the unit vector v_n in $T_m\mathcal{S}^{d-1}$ such that $\mu^m_{\alpha_n,u_n}$ belongs to the great half-circle $\{(\cos t)m + (\sin t)v_n : t \in [0, \pi]\}$) converges to u.*

In the Euclidean case, there is no guarantee that the distance between geometric quantiles of order $\alpha \to 1$ in direction u and the halfline $\{m + ru : r \geq 0\}$ converges to zero, that is, it may be so that these extreme quantiles are not eventually on the halfline with direction u originating from the geometric median m; see Figure 1(a)–(b) in [83] for examples. Interestingly, the spherical result in Theorem 4.3.3(ii) is thus stronger than the corresponding Euclidean result.

We conclude this section with a graphical illustration for $d = 3$. We consider the von Mises–Fisher distribution with location $\theta = (0, 0, 1)$ and concentration $\kappa = 1$, that is, the distribution, P_1 say, of

$$X = Z\theta + \sqrt{1 - Z^2}\begin{pmatrix} S \\ 0 \end{pmatrix},$$

where Z and S are mutually independent, Z admits the density $z \mapsto c_\kappa \exp(\kappa z)\mathbb{I}[-1 \leq z \leq 1]$ with respect to the Lebesgue measure (c_κ is a normalizing constant), and S is uniformly distributed over $\mathcal{S}^1$. We also consider the probability measure P_2 obtained when S rather results from projecting radially onto $\mathcal{S}^1$ a bivariate normal random vector with mean zero and covariance matrix $\Sigma = \text{diag}(25, 1)$ (in the terminology of [99], S thus follows an *angular Gaussian distribution* with shape matrix proportional to Σ). Both for $\ell = 1$ and $\ell = 2$, Figure 4.1 then draws the quantile curves $\alpha(\in [0, 1]) \mapsto \mu^m_{\alpha,u}(P_\ell)$ in each of the eight directions $u = (\cos(k\pi/4), \sin(k\pi/4), 0)$, $k = 0, 1, \ldots, 7$. In the rotationally symmetric setup $\ell = 1$, these quantile curves are geodesics (great half-circles) from m to $-m$, which actually illustrates Theorem 4.5.2 below. For $\ell = 2$, quantile curves are geodesics for $k = \{0, 2, 4, 6\}$ only, and the four other quantile curves are not contained in a plane, which reflects the non-rotational symmetry of this probability measure.

4.3.2 The spherical quantile function

The results of the previous section allow us to formally define the spherical quantile function associated with our quantile concept and to study some of its properties. For

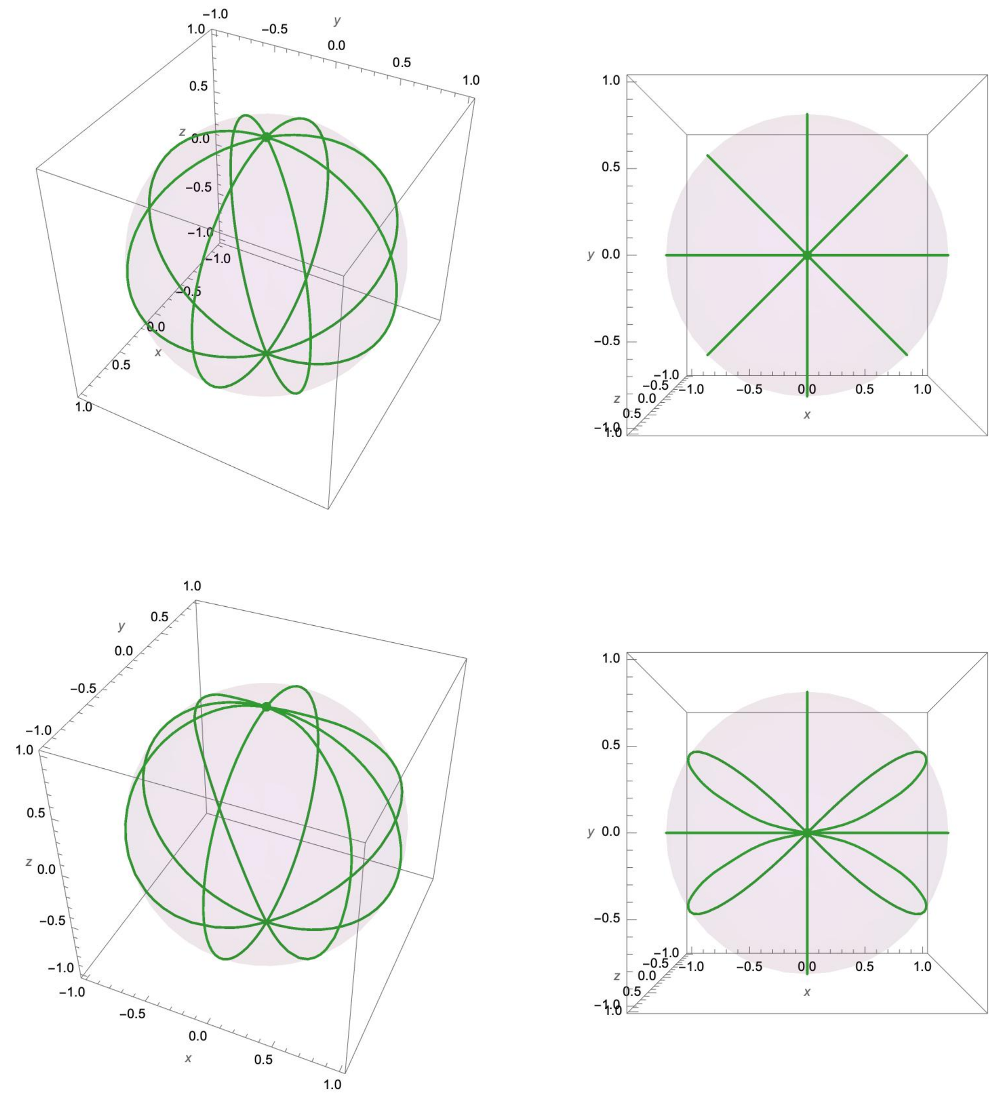

Figure 4.1: Quantile curves $\alpha(\in [0,1]) \mapsto \mu^m_{\alpha,u}$ in each of the eight directions $u = (\cos(k\pi/4), \sin(k\pi/4), 0)$, $k = 0, 1, \ldots, 7$, for the rotationally symmetric probability measure (P_1) and the non-rotationally symmetric one (P_2) described in the last paragraph of Section 4.3.1 (top and bottom row, respectively); in each case, the second column offers a view from above the Fréchet median m, that is marked as a green dot.

any $\mu \in \mathcal{S}^{d-1}$ and any $r \in (0, 1]$, let $\mathcal{B}_{\mu,r} = \{z \in T_\mu\mathcal{S}^{d-1} : \|z\| < r\}$ and $\mathcal{B}^\infty_{\mu,r} := \mathcal{B}_{\mu,r} \cup \{u^\infty_{\mu,r}\}$, where $u^\infty_{\mu,r}$ is a single element identifying all points in the closed annulus $\overline{\mathcal{B}_{\mu,1}} \backslash \mathcal{B}_{\mu,r}$ (here, $\bar{A}$ is the closure of A with respect to the usual topology). We endow the space $\mathcal{B}^\infty_{\mu,r}$

with the metric $\delta_{\mu,r}$ defined by

$$\delta_{\mu,r}(z_1, z_2) := \begin{cases} \|z_1 - z_2\| & \text{if } z_1, z_2 \in \mathcal{B}_{\mu,r} \\ r - \|z_1\| & \text{if } z_1 \in \mathcal{B}_{\mu,r} \text{ and } z_2 = u^{\infty}_{\mu,r} \\ r - \|z_2\| & \text{if } z_1 = u^{\infty}_{\mu,r} \text{ and } z_2 \in \mathcal{B}_{\mu,r} \\ 0 & \text{if } z_1 = z_2 = u^{\infty}_{\mu,r} \end{cases}$$

(it is easy to check that $\delta_{\mu,r}$ is a proper metric on $\mathcal{B}^{\infty}_{\mu,r}$). The corresponding norm on $\mathcal{B}^{\infty}_{\mu,r}$ is then defined through $\|z\|_{\mu,r} := \delta_{\mu,r}(0, z)$. Elements of $\mathcal{B}^{\infty}_{\mu,r}$ that belong to $\mathcal{B}_{\mu,r}$ will often be written as $z = \alpha u$, where $\alpha \in [0, r)$ and u is a unit vector of $T_\mu \mathcal{S}^{d-1}$. The quantile function is then formally defined as follows.

Definition 4.3.4. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and let m be a Fréchet median of P. Assume that P is not concentrated on a great circle containing m. Then, the m-quantile function of P is the map*

$$Q = Q^m_P : \mathcal{B}^{\infty}_{m,p_m} \to \mathcal{S}^{d-1}$$

defined through $Q(\alpha u) = \mu^m_{\alpha,u}$ for $\alpha u \in \mathcal{B}_{m,p_m}$ and $Q(u^{\infty}_{m,p_m}) = -m$.

This definition is motivated by the fact that $\mu^m_{\alpha,u} = -m$ for any $\alpha \in [p_m, 1]$ and any unit vector $u \in T_m \mathcal{S}^{d-1}$ (see Definition 4.2.2), so that identifying all points in the closed annulus $\overline{\mathcal{B}_{m,1}} \setminus \mathcal{B}_{m,p_m}$ to u^{∞}_{m,p_m} leads to the above definition. Observe that, in the important case where $-m$ is not an atom of P, $u^{\infty}_{m,p_m} = u^{\infty}_{m,1}$ simply identifies the points in the boundary of $\mathcal{B}_{m,1}$, that is, those belonging to the unit circle in $T_m \mathcal{S}^{d-1}$.

We turn to continuity of Q. We did not define the quantile function in the circular case $d = 2$, as our assumption that guarantees uniqueness of quantiles is never satisfied on the circle (for $d = 2$, the unit circle is itself a great circle through m that will always have P-probability one). Yet, a circular quantile function could similarly be defined once a convention has been taken to identify a unique quantile, such as, e.g., the classical infimum-based one in the univariate Euclidean case. The resulting circular quantile function may of course fail to be continuous (in particular, it will not be continuous for empirical probability measures). In contrast, for $d \geq 3$, the quantile function is continuous for any probability measure P, even for an empirical probability measure P. We have the following result.

Theorem 4.3.5. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and let m be a Fréchet median of P. Assume that P is not concentrated on a great circle containing m. Then the quantile function $Q = Q^m_P : \mathcal{B}^{\infty}_{m,p_m} \to \mathcal{S}^{d-1}$ is continuous (here, $\mathcal{B}^{\infty}_{m,p_m}$ is equipped with the metric δ_{m,p_m}).*

We will later show that, for any probability measure P on $\mathcal{S}^{d-1}$ with $d \geq 3$, the quantile function $Q^m_P : \mathcal{B}^{\infty}_{m,p_m} \to \mathcal{S}^{d-1}$ is actually surjective and that it may fail to be injective for atomic probability measures only; see Theorem 4.4.5. These further results, as well as the important result stating that the quantile function characterizes the underlying probability measure, require the spherical rank concept that will be introduced in the next section.

4.4 Gradient conditions and spherical ranks

The main goal of this section is to introduce a *spherical rank function*, that, under mild assumptions on the underlying probability measure, will be the inverse map of the spherical quantile function considered above. As we will see, this rank function is the right tool to obtain further results on the quantile function. The rank function is intimately linked to the gradient condition associated with the spherical quantiles $\mu^m_{\alpha,u}(P)$, a gradient condition that itself will follow from the directional derivatives of the objective function $M^{m,P}_{\alpha,u}$ defining these quantiles; see Definition 4.2.2.

Theorem 4.4.1. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and let m be a Fréchet median of P. Fix $\alpha \in [0, p_m)$ (still with $p_m := 1 - P[\{-m\}]$) and a unit vector $u \in T_m\mathcal{S}^{d-1}$. Fix $\mu \in \mathcal{S}^{d-1}_{-m}$ and a unit vector v in $T_\mu\mathcal{S}^{d-1}$. Let $\varphi : [0,\pi] \to \mathcal{S}^{d-1}$ be a geodesic path such that $\varphi(0) = \mu$ and $\dot\varphi(0) = v$. Then, the directional derivative*

$$\frac{\partial M^{m,P}_{\alpha,u}}{\partial v}(\mu) = \lim_{t \overset{>}{\to} 0} \frac{M^{m,P}_{\alpha,u}(\varphi(t)) - M^{m,P}_{\alpha,u}(\mu)}{t} \tag{4.4.5}$$

exists and is given by

$$\frac{\partial M^{m,P}_{\alpha,u}}{\partial v}(\mu) = \frac{1}{p_m}\big(d\pi_m(\mu)v\big)'\Big\{p_m \mathrm{E}\Big[\frac{\pi_m(\mu) - \pi_m(X)}{\|\pi_m(\mu) - \pi_m(X)\|}\xi_{X,\mu}\Big] - \alpha u\Big\}$$
$$+\frac{1}{p_m}\|d\pi_m(\mu)v\|P[\{\mu\}],$$

where X is an $\mathcal{S}^{d-1}_{-m}$-valued random vector with distribution P_{-m} and where we let $\xi_{x,y} = \mathbb{I}[x \neq y]$. Above, $d\pi_m(\mu) : T_\mu\mathcal{S}^{d-1} \to T_m\mathcal{S}^{d-1}$ is the differential of the map $\pi_m : \mathcal{S}^{d-1}_{-m} \to T_m\mathcal{S}^{d-1}$ in (4.2.3) (we refer to Lemma 4.8.7 for an explicit expression).

For $\alpha \in [0, p_m)$ and a unit vector u in $T_m\mathcal{S}^{d-1}$, any m-quantile of order α in direction u by definition belongs to $\mathcal{S}^{d-1}_{-m}$ and minimizes the objective function $M^{m,P}_{\alpha,u}$ over $\mathcal{S}^{d-1}_{-m}$. As we will show in Lemma 4.8.8, $\mu (\in \mathcal{S}^{d-1}_{-m})$ is an m-quantile of order α in direction u for P *if and only if* the directional derivative in (4.4.5) is non-negative for any unit vector v in $T_\mu\mathcal{S}^{d-1}$. Theorem 4.4.1 then allows us to obtain the gradient condition provided in the following result.

Theorem 4.4.2. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and let m be a Fréchet median of P. Fix $\alpha \in [0, p_m)$ and a unit vector u in $T_m\mathcal{S}^{d-1}$. Then, $\mu (\in \mathcal{S}^{d-1}_{-m})$ is an m-quantile of order α in direction u for P if and only if*

$$\Big\|p_m \mathrm{E}\Big[\frac{\pi_m(\mu) - \pi_m(X)}{\|\pi_m(\mu) - \pi_m(X)\|}\xi_{X,\mu}\Big] - \alpha u\Big\| \leq P[\{\mu\}], \tag{4.4.6}$$

where X is an $\mathcal{S}^{d-1}_{-m}$-valued random vector with distribution P_{-m}.

Fix $\mu \in \mathcal{S}^{d-1}_{-m}$, $\alpha \in [0, p_m)$ and a unit vector u in $T_m\mathcal{S}^{d-1}$, and assume that P is not concentrated on a great circle containing m, so that $\mu^m_{\alpha,u} = Q(\alpha u)$ is unique (Theorem 4.3.1). Denoting for a moment the quantity inside the norm in (4.4.6) as $R(\mu) - \alpha u$, Theorem 4.4.2 shows that $R(\mu) = \alpha u$ implies $\mu = Q(\alpha u)$. Thus, the resulting function R is a natural candidate to be the inverse map of Q. We adopt the following definition.

Definition 4.4.3. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and let m be a Fréchet median of P. Assume that P is not concentrated on a great circle containing m. Let X be an $\mathcal{S}^{d-1}_{-m}$-valued random vector with distribution P_{-m}. Then, the rank function of P is the map $R = R^m_P : \mathcal{S}^{d-1} \to \mathcal{B}^\infty_{m,p_m}$ such that*

$$R(\mu) = p_m \mathrm{E}\Big[\frac{\pi_m(\mu) - \pi_m(X)}{\|\pi_m(\mu) - \pi_m(X)\|}\xi_{X,\mu}\Big] \tag{4.4.7}$$

for $\mu \in \mathcal{S}^{d-1}_{-m}$ and $R(-m) = u^\infty_{m,p_m}$.

In the framework of this definition, Lemma 4.8.4 and Corollary 4.8.3 together entail that the distribution of $\pi_m(X)$ is not concentrated on a line of $\mathbb{R}^d$, so that the proof of Proposition 2.1 in [42] ensures that $\|R(\mu)\| < p_m$ for any $\mu \in \mathcal{S}^{d-1}_{-m}$; this justifies that the rank function R indeed takes its values in $\mathcal{B}^\infty_{m,p_m}$ and, less importantly, this also shows that $-m$ is the only location on the sphere that is given rank u^∞_{m,p_m}. Like the quantile function defined in the previous section, the rank function is then always continuous for $d \geq 3$.

Theorem 4.4.4. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and let m be a Fréchet median of P. Assume that P is not concentrated on a great circle containing m. Then the rank function $R = R^m_P : \mathcal{S}^{d-1} \to \mathcal{B}^\infty_{m,p_m}$ is continuous (again, $\mathcal{B}^\infty_{m,p_m}$ is equipped with the metric δ_{m,p_m}).*

By using this rank function, we can show that the quantile function $Q = Q^m_P$ is always a surjective map from $\mathcal{B}^\infty_{m,p_m}$ to $\mathcal{S}^{d-1}$ and that, under the further assumption that P is non-atomic, Q is a one-to-one map, whose inverse map is the corresponding rank function R. More precisely, we have the following result.

Theorem 4.4.5. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and let m be a Fréchet median of P. Assume that P is not concentrated on a great circle containing m. Then, (i) $Q^m_P : \mathcal{B}^\infty_{m,p_m} \to \mathcal{S}^{d-1}$ is a surjective map. (ii) If P is also non-atomic, then $Q^m_P : \mathcal{B}^\infty_{m,p_m} \to \mathcal{S}^{d-1}$ is a homeomorphism, with inverse given by $R^m_P : \mathcal{S}^{d-1} \to \mathcal{B}^\infty_{m,p_m}$.*

As the following result shows, the rank function $R = R^m_P$ actually characterizes the probability measure P, so that, under the assumptions that guarantee that the quantile and rank functions are inverse maps of one another, the quantile function also characterizes the underlying probability measure.

Theorem 4.4.6. *(i) Assume that $P_1, P_2 \in \mathcal{P}_{d-1}$ admit a common Fréchet median m and are not concentrated on a great circle containing m. Then, $R^m_{P_1} = R^m_{P_2}$ if and only if $P_1 = P_2$. (ii) Assume further that P_1 and P_2 are non-atomic. Then, $Q^m_{P_1} = Q^m_{P_2}$ if and only if $P_1 = P_2$.*

Inspection of the proof of Theorem 4.4.6(i) reveals that the result actually does not require the assumption that distributions are not concentrated on any great circle containing m (in such a case, rank functions are still defined as in Definition 4.4.3, but they are no more guaranteed to take their values in $\mathcal{B}^\infty_{m,p_m}$). This assumption, however, cannot be dropped in Theorem 4.4.6(ii) since quantile functions are not properly defined when this assumption is violated.

Figure 4.2 provides a graphical illustration of the proposed spherical ranks for the probability measures P_1 and P_2 already considered in Figure 4.1. More precisely, the figure represents the spherical ranks $R^m_{P_\ell}(\mu)$ associated with 15 locations μ on each of the 8 geodesics $\{(\cos\varphi)\theta + (\sin\varphi)u : \varphi \in [0,\pi]\}$ associated with $u = (\cos(k\pi/4), \sin(k\pi/4), 0)$, $k = 0, 1, \ldots, 7$. For easier visualisation, these ranks, that take values of the form $\alpha u = \alpha(u_1, u_2, 0)$ in $B^\infty_{m,1}$ (recall that $m = \theta = (0,0,1)$), are both drawn as arrows with length α and direction u located at the 15×8 points $\mu = (\cos\varphi)\theta + (\sin\varphi)u$ considered in $\mathcal{S}^2$ (left panels) or as arrows with length α and direction (u_1, u_2) located at the corresponding points $\varphi(u_1, u_2)$ in $\mathbb{R}^2$ (right panels). These ranks are to be thought as the duals of the quantiles in Figure 4.1: they give, along each of the corresponding geodesics, the order α and direction u for which the proposed spherical quantile is the given point on the geodesic. As expected, the norm of $R^m_{P_\ell}(\mu)$ converges to one as μ converges to $-m$. Clearly, ranks have the same direction as the corresponding geodesic in the rotationally symmetric case P_1, but this is the case for only half of the geodesics considered in the other case P_2.

4.5 Spherical depth

If a location μ on the unit sphere is an m-quantile of order α in direction u for the probability measure P at hand, then the larger α is, the more outlying μ is with respect to P. In other words, $\alpha = \|R(\mu)\|_{m,p_m}$ measures the *outlyingness* of μ with respect to P (here, $\|\cdot\|_{m,p_m}$ is the norm defined in Section 4.3.2). Thus, $1 - \|R(\mu)\|_{m,p_m}$ is a measure of centrality of μ with respect to P, which leads to the following definition.

Definition 4.5.1. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and let m be a Fréchet median of P. Assume that P is not concentrated on a great circle containing m. Then, the depth function of P is the map $D = D^m_P : \mathcal{S}^{d-1} \to [1-p_m, 1]$ such that $D(\mu) = 1 - \|R(\mu)\|_{m,p_m}$ for any $\mu \in \mathcal{S}^{d-1}$ (recall that, by definition, $\|R(-m)\|_{m,p_m} = \|u^\infty_{m,p_m}\|_{m,p_m} = p_m$).*

For any $d \geq 3$, the depth function D^m_P is continuous over $\mathcal{S}^{d-1}$ as soon as P is not concentrated on a great circle containing m (this is a direct corollary of Theorem 4.4.4). Thus, if μ diverges to "infinity" (with respect to m), that is, if it converges to $-m$, then $D^m_P(\mu)$ converges to $1-p_m$. This is a bit in contrast to depth functions in Euclidean spaces, for which a classical requirement is that the depth of μ converges to zero as $\|\mu\|$ diverges to infinity; see Property P4 in [108]. This "vanishing at infinity" property is a natural requirement indeed in Euclidean spaces since such spaces cannot contain probability mass at infinity. We argue that since, in contrast, spheres may have an atom at $-m$, it is also natural that the depth does not vanish at $-m$ in such cases. We stress, however, that in the important case $p_m = 1$ where there is no atom at $-m$, then the proposed spherical depth is indeed vanishing at infinity, which is quite natural—interestingly, recent proposals for depth on general metric spaces actually rather impose this vanishing at infinity for unbounded spaces only; see, e.g., [24].

In the same line of thought, note that, even in the case where the proposed spherical depth could in principle be equal to zero (that is, in the case $p_m = 1$ where there is no atom at $-m$), the properties of the rank function entail that zero depth will be achieved at $-m$ only. In other words, irrespective of the probability measure P at

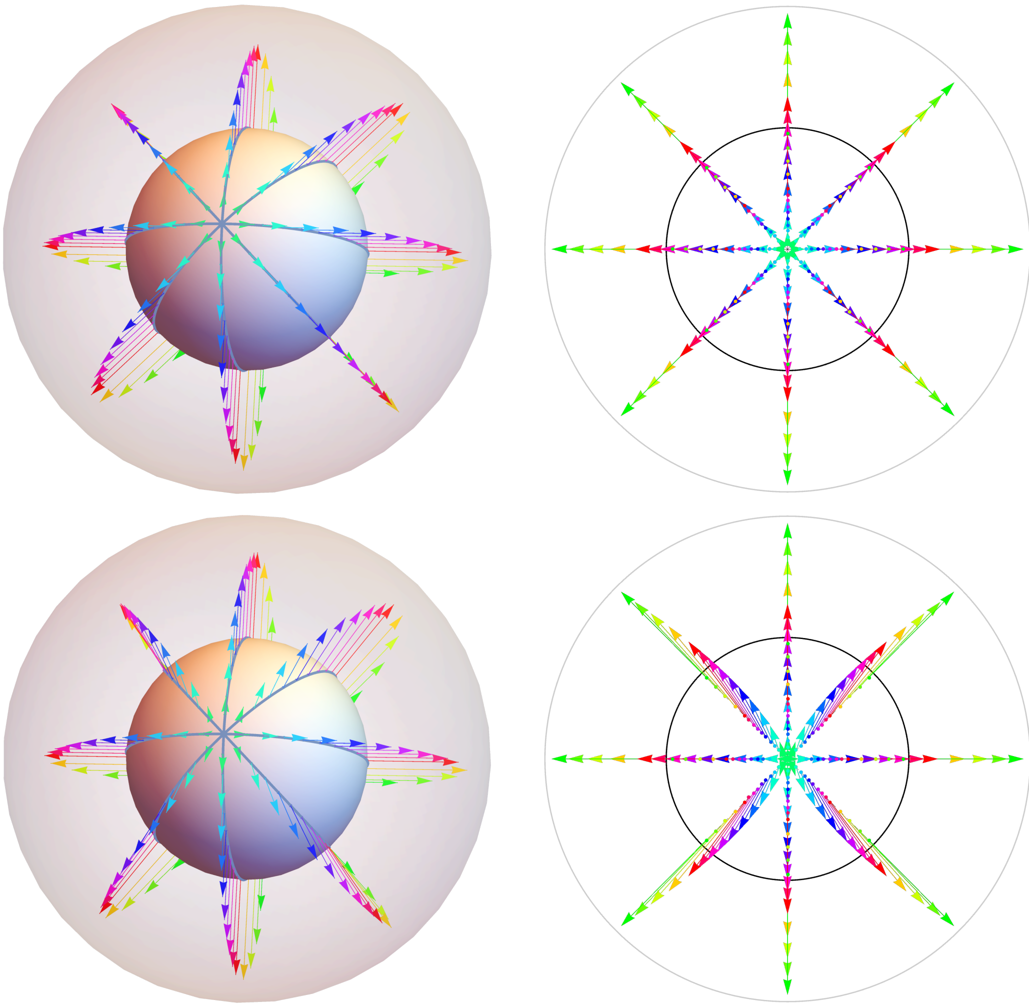

Figure 4.2: Ranks $R^m_{P_\ell}(\mu)$ associated with 15 locations μ on each of the 8 geodesics $\{(\cos\varphi)\theta + (\sin\varphi)u : \varphi \in [0, \pi]\}$ associated with $u = (\cos(k\pi/4), \sin(k\pi/4), 0)$, $k = 0, 1, \ldots, 7$. Ranks, that are of the form $\alpha u = \alpha(u_1, u_2, 0)$, here are drawn as arrows with length α and direction u located at the 15×8 points $\mu = (\cos\varphi)\theta + (\sin\varphi)u$ considered in $\mathcal{S}^2$ (left panels) or as arrows with length α and direction (u_1, u_2) located at the corresponding points $\varphi(u_1, u_2)$ in $\mathbb{R}^2$ (right panels); the top and bottom rows correspond to the rotationally symmetric probability measure (P_1) and the non-rotationally symmetric one (P_2) already considered in Figure 4.1, respectively.

hand, our depth has the "non-vanishing property" to stay strictly positive over $\mathcal{S}^{d-1}_{-m}$. This is a very desirable property in some inferential applications of depth, such as, e.g., supervised classification. In depth-based classification, an observed location $\mu = x$ is assigned to a probability measure P_1 rather than P_2 if the depth of x with respect to P_1 is larger than the depth of x with respect to P_2. Most depths available on the sphere will provide zero depth values when x is outside the convex hull of the supports of P_1 and P_2, which then makes it impossible to classify x based on depth. Interestingly, the above non-vanishing property of our spherical depth will avoid this problem (we refer to [34] for an interesting discussion on this non-vanishing issue).

Now, for any depth function, be it in a Euclidean space or a non-Euclidean one, it is natural to consider the corresponding depth regions, that collect the locations μ with a depth that is larger than or equal to a given level α. In other words, the α-depth region is

$$\mathcal{R}_P^m(\alpha) := \{\mu \in \mathcal{S}^{d-1} : D_P^m(\mu) \geq \alpha\},$$

and the corresponding depth contour is then the boundary, $\mathcal{C}_P^m(\alpha) := \partial\mathcal{R}_P^m := \{\mu \in \mathcal{S}^{d-1} : D_P^m(\mu) = \alpha\}$ of this depth region. Obviously, depth regions form a collection of nested subsets of the unit sphere $\mathcal{S}^{d-1}$, the most inner one, $\mathcal{R}_P^m(1)$, being $\{m\}$, and the most outer one, $\mathcal{R}_P^m(1 - p_m)$, being the sphere itself. The shape of the depth contours reflects the "structure" of the underlying probability measure. In particular, we have the following result.

Theorem 4.5.2. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and let m be a Fréchet median of P. Assume that P is rotationally symmetric about m. Then (i) for $\alpha = 0$ and any unit vector $u \in T_m\mathcal{S}^{d-1}$, the unique m-quantile $\mu_{\alpha,u}^m$ is m; (ii) for any $\alpha \in [0, 1]$ and any unit vector $u \in T_m\mathcal{S}^{d-1}$, the unique m-geometric quantile $\mu_{\alpha,u}^m$ belongs to the meridian $\{(\cos t)m + (\sin t)u : t \in [0, \pi]\}$; (iii) for any unit vector $u \in T_m\mathcal{S}^{d-1}$, the map $\alpha \mapsto d(\mu_{\alpha,u}^m, m)$ is monotone non-decreasing over $[0, 1]$; (iv) when P is not concentrated on $\{-m, m\}$, each depth contour $\mathcal{C}_P^m(\alpha)$ is of the form $\{\mu \in \mathcal{S}^{d-1} : \mu' m = c_\alpha\}$, and the map $\alpha \mapsto c_\alpha$ is monotone non-decreasing.*

A key ingredient in the proof of Theorem 4.5.2 is the following rotation-equivariance result for spherical quantiles, which is of independent interest.

Theorem 4.5.3. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and let m be a Fréchet median of P. Let O be a $d \times d$ orthogonal matrix such that $Om = m$ and denote as P_O the distribution of OX when X has distribution P. Fix $\alpha \in [0, 1]$ and a unit vector u in $T_m\mathcal{S}^{d-1}$. Then, μ is an m-quantile of order α in direction u for P if and only if $O\mu$ is an m-quantile of order α in direction Ou for P_O.*

We stress that, in Theorem 4.5.2(ii), m may be an m-geometric quantile of order $\alpha > 0$ in direction u, provided that m is an atom of P. Actually, it is easy to prove that the largest α for which m is an m-geometric quantile of order α in (any) direction u is $P[\{m\}]$. More importantly, Theorem 4.5.2(iv) shows that a probability measure P that is rotationally symmetric about m provides depth contours that are themselves invariant under rotations fixing m; of course, this is also the case for the corresponding depth regions, that are spherical caps centered at m. Departures from rotational symmetry will result into depth regions that exhibit other shapes, which is an advantage over the depth regions from [55] that are "concentric" spherical caps even for probability measures that are not rotationally symmetric. This is illustrated in Figure 4.2 that draws the depth contours $C_P^m(\alpha)$, $\alpha = 0, 0.2, 0.4, 0.6, 0.8$, for both probability measures that were considered in Figures 4.1–4.2.

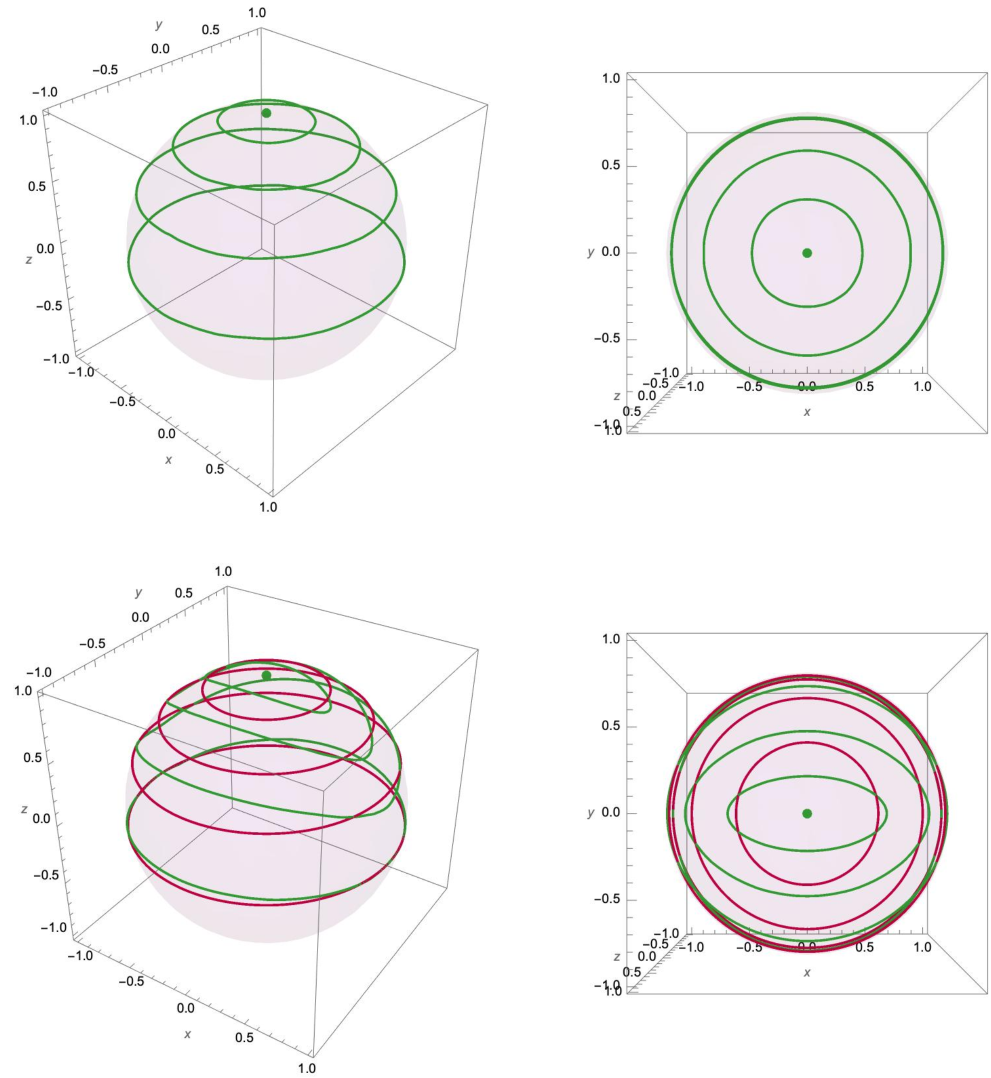

Figure 4.3: Depth contours $C_P^m(\alpha)$, $\alpha = 0, 0.2, 0.4, 0.6, 0.8$, computed from the first probability measure (top row) and second one (bottom row) in Figure 4.1; in each case, the second column offers a view from above the Fréchet median, that is marked as a green dot. In the bottom row, the quantile contours from [55] containing the same probability mass as the proposed contours are plotted in red (these are not plotted in the top row since, in the rotationally symmetric setup considered there, those contours coincide with the proposed ones).

4.6 Asymptotics

In the sample case, evaluation of our spherical quantiles requires estimating the population Fréchet median(s), and we therefore start this section by studying the asymptotic behaviour of sample Fréchet medians. When a random sample $X_1, \ldots, X_n$ is available, a sample Fréchet median is defined as a Fréchet median of the corresponding empirical probability measure P_n, that is, as a minimizer of

$$\mu \mapsto \frac{1}{n}\sum_{i=1}^{n} d(\mu, X_i) = \frac{1}{n}\sum_{i=1}^{n} \arccos(\mu' X_i)$$

over $\mathcal{S}^{d-1}$. When observations are randomly sampled from a probability measure P admitting a unique Fréchet median, we have the following almost sure consistency result.

Theorem 4.6.1. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and assume that P has a unique Fréchet median m. Let $X_1, X_2, \ldots$ be mutually independent random vectors with distribution P, and let $\hat{m}_n$ be an arbitrary sample Fréchet median associated with $X_1, \ldots, X_n$. Then, $\hat{m}_n \to m$ almost surely as n diverges to infinity.*

Under the further assumption that P admits a density, we have the following Bahadur representation result, which of course guarantees asymptotic normality of sample Fréchet medians.

Theorem 4.6.2. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and assume that P has a unique Fréchet median m. Assume further that P admits a density with respect to the surface area measure on $\mathcal{S}^{d-1}$ and that*

$$K := \mathrm{E}\bigg[\frac{m'X}{\|(I_d - mm')X\|}\bigg(I_d - \frac{(I_d - mm')XX'(I_d - mm')}{\|(I_d - mm')X\|^2}\bigg)\xi_{X,\pm m}\bigg]$$

exists, is finite and is invertible; here, X is an $\mathcal{S}^{d-1}$-valued random vector with distribution P and $\xi_{x,\pm y} := \mathbb{I}[x \notin \{\pm y\}]$. Let $X_1, X_2, \ldots$ be mutually independent random vectors with distribution P, and let $\hat{m}_n$ be an arbitrary sample Fréchet median associated with $X_1, \ldots, X_n$. Then,

$$\sqrt{n}(\hat{m}_n - m) = K^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{(I_d - mm')X_i}{\|(I_d - mm')X_i\|}\xi_{X_i,\pm m} + o_{\mathrm{P}}(1)$$

as n diverges to infinity. Moreover, $\sqrt{n}(\hat{m}_n - m) = (I_d - mm')\sqrt{n}(\hat{m}_n - m) + o_{\mathrm{P}}(1)$ as n diverges to infinity.

We now turn to the asymptotic behaviour of sample spherical quantiles. Before defining these sample quantiles, we need to address the following issue. Quite naturally, sample quantiles will be relative to a sample Fréchet median $\hat{m}_n$, hence will involve a direction u in $\in T_{\hat{m}_n}\mathcal{S}^{d-1}$, whereas directions associated with population quantiles belong to a different tangent vector space, namely $T_m\mathcal{S}^{d-1}$. To investigate the asymptotic behaviour of sample quantiles, we thus need to match the directions associated with different locations on $\mathcal{S}^{d-1}$. We argue that a natural way to pick a particular direction u at a location $\mu \in \mathcal{S}^{d-1}$ is to consider the direction pointing to a given target location $\mu_* \in \mathcal{S}^{d-1}$, that is, the unit vector $u = U_{\mu_*}(\mu) \in T_\mu\mathcal{S}^{d-1}$ such that the geodesic from μ to μ_* is $\{(\cos t)\mu + (\sin t)u : t \in [0, d(\mu, \mu_*)]\}$. This scheme, that identifies a unique direction u at any $\mu \in \mathcal{S}^{d-1} \setminus \{\pm\mu_*\}$, indeed allows us to match directions associated with different locations on the sphere, in the sense that, if the location μ

is changed into $\tilde{\mu}$, then the direction $u = U_{\mu_*}(\mu) \in T_\mu \mathcal{S}^{d-1}$ is changed accordingly into $\tilde{u} = U_{\mu_*}(\tilde{\mu}) \in T_{\tilde{\mu}} \mathcal{S}^{d-1}$.

Assume then that a random sample $X_1, \ldots, X_n$ from a probability measure P is available. We will rely on sample splitting and first estimate m by the sample Fréchet median $\hat{m}_n$ of $X_1, \ldots, X_{\lfloor cn \rfloor}$, where $c \in (0,1)$ is fixed. As explained above, we focus on directions u pointing to a given target point $\mu_* \in \mathcal{S}^{d-1}$ and then estimate, for any $\alpha \in [0,1)$, the population quantile $\mu^m_{\alpha,U(m)}$ (from now on, we simply write $U(\mu)$ rather than $U_{\mu_*}(\mu)$) by the minimizer $\hat{\mu}^{\hat{m}_n}_{\alpha,U(\hat{m}_n)}$ of the objective function

$$\mu \mapsto \frac{1}{n - \lfloor cn \rfloor} \sum_{i=\lfloor cn \rfloor+1}^{n} \left\{ \|\pi_{\hat{m}_n}(X_i) - \pi_{\hat{m}_n}(\mu)\| - \|\pi_{\hat{m}_n}(X_i)\| - \alpha (U(\hat{m}_n))' \pi_{\hat{m}_n}(\mu) \right\}$$

over $\mathcal{S}^{d-1}_{-\hat{m}_n}$ (uniqueness is guaranteed as soon as the X_i's, $i = \lfloor cn \rfloor + 1, \ldots, n$, do not all belong to a common great circle containing $\hat{m}_n$, which, under the assumptions of Theorem 4.6.3 below, occurs with probability one; see Theorem 4.3.1). Of course, for $\alpha = 1$, we simply put $\hat{\mu}^{\hat{m}_n}_{\alpha,u} := -\hat{m}_n$ for any $u \in T_{\hat{m}_n}\mathcal{S}^{d-1}$.

Theorem 4.6.3. *Fix a probability measure $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, that satisfies the assumptions of Theorem 4.6.2. Let $X_1, X_2, \ldots$ be mutually independent random vectors with distribution P, fix $c \in (0,1)$, and let $\hat{m}_n$ be an arbitrary sample Fréchet median associated with $X_1, \ldots, X_{\lfloor cn \rfloor}$. Fix $\alpha \in [0,1)$ and a target point $\mu_* \in \mathcal{S}^{d-1} \setminus \{\pm m\}$. Then, writing $u = U(m)$, $u_n = U(\hat{m}_n)$, and $q := \pi_m(\mu^m_{\alpha,u})$, and denoting as $J_x(g)$ the Jacobian matrix of g at x,*

$$\begin{aligned}\sqrt{n}(\hat{\mu}^{\hat{m}_n}_{\alpha,u_n} - \mu^m_{\alpha,u}) &\\ &= \left\{ (m'\mu^m_{\alpha,u}) I_d + J_q(\pi_m^{-1}) V^{-1} (\alpha J_m(U) - L) \right\} \sqrt{n}(\hat{m}_n - m) \\ &\quad + \frac{1}{1-c} J_q(\pi_m^{-1}) V^{-1} \frac{1}{\sqrt{n}} \sum_{i=\lfloor cn \rfloor+1}^{n} \left(\frac{\pi_m(X_i) - q}{\|\pi_m(X_i) - q\|} \xi_{\pi_m(X_i),q} + \alpha u \right) + o_{\mathrm{P}}(1)\end{aligned} \tag{4.6.8}$$

as n diverges to infinity, with

$$V := \mathrm{E}\left[\frac{1}{\|\pi_m(X) - q\|} \left(I_d - \frac{(\pi_m(X) - q)(\pi_m(X) - q)'}{\|\pi_m(X) - q\|^2} \right) \xi_{\pi_m(X),\mu} \right]$$

and

$$L := \mathrm{E}\left[\frac{\xi_{\pi_m(X),q}\,\xi_{X,-m}}{(1 + m'X)\|\pi_m(X) - q\|} \left(I_d - \frac{(\pi_m(X) - q)(\pi_m(X) - q)'}{\|\pi_m(X) - q\|^2} \right) \{(m'X) I_d + (m + q) X'\} \right].$$

In the Bahadur representation (4.6.8), the first term of the righthand side is associated with estimation of the Fréchet median m, whereas the second term corresponds to spherical quantile estimation[2] (if one would know the population Fréchet median m, then all observations could be used to estimate the quantile, i.e., one could take $c = 0$, and only the second term would show in the Bahadur representation of the resulting spherical quantiles). Two remarks are in order:

[2] If one forgets the scalar factor $1/(1-c)$ and the Jacobian matrix π_m^{-1}, then this second term provides exactly the asymptotics in the Euclidean case for the pushed forward probability measure $\pi_m \# P_{-m}$ for a fixed m.

(i) At first, it may be puzzling that, for $\alpha = 0$, the Bahadur representation does not simply reduce to $\sqrt{n}(\hat{m}_n - m)$. From Theorem 4.3.2, this should indeed be the case if $\hat{\mu}_n$ would be computed from a Fréchet median of the observations showing in the second term of the Bahadur representation. However, this is almost surely not the case due to sample splitting.

(ii) [55] also obtained a two-term Bahadur representation of this form for their spherical cap quantiles, and they actually showed that the first term vanishes under rotational symmetry (see their Proposition 3.2). In contrast, rotational symmetry will not put to zero the first term in the righthand side of (4.6.8), which is due to the fact that the directional nature (in u) of our quantiles breaks the natural symmetry.

Finally, note that, from sample splitting, both terms in the Bahadur representation (4.6.8) are mutually independent, which allows one to derive the variance in the corresponding asymptotically normal distribution by simply summing the asymptotic variances of each term.

4.7 An inferential application

In this section, we consider the problem of testing the null hypothesis that a probability measure $P \in \mathcal{P}_{d-1}$ is rotationally symmetric with respect to a specified median location m (for the sake of simplicity, we assume throughout that P admits a density with respect to the surface area measure on $\mathcal{S}^{d-1}$). It follows from Theorem 4.5.2 that, under the null hypothesis, $Q_P^m(\alpha u) = (\cos\varphi_\alpha)m + (\sin\varphi_\alpha)u =: z^m_{\varphi_\alpha,u}$ for some $\varphi_\alpha \in [0,\pi]$. Since, under the assumptions adopted here, R_P and Q_P are inverse maps of one another, this implies that

$$R_P^m(z^m_{\varphi,u}) = \lambda_\varphi u, \qquad \text{with } \lambda_\varphi := \int_{\mathcal{U}_m} \|R_P^m(z^m_{\varphi,u})\| \, d\sigma_m(u),$$

where $\mathcal{U}_m$ denotes the collection of unit vectors in $T_m\mathcal{S}^{d-1}$ and σ_m is the surface area measure on $\mathcal{U}_m$. It is then expected that

$$T_P^m := \int_0^\pi \int_{\mathcal{U}_m} \left\| R_P^m(z^m_{\varphi,u}) - \left(\int_{\mathcal{U}_m} \|R_P(z^m_{\varphi,v})\| \, d\sigma_m(v) \right) u \right\|^2 d\sigma_m(u) \, d\varphi$$

measures deviations from rotational symmetry about m. We have the following result.

Theorem 4.7.1. *Let $P \in \mathcal{P}_{d-1}$ admit a density on $\mathcal{S}^{d-1}$. Then, $T_P^m = 0$ if and only if P is rotationally symmetric with respect to m.*

Assume now that a random sample $X_1, \ldots, X_n$ from P is available and denote the corresponding empirical measure by P_n. Theorem 4.7.1 suggests that the test rejecting the null hypothesis of rotational symmetry about m for large values of

$$T_{P_n}^m := \int_0^\pi \int_{\mathcal{U}_m} \left\| R_{P_n}^m(z^m_{\varphi,u}) - \left(\int_{\mathcal{U}_m} \|R_{P_n}(z^m_{\varphi,v})\| \, d\sigma_m(v) \right) u \right\|^2 d\sigma_m(u) \, d\varphi \qquad (4.7.9)$$

is an omnibus test (i.e., is consistent against any alternative). Deriving the asymptotic null distribution of this test statistic would obviously require a stochastic process version of the asymptotic result in Theorem 4.6.3. Not only is such a result beyond the scope of the present work, but it would also provide an asymptotic distribution that depends on the particular null distribution P at hand. Here, we favour a more efficient approach relying on exact distribution-freeness. Let $R_n = (R_{n1}, \ldots, R_{nn})$ and $U_n = (U_{n1}, \ldots, U_{nn})$, where R_{ni} is the rank of $X_i'm$ among $X_1'm, \ldots, X_n'm$ and $U_{ni} := (I_d - mm')X_i/\|(I_d - mm')X_i\|$, $i = 1, \ldots, n$. Under the null hypothesis of rotational symmetry about m, R_n is uniformly distributed over all permutations of $\{1, \ldots, n\}$, the U_{ni}'s form a random sample from the uniform distribution over $\mathcal{U}_m$, and R_n and U_n are mutually independent. As a corollary, denoting as $\tilde{P}_n$ the empirical probability measure associated with the transformed sample

$$\tilde{X}_i = \frac{R_{ni}}{n+1} m + \sqrt{1 - \left(\frac{R_{ni}}{n+1}\right)^2}\, U_{ni} \qquad i = 1, \ldots, n,$$

the test statistic $T^m_{\tilde{P}_n}$ is distribution-free under the null hypothesis (note that the $\tilde{X}_i$'s form a random sample from a distribution that is rotationally symmetric about m if and only if the X_i's do). Thanks to distribution-freeness, critical values can of course be arbitrarily well approximated through simulations; more precisely, at level $\alpha \in (0, 1)$, the corresponding test will reject the null hypothesis if and only if

$$T^m_{\tilde{P}_n} > c_\alpha(G), \tag{4.7.10}$$

where $c_\alpha(G)$ is the sample $(1 - \alpha)$-quantile in a collection of G mutually independent values of $T^m_{\tilde{P}_n}$ under the null hypothesis (from distribution-freeness, these G values can be obtained by simulating from an arbitrary distribution that is rotationally symmetric about m).

We explore the finite-sample performances of this test through the following Monte Carlo exercise. For each value of $\ell = \{0, 1, 2, 3, 4\}$, we generated $M = 5{,}000$ independent random samples of size $n = 200$ from four different distributions indexed by ℓ; in each case, $\ell = 0$ will correspond to the null hypothesis of rotational symmetry about $m = (0, 0, 1)$, whereas $\ell = 1, 2, 3, 4$ will provide increasingly severe alternatives. The four distributions are as follows:

(i) *Tangent von Mises–Fisher:* for $\kappa = 1$, the first distribution is the one of

$$Zm + \sqrt{1 - Z^2} \begin{pmatrix} S \\ 0 \end{pmatrix},$$

where Z and S are mutually independent, Z admits the density $z \mapsto c_\kappa \exp(\kappa z)\mathbb{I}[-1 \leq z \leq 1]$ (c_κ is a normalizing constant), and S follows a von Mises–Fisher distribution with location $(1, 0)'$ and concentration $\eta_\ell = \ell/10$;

(ii) *Tangent elliptical:* this distribution is the same as in (i), but for the fact that S rather results from projecting radially onto $\mathcal{S}^1$ a bivariate normal random vector with mean zero and covariance matrix $\Sigma_\ell = \text{diag}(1 + \ell/2, 1)$;

(iii) *Beta in longitude:* the third distribution is the same as in (i)–(ii), but for the fact that $S = (\cos T, \sin T)'$, where $T/(2\pi)$ is $\text{Beta}(s_\ell, s_\ell)$, with $s_\ell = 1 + \ell/8$;

(iv) *Dependence in longitude-latitude:* this last distribution is the one of $((\cos T)(\sin W), (\sin T)(\sin W), \cos W)'$, where T is uniform over $[0, 2\pi]$ and W, conditional on $[T = t]$, is uniform over $[0, \pi\{T(2\pi - T)/\pi^2\}^{\ell/2}]$.

In each sample, we performed the following five tests at nominal level $\alpha = 5\%$: (1) the proposed distribution-free test above, where the critical value was obtained from $G = 100,000$ independent random samples generated from the von Mises–Fisher distribution with location $m = (0, 0, 1)$ and concentration $\kappa = 1$ (both to obtain its critical value then to perform the test, evaluation of the integrals in (4.7.9) was done along regular grids of size 30 for both u and φ); (2) the semiparametric test from [57]; (3)–(4) The "location" and "scatter" tests from [36], that are optimal against tangent von Mises–Fisher alternatives and tangent elliptical alternatives, respectively; (5) the test of rotational symmetry based on celebrated Kuiper's test of uniformity over $\mathcal{S}^{p-2}$; see Page 99 in [67] (see also [36] for more details).

The resulting rejection frequencies are plotted against ℓ in Figure 4.4. In line with distribution-freeness, the proposed test shows the target size under the null hypothesis in all cases (i)–(iv). Clearly, the test exhibits power against the four types of alternatives considered (which was expected in view of Theorem 4.7.1), but these simulations reveal that it is the only test that does so among the five tests considered here (the scatter test is blind to alternatives of type (i), the LV test and location test to alternatives of type (ii), and the Kuiper test to alternatives of type (iv)). The proposed test performs very well against alternatives of type (i) and (iii), particularly so for type (i) since it competes almost equally with the optimal location test for such alternatives.

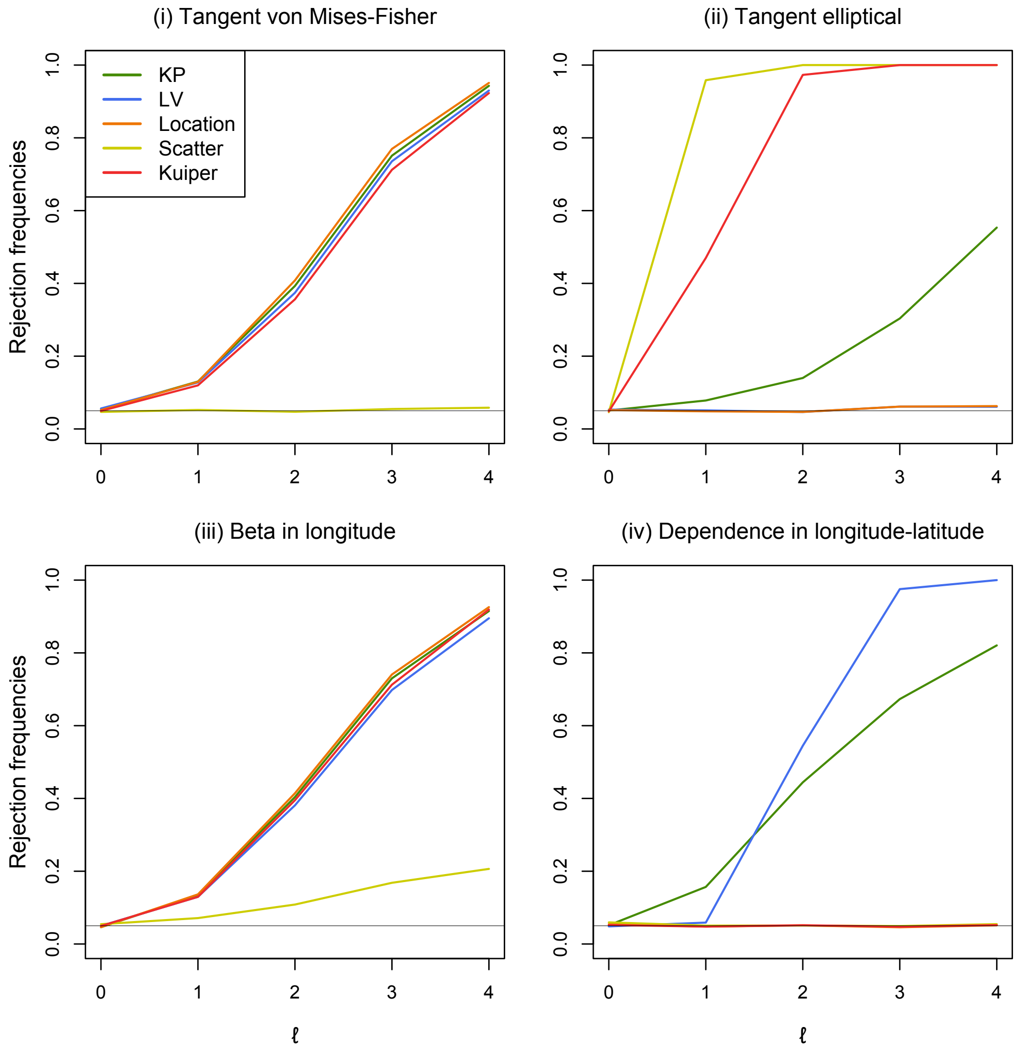

Figure 4.4: Rejection frequencies of five tests of rotational symmetry about $m = (0, 0, 1)$ (each test is performed at nominal level $\alpha = 5\%$) based on $M = 5,000$ mutually independent random samples of size $n = 200$ drawn from four different models (i)–(iv); in each case, $\ell = 0$ corresponds to the null hypothesis and $\ell = 1, 2, 3, 4$ provide increasingly severe alternatives. We refer to Section 4.7 for details on the five tests and the four models used here.

4.8 Appendix: proofs

Some preliminary results

Some of the proofs in the next sections will require the results below.

Lemma 4.8.1. *Let $z_0 \in \mathbb{R}^d$ and $V = (v_1 \ \dots \ v_k)$ be a $d \times k$ matrix that has orthonormal columns, and consider the corresponding k-dimensional affine hyperplane $H = \{z_0 + Vy : y \in \mathbb{R}^k\}$. Let P be a probability measure over $\mathbb{R}^d$ such that $P[H] = 1$. Fix $\alpha \in [0, 1)$ and fix $u \in \mathcal{S}^{d-1}$ such that u is a linear combination of the v_ℓ's. Then, any geometric quantile of order α in direction u for P belongs to H, and is of the form $z_0 + V\xi_{\alpha,u}$,* *where $\xi_{\alpha,u}$ is a geometric quantile of order α in direction $V'u$ for the probability measure on $\mathbb{R}^k$ defined through $Q[B] = P[z_0 + VB]$ for any Borel set B of $\mathbb{R}^k$.*

PROOF OF LEMMA 4.8.1. (i) Fix $\mu \notin H$. Let Z be a random d-vector with distribution P and define the random k-vector Y through $Z =: \mu_H + VY$, where μ_H is the orthogonal projection of μ onto H. Define further $w := (\mu_H - \mu)/c$, with $c := \|\mu_H - \mu\|$. Since $u'w = 0$, we then have (recall the definition of the [17] objective function in (4.1.1))

$$\begin{aligned}\frac{O^P_{\alpha,u}(\mu + hw) - O^P_{\alpha,u}(\mu)}{h} &= -\alpha u'w + \int_{\mathbb{R}^d} \frac{\|z - (\mu + hw)\| - \|z - \mu\|}{h}\, dP(z)\\ &= \int_{\mathbb{R}^k} \frac{\|(\mu_H + Vy) - (\mu + hw)\| - \|(\mu_H + Vy) - \mu\|}{h}\, dP^Y(y)\\ &= \int_{\mathbb{R}^k} \frac{\|Vy + cw - hw\| - \|Vy + cw\|}{h}\, dP^Y(y).\end{aligned}$$

This yields

$$\frac{O^P_{\alpha,u}(\mu + hw) - O^P_{\alpha,u}(\mu)}{h} + \int_{\mathbb{R}^k} \frac{w'(Vy + cw)}{\|Vy + cw\|}\, dP^Y(y) = \int_{\mathbb{R}^k} g_h(y)\, dP^Y(y),$$

where

$$\begin{aligned}g_h(y) \ &:= \ \frac{\|Vy + cw - hw\| - \|Vy + cw\|}{h} + \frac{w'(Vy + cw)}{\|Vy + cw\|}\\ &= \ \frac{h^2 - 2hw'(Vy + cw)}{h(\|Vy + cw - hw\| + \|Vy + cw\|)} + \frac{w'(Vy + cw)}{\|Vy + cw\|}.\end{aligned}$$

Clearly, $y \mapsto |g_h(y)|$ is, for $h \in (0, 1)$ say, upper-bounded by the function $y \mapsto (1/\|Vy + cw\|) + 3$ that is P^Y-integrable and does not depend on h (integrability follows from the fact that $\|Vy + cw\|^2 = \|Vy\|^2 + c^2 \geq c^2$). Moreover, $g_h(y) \to 0$ as $h \to 0$ for any y. Lebesgue's dominated convergence theorem thus shows that the directional derivative of $O^P_{\alpha,u}$ at μ in direction w exists and is given by

$$\frac{\partial O^P_{\alpha,u}}{\partial w}(\mu) = -\int_{\mathbb{R}^k} \frac{w'(Vy + cw)}{\|Vy + cw\|}\, dP^Y(y) = -\int_{\mathbb{R}^k} \frac{c}{\|Vy + cw\|}\, dP^Y(y) < 0.$$

Therefore, μ is not a geometric quantile of order α in direction u for P.

(ii) We just showed that all geometric quantiles of order α in direction u for P belong to H, hence are of the form $z_0 + V\xi_{\alpha,u}$. In this part of the proof, define the random k-vector Y through $Z =: z_0 + VY$. Clearly, Y has distribution Q. By definition, $\xi_{\alpha,u}$ thus minimizes

$$\begin{aligned}\xi \mapsto O^P_{\alpha,u}(z_0 + V\xi) &= \int_{\mathbb{R}^d} \{\|z - (z_0 + V\xi)\| - \|z\| - \alpha u'(z_0 + V\xi)\}\, dP(z)\\ &= -\alpha u'z_0 + \int_{\mathbb{R}^k} \{\|y - \xi\| - \|z_0 + Vy\| - \alpha u'V\xi\}\, dQ(y)\end{aligned}$$

(where we used the fact that $V'V = I_k$), or, equivalently, minimizes

$$\xi \mapsto \int_{\mathbb{R}^k} \{\|y - \xi\| - \|y\| - \alpha(V'u)'\xi\}\, dQ(y),$$

hence is a geometric quantile of order α in direction $V'u$ for Q. ■

Lemma 4.8.2. *Assume that $P \in \mathcal{P}_{d-1}$ is concentrated on the spherical cap $C = \{x \in \mathcal{S}^{d-1} : p'x \geq h\}$, with $p \in \mathcal{S}^{d-1}$ and $h > 0$. Then, any Fréchet median of P belongs to C.*

Proof of Lemma 4.8.2. Since the collection of Fréchet medians of P is equivariant under orthogonal transformations, we may assume without any loss of generality that $p = (0, \ldots, 0, 1)$, so that

$$C = \{x = (x_1, \ldots, x_d) \in \mathcal{S}^{d-1} : x_d \geq h\}.$$

Fix then $\mu = (0, \ldots, 0, \mu_{d-1}, \mu_d) \in \mathcal{S}^{d-1} \setminus C$, with $\mu_{d-1} \geq 0$. Since $\mu \in \mathcal{S}^{d-1}$, we have that

$$\mu = (0, \ldots, 0, (1 - \mu_d^2)^{1/2}, \mu_d). \tag{4.8.11}$$

We first show that there exists $\tilde{\mu} \in \mathcal{S}^{d-1}$ such that $g_Q(\tilde{\mu}) < g_Q(\mu)$ for any probability measure Q concentrated on C. To do so, we need to consider two cases. (a) $|\mu_d| < |h|$ (i.e., $-h < \mu_d < h$). Let then $\tilde{\mu}$ be the intersection of the boundary circle of C with the geodesic path from μ to p, that is,

$$\tilde{\mu} = (0, \ldots, 0, (1 - h^2)^{1/2}, h). \tag{4.8.12}$$

Fix an arbitrary $x = (x_1, \ldots, x_d) \in C$ and note that

$$\tilde{\mu}'x - \mu'x = ((1 - h^2)^{1/2} - (1 - \mu_d^2)^{1/2})x_{d-1} + (h - \mu_d)x_d. \tag{4.8.13}$$

Irrespective of the sign of x_{d-1}'s coefficient in (4.8.13), we have

$$\begin{aligned}\tilde{\mu}'x - \mu'x &\geq -|(1 - h^2)^{1/2} - (1 - \mu_d^2)^{1/2}|(1 - x_d^2)^{1/2} + (h - \mu_d)x_d \\ &\geq -|(1 - h^2)^{1/2} - (1 - \mu_d^2)^{1/2}|(1 - h^2)^{1/2} + (h - \mu_d)h. \end{aligned} \tag{4.8.14}$$

Since $|\mu_d| < |h|$, this yields

$$\begin{aligned}\tilde{\mu}'x - \mu'x &\geq ((1 - h^2)^{1/2} - (1 - \mu_d^2)^{1/2})(1 - h^2)^{1/2} + (h - \mu_d)h \\ &= (1 - \mu_d h) - (1 - \mu_d^2)^{1/2}(1 - h^2)^{1/2} \\ &> 0,\end{aligned}$$

where the last inequality results from the fact that it is easy to check that $(1 - \mu_d h)^2 > (1 - \mu_d^2)(1 - h^2)$ if and only if $\mu_d \neq h$. Thus, we proved that $\tilde{\mu}'x > \mu'x$ for any $x \in C$. Since $t \mapsto \arccos(t)$ is monotone decreasing on $[-1, 1]$, we then have

$$d(x, \tilde{\mu}) = \arccos(\tilde{\mu}'x) < \arccos(\mu'x) = d(x, \mu)$$

for any $x \in C$. For any probability measure Q that is concentrated on C, this yields $g_Q(\tilde{\mu}) < g_Q(\mu)$, as was to be shown. (b) $|\mu_d| \geq h$, so that $\mu_d \leq -h$ (since $\mu \notin C$, we have $\mu_d < h$). Let then $\tilde{\mu}$ be the reflection of μ with respect to the $(d-1)$-dimensional hyperplane orthogonal to p, that is,

$$\tilde{\mu} = (0, \ldots, 0, (1 - \mu_d^2)^{1/2}, -\mu_d).$$

For any $x = (x_1, \ldots, x_d) \in C$, we have $\tilde{\mu}'x - \mu'x = -2\mu_d x_d = 2|\mu_d|x_d > 0$, so that the same argument as in case (a) shows that $g_Q(\tilde{\mu}) < g_Q(\mu)$ for any probability measure Q that is concentrated on C.

For any $\mu \in \mathcal{S}^{d-1} \backslash C$ of the form (4.8.11), we have thus identified a corresponding $\tilde{\mu} \in \mathcal{S}^{d-1}$ such that $g_Q(\tilde{\mu}) < g_Q(\mu)$ for any probability measure Q that is concentrated on C. Fix then an arbitrary $\xi \in \mathcal{S}^{d-1} \setminus C$. Pick a $d \times d$ orthogonal matrix O such that $Op = p$ and $O\xi = \mu$, where μ is of the form given in (4.8.11) (so that $\mu_d = \xi' p$). Denoting as P_O the distribution of OX when X has distribution P, we have that $P_O[C] = P[C] = 1$, so that

$$g_P(\xi) = g_{P_O}(\mu) > g_{P_O}(\tilde{\mu}) = g_P(O'\tilde{\mu}).$$

This shows that ξ cannot be a Fréchet median of P. ■

Corollary 4.8.3. *Let $P \in \mathcal{P}_{d-1}$ and m be Fréchet median of P. Assume that P is concentrated on a circle $\mathcal{C}$ containing $-m$. Then, $\mathcal{C}$ is a great circle containing m.*

Proof of Corollary 4.8.3. Since any great circle containing $-m$ also contains m, it is sufficient to establish that $\mathcal{C}$ is a great circle. Assume, ad absurdum, that $\mathcal{C}$ is not a great circle. Thus, $\mathcal{C}$ is contained in some spherical cap $C := \{x \in \mathcal{S}^{d-1} : p'x \geq h\}$, with $p \in \mathcal{S}^{d-1}$ and $h > 0$. Since $-m \in C$, we have $p'm \leq -h < h$, so that $m \notin C$. Since P is concentrated on $\mathcal{C}$, hence on C, Lemma 4.8.2 ensures that $m \in C$, a contradiction. ■

Lemma 4.8.4. *Fix $d \geq 3$ and let $\mu \in \mathcal{S}^{d-1}$. Then, for any line $\mathcal{L}$ in $T_\mu \mathcal{S}^{d-1}$, the set $\pi_\mu^{-1}(\mathcal{L}) \cup \{-\mu\}$ is a circle on $\mathcal{S}^{d-1}$.*

Proof of Lemma 4.8.4. Since $\mathcal{L}$ is included in $T_\mu \mathcal{S}^{d-1}$, it does not contain $-\mu$, so that there is a unique plane, Π say, containing $\mathcal{L}$ and $-\mu$. Clearly, $\pi_\mu^{-1}(\mathcal{L}) \cup \{-\mu\}$ is then the intersection between $\mathcal{S}^{d-1}$ and Π. Since Π is a two-dimensional affine subspace of $\mathbb{R}^d$, there exist $\mu_1, \ldots, \mu_{d-2} \in \mathcal{S}^{d-1}$ and $h_1, \ldots, h_{d-2} \in \mathbb{R}$ such that $x \in \Pi$ if and only if

$$\mu_\ell' x = h_\ell, \quad \ell = 1, \ldots, d-2. \tag{4.8.15}$$

Note that the matrix $M = (\mu_1 \ldots \mu_{d-2})$ must have full rank (otherwise, the $d-2$ hyperplanes described by (4.8.15) either do not intersect or have an intersection that has a dimension strictly larger than 2). Now, pick the vector $\lambda = (\lambda_1, \ldots, \lambda_{d-2})' \in \mathbb{R}^{d-2}$ such that

$$q = \sum_{\ell=1}^{d-2} \lambda_\ell \mu_\ell = M\lambda$$

belongs to Π (writing $h = (h_1, \ldots, h_{d-2})'$, existence and uniqueness of λ follows from the fact that this imposes that $h = M'q = (M'M)\lambda$, where the square matrix $M'M$ has full rank $d-2$). Note that if $x \in \Pi \cap \mathcal{S}^{d-1}$, then

$$(x-q)'(x-q) = x'x - 2(M\lambda)'x + (M\lambda)'(M\lambda) = 1 - h'(M'M)^{-1}h.$$

Therefore, the intersection $\Pi \cap \mathcal{S}^{d-1}$ is the collection of $x \in \mathbb{R}^d$ satisfying

$$\begin{cases} \mu_\ell' x = h_\ell, \ \ell = 1, \ldots, d-2, \\ \|x - q\| = \sqrt{1 - h'(M'M)^{-1}h}\,(=: r), \end{cases}$$

which, as the intersection between the sphere centered at q with radius r and the plane Π that contains the center of this sphere, is a circle in Π (note that r must be well-defined above; otherwise $\Pi \cap \mathcal{S}^{d-1}$ would be void, which would contradict the fact that $-\mu$ belongs to the intersection). ■

Proofs for Section 4.2

The proof of Theorem 4.2.3 requires the following result.

Lemma 4.8.5. *Let $v, w \in \mathbb{R}^d \setminus \{0\}$. Then*

$$\left\| \frac{v}{\|v\|} - \frac{w}{\|w\|} \right\| \leq 2 \min\left(\frac{\|v-w\|}{\|v\|}, \frac{\|v-w\|}{\|w\|} \right).$$

Proof of Theorem 4.2.3. Fix $\alpha \in [0,1)$ and a unit vector $u \in T_m\mathcal{S}^{d-1}$. We first prove that, as $\ell \to \infty$, $M_{\alpha,u}^{m,P_\ell}$ converges to $p_m M_{\alpha/p_m,u}^{m,P_{-m}}$ uniformly on compacts sets of $\mathcal{S}_{-m}^{d-1}$, where $M_{\alpha,u}^{m,P}$ is the objective function in Definition 4.2.2. Equivalently, we prove that, for any $r > 0$,

$$c_{\ell,r} := \sup_{\mu \in \pi_m^{-1}(B_r)} \left| M_{\alpha,u}^{m,P_\ell}(\mu) - p_m M_{\alpha/p_m,u}^{m,P_{-m}}(\mu) \right| = o(1) \tag{4.8.16}$$

as $\ell \to \infty$, where we let $B_r := \{z \in T_m\mathcal{S}^{d-1} : \|z\| \leq r\}$. To do so, let Y be an $\mathcal{S}_{-m}^{d-1}$-valued random vector with distribution P_{-m} and, for any ℓ, let Y_ℓ be an $\mathcal{S}^{d-1}$-valued random vector with distribution Q_ℓ. Since these distributions do not have an atom at $-m$, the corresponding projections onto $T_m\mathcal{S}^{d-1}$, namely $Z := \pi_m(Y)$ and $Z_\ell := \pi_m(Y_\ell)$, are well-defined with probability one.

Before proceeding, we make a couple of comments on the Z_ℓ's. First note that Condition (ii) in the statement of the theorem implies that, with probability one,

$$\|Z_\ell\| = \frac{\|Y_\ell - (m'Y_\ell)m\|}{1 + m'Y_\ell} = \sqrt{\frac{1 - m'Y_\ell}{1 + m'Y_\ell}} \geq \sqrt{\frac{1 - \cos c}{1 + m'Y_\ell}}. \tag{4.8.17}$$

Since $y \mapsto \sqrt{(1 + m'y)/(1 - \cos c)}$ is a continuous and bounded mapping on $\mathcal{S}^{d-1}$, weak convergence of Q_ℓ to the Dirac measure δ_{-m} at $-m$ thus implies that

$$\begin{aligned} \mathrm{E}\left[\frac{1}{\|Z_\ell\|} \mathbb{I}[Z_\ell \neq 0] \right] &\leq \int_{\mathcal{S}^{d-1}} \sqrt{\frac{1 + m'y}{1 - \cos c}}\, dQ_\ell(y) \\ &\to \int_{\mathcal{S}^{d-1}} \sqrt{\frac{1 + m'y}{1 - \cos c}}\, d\delta_{-m}(y) = 0, \end{aligned} \tag{4.8.18}$$

as $\ell \to \infty$ (finiteness of $\mathrm{E}[(1/\|Z_\ell\|)\mathbb{I}[Z_\ell \neq 0]]$ for any ℓ obviously follows from (4.8.17)). Finally, note that since Z_ℓ takes its values in $T_m\mathcal{S}^{d-1}$ and is such that OZ_ℓ and Z_ℓ are equal in distribution for any $d \times d$ orthogonal transformation fixing m (this readily follows from the rotational symmetry of Q_ℓ), we have

$$\mathrm{E}\bigg[\frac{Z_\ell}{\|Z_\ell\|}\mathbb{I}[Z_\ell \neq 0]\bigg] = 0, \quad \ell = 1, 2, \ldots \tag{4.8.19}$$

Now, aiming at proving (4.8.16), observe that, for any $\mu \in \mathcal{S}^{d-1}_{-m}$ and any ℓ,

$$\begin{aligned}
M^{m,P_\ell}_{\alpha,u}(\mu) &= p_m M^{m,P_{-m}}_{\alpha,u}(\mu) + (1-p_m) M^{m,Q_\ell}_{\alpha,u}(\mu)\\
&= p_m \mathrm{E}[\|Z - \pi_m(\mu)\| - \|Z\| - \alpha u'\pi_m(\mu)]\\
&\qquad\qquad + (1-p_m)\mathrm{E}[\|Z_\ell - \pi_m(\mu)\| - \|Z_\ell\| - \alpha u'\pi_m(\mu)]\\
&= p_m \mathrm{E}[\|Z - \pi_m(\mu)\| - \|Z\|] - \alpha u'\pi_m(\mu)\\
&\qquad\qquad + (1-p_m)\mathrm{E}[\|Z_\ell - \pi_m(\mu)\| - \|Z_\ell\|]\\
&= p_m M^{m,P_{-m}}_{\alpha/p_m,u}(\mu) + (1-p_m)\mathrm{E}[\|Z_\ell - \pi_m(\mu)\| - \|Z_\ell\|].
\end{aligned} \tag{4.8.20}$$

For any $\mu \in \mathcal{S}^{d-1}_{-m}$, let $S_\mu := \{\lambda \pi_m(\mu) : \lambda \in [0,1]\}$ be the line segment from 0 to $\pi_m(\mu)$ in $T_m\mathcal{S}^{d-1}$. Since

$$\frac{d}{dt}\|z - t\pi_m(\mu)\| = -\frac{(\pi_m(\mu))'(z - t\pi_m(\mu))}{\|z - t\pi_m(\mu)\|}$$

as soon as $z - t\pi_m(\mu) \neq 0$, the fundamental theorem of calculus yields that, for any $\mu \in \mathcal{S}^{d-1}_{-m}$ and any $z \in \mathbb{R}^d \setminus S_\mu$,

$$\|z - \pi_m(\mu)\| - \|z\| = -\int_0^1 \frac{(\pi_m(\mu))'(z - t\pi_m(\mu))}{\|z - t\pi_m(\mu)\|}\, dt.$$

Therefore,

$$\begin{aligned}
&\mathrm{E}[(\|Z_\ell - \pi_m(\mu)\| - \|Z_\ell\|)\mathbb{I}[Z_\ell \notin S_\mu]]\\
&\quad = -\mathrm{E}\bigg[\bigg(\int_0^1 \frac{(\pi_m(\mu))'(Z_\ell - t\pi_m(\mu))}{\|Z_\ell - t\pi_m(\mu)\|}\, dt\bigg)\mathbb{I}[Z_\ell \notin S_\mu]\bigg]\\
&\quad = -(\pi_m(\mu))'\mathrm{E}\bigg[\frac{Z_\ell}{\|Z_\ell\|}\mathbb{I}[Z_\ell \notin S_\mu]\bigg]\\
&\qquad\qquad - \mathrm{E}\bigg[\bigg(\int_0^1 (\pi_m(\mu))'\bigg(\frac{(Z_\ell - t\pi_m(\mu))}{\|Z_\ell - t\pi_m(\mu)\|} - \frac{Z_\ell}{\|Z_\ell\|}\bigg)\, dt\bigg)\mathbb{I}[Z_\ell \notin S_\mu]\bigg],
\end{aligned}$$

so that Lemma 4.8.5 provides

$$\begin{aligned}
&\bigg|\mathrm{E}[(\|Z_\ell - \pi_m(\mu)\| - \|Z_\ell\|)\mathbb{I}[Z_\ell \notin S_\mu]] + (\pi_m(\mu))'\mathrm{E}\bigg[\frac{Z_\ell}{\|Z_\ell\|}\mathbb{I}[Z_\ell \notin S_\mu]\bigg]\bigg|\\
&\qquad \leq 2\mathrm{E}\bigg[\bigg(\int_0^1 \frac{t\|\pi_m(\mu)\|}{\|Z_\ell\|}\, dt\bigg)\mathbb{I}[Z_\ell \notin S_\mu]\bigg] \leq \|\pi_m(\mu)\|\mathrm{E}\bigg[\frac{1}{\|Z_\ell\|}\mathbb{I}[Z_\ell \neq 0]\bigg]
\end{aligned} \tag{4.8.21}$$

Using (4.8.19) and the fact that $P[Z_\ell = 0] = 0$ by assumption, we then have that

$$\begin{aligned}
&\big|\mathrm{E}[\|Z_\ell - \pi_m(\mu)\| - \|Z_\ell\|]\big| \\
&\quad = \left|\mathrm{E}[\|Z_\ell - \pi_m(\mu)\| - \|Z_\ell\|] + (\pi_m(\mu))'\mathrm{E}\left[\frac{Z_\ell}{\|Z_\ell\|}\mathbb{I}[Z_\ell \neq 0]\right]\right| \\
&\quad \leq \left|\mathrm{E}[(\|Z_\ell - \pi_m(\mu)\| - \|Z_\ell\|)\mathbb{I}[Z_\ell \notin S_\mu]] + (\pi_m(\mu))'\mathrm{E}\left[\frac{Z_\ell}{\|Z_\ell\|}\mathbb{I}[Z_\ell \notin S_\mu]\right]\right| \\
&\quad + \left|\mathrm{E}[(\|Z_\ell - \pi_m(\mu)\| - \|Z_\ell\|)\mathbb{I}[Z_\ell \in S_\mu\backslash\{0\}]] + (\pi_m(\mu))'\mathrm{E}\left[\frac{Z_\ell}{\|Z_\ell\|}\mathbb{I}[Z_\ell \in S_\mu\backslash\{0\}]\right]\right| \\
&\quad \leq \|\pi_m(\mu)\|\mathrm{E}\left[\frac{1}{\|Z_\ell\|}\mathbb{I}[Z_\ell \neq 0]\right] + 2\|\pi_m(\mu)\|P[Z_\ell \in S_\mu\backslash\{0\}], \qquad (4.8.22)
\end{aligned}$$

where the last inequality follows from using (4.8.21), the triangle inequality and the Cauchy–Schwarz inequality. Now, Markov's inequality yields that, for any $\mu \in \mathcal{S}^{d-1}_{-m} \setminus \{m\}$,

$$\begin{aligned}
P[Z_\ell \in S_\mu\backslash\{0\}] &\leq P[0 < \|Z_\ell\| \leq \|\pi_m(\mu)\|] \\
&= P\left[\frac{1}{\|Z_\ell\|}\mathbb{I}[Z_\ell \neq 0] \geq \frac{1}{\|\pi_m(\mu)\|}\right] \leq \|\pi_m(\mu)\|E\left[\frac{1}{\|Z_\ell\|}\mathbb{I}[Z_\ell \neq 0]\right]; \quad (4.8.23)
\end{aligned}$$

note that since $S_\mu\backslash\{0\} = \emptyset$ for $\mu = m$, (4.8.23) actually holds for any $\mu \in \mathcal{S}^{d-1}_{-m}$. Therefore, (4.8.22) yields

$$\big|\mathrm{E}[\|Z_\ell - \pi_m(\mu)\| - \|Z_\ell\|]\big| \leq (\|\pi_m(\mu)\| + 2\|\pi_m(\mu)\|^2)\mathrm{E}\left[\frac{1}{\|Z_\ell\|}\mathbb{I}[Z_\ell \neq 0]\right]$$

for any $\mu \in \mathcal{S}^{d-1}_{-m}$. From (4.8.20), this provides

$$\begin{aligned}
&\left|M^{m,P_\ell}_{\alpha,u}(\mu) - p_m M^{m,P_{-m}}_{\alpha/p_m,u}(\mu)\right| \\
&\qquad \leq (1-p_m)(\|\pi_m(\mu)\| + 2\|\pi_m(\mu)\|^2)\mathrm{E}\left[\frac{1}{\|Z_\ell\|}\mathbb{I}[Z_\ell \neq 0]\right],
\end{aligned}$$

which, in view of (4.8.18), establishes (4.8.16).

We can now proceed with the proof of (4.2.4), which we split into two cases. (A) Consider first the case $\alpha \in [0, p_m)$. For such a value of α, Lemma 3.9.6 of Chapter 3 guarantees the existence of positive constants C and R such that

$$M^{m,P_{-m}}_{\alpha/p_m,u}(\mu) > C\|\pi_m(\mu)\| \qquad (4.8.24)$$

for any $\mu \notin \pi_m^{-1}(B_R)$. In particular,

$$M^{m,P_{-m}}_{\alpha/p_m,u}(\mu) > 0 = M^{m,P_{-m}}_{\alpha/p_m,u}(m)$$

for any $\mu \notin \pi_m^{-1}(B_R)$. Since the (unique; see Theorem 4.3.1, whose proof is independent of the present proof) quantile $\mu^m_{\alpha,u}(P)$ minimizes $\mu \mapsto M^{m,P_{-m}}_{\alpha/p_m,u}(\mu)$ by definition, this implies that $\mu^m_{\alpha,u}(P)$ belongs to B_R.

We need to ensure that all quantiles $\mu^m_{\alpha,u}(P_\ell)$ belong to B_R, too. Before doing so, note that since Q_ℓ is rotationally symmetric on $\mathcal{S}^{d-1}$ and satisfies $Q_\ell[\mathcal{S}^{d-1}\setminus\{\pm m\}]=1$, the probability measures Q_ℓ are not concentrated on a great circle of $\mathcal{S}^{d-1}$ containing m. Since this trivially extends to P_ℓ, Theorem 4.3.1(ii) ensures uniqueness of $\mu^m_{\alpha,u}(P_\ell)$ for any ℓ. Now, for any $\mu\in\mathcal{S}^{d-1}_{-m}\setminus\{m\}$,

$$\frac{Z_\ell}{\|Z_\ell\|}\mathbb{I}[Z_\ell\in S_\mu\backslash\{0\}]=\frac{\pi_m(\mu)}{\|\pi_m(\mu)\|}\mathbb{I}[Z_\ell\in S_\mu\backslash\{0\}]$$

almost surely, so that the triangle inequality yields

$$\begin{aligned}\mathrm{E}[(\|Z_\ell-\pi_m(\mu)\|-\|Z_\ell\|)\mathbb{I}[Z_\ell\in S_\mu\backslash\{0\}]]+(\pi_m(\mu))'\mathrm{E}\bigg[\frac{Z_\ell}{\|Z_\ell\|}\mathbb{I}[Z_\ell\in S_\mu\backslash\{0\}]\bigg]\\=\mathrm{E}\big[(\|Z_\ell-\pi_m(\mu)\|-\|Z_\ell\|+\|\pi_m(\mu)\|)\mathbb{I}[Z_\ell\in S_\mu\backslash\{0\}]\big]\geq 0\end{aligned}$$

(which also trivially holds for $\mu=m$ since we then have $S_\mu\backslash\{0\}=\emptyset$). This entails that, for any $\mu\in\mathcal{S}^{d-1}_{-m}$,

$$\begin{aligned}&\mathrm{E}[\|Z_\ell-\pi_m(\mu)\|-\|Z_\ell\|]\\&\quad=\mathrm{E}[(\|Z_\ell-\pi_m(\mu)\|-\|Z_\ell\|)]+(\pi_m(\mu))'E\bigg[\frac{Z_\ell}{\|Z_\ell\|}\mathbb{I}[Z_\ell\neq 0]\bigg]\\&\quad\geq\mathrm{E}[(\|Z_\ell-\pi_m(\mu)\|-\|Z_\ell\|)\mathbb{I}[Z_\ell\notin S_\mu]]+(\pi_m(\mu))'\mathrm{E}\bigg[\frac{Z_\ell}{\|Z_\ell\|}\mathbb{I}[Z_\ell\notin S_\mu]\bigg]\\&\quad\geq-\|\pi_m(\mu)\|\mathrm{E}\bigg[\frac{1}{\|Z_\ell\|}\mathbb{I}[Z_\ell\neq 0]\bigg]=:-\|\pi_m(\mu)\|\varepsilon_\ell,\end{aligned}$$

where the last inequality follows from (4.8.21). Using (4.8.20) and (4.8.24), this provides

$$\begin{aligned}M^{m,P_\ell}_{\alpha,u}(\mu)&\geq p_mM^{m,P_{-m}}_{\alpha/p_m,u}(\mu)-(1-p_m)\|\pi_m(\mu)\|\varepsilon_\ell\\&>\big(Cp_m-(1-p_m)\varepsilon_\ell\big)\|\pi_m(\mu)\|\end{aligned}$$

for any $\mu\notin\pi_m^{-1}(B_R)$. Thus, with ℓ_0 large enough so that $(1-p_m)\varepsilon_\ell\leq Cp_m$ (such an ℓ_0 exists since $p_m>0$ and $\varepsilon_\ell\to 0$ as $\ell\to\infty$), we have that, for any $\ell\geq\ell_0$, $M^{m,P_\ell}_{\alpha,u}(\mu)>0=M^{m,P_\ell}_{\alpha,u}(m)$ for any $\mu\notin\pi_m^{-1}(B_R)$. This implies that, for any $\ell\geq\ell_0$, the quantile $\mu^m_{\alpha,u}(P_\ell)$, that minimizes $\mu\mapsto M^{m,P_\ell}_{\alpha,u}(\mu)$, belongs to $\pi_m^{-1}(B_R)$.

We can now conclude in the present case where $\alpha\in[0,p_m)$. To make the notation lighter, write $q_\ell:=\mu^m_{\alpha,u}(P_\ell)$. Let then (q_{ℓ_k}) be an arbitrary subsequence of (q_ℓ). Since the latter sequence takes values in $\pi_m^{-1}(B_R)$ for $\ell\geq\ell_0$, compactness of $\pi_m^{-1}(B_R)$ entails that (q_{ℓ_k}) admits a subsequence $(q_{\ell_{k_j}})$ taking values in $\pi_m^{-1}(B_R)$ and converging in $\pi_m^{-1}(B_R)$, to q say. For any $\mu\in\mathcal{S}^{d-1}_{-m}$, we then have (see (4.8.16) for the definition of $c_{\ell,r}$)

$$\begin{aligned}&p_mM^{m,P_{-m}}_{\alpha/p_m,u}(q_{\ell_{k_j}})-c_{\ell_{k_j},R}\\&\quad\leq M^{m,P_{\ell_{k_j}}}_{\alpha,u}(q_{\ell_{k_j}})\leq M^{m,P_{\ell_{k_j}}}_{\alpha,u}(\mu)\leq p_mM^{m,P_{-m}}_{\alpha/p_m,u}(\mu)+c_{\ell_{k_j},\|\mu\|}.\end{aligned}$$

From (4.8.16) and the continuity of $\mu \mapsto M^{m,P_{-m}}_{\alpha/p_m,u}(\mu)$ over $\pi_m^{-1}(B_R)$ (continuity results from Lemma 3.9.5 in Chapter 3 and from the diffeomorphic nature of π_m), taking limits as $j \to \infty$ then yields that

$$M^{m,P_{-m}}_{\alpha/p_m,u}(q) \leq M^{m,P_{-m}}_{\alpha/p_m,u}(\mu)$$

for any $\mu \in \mathcal{S}^{d-1}_{-m}$ (recall that $p_m > 0$). By definition, we then have that $q = \mu^m_{\alpha,u}(P)$ (recall that $\mu^m_{\alpha,u}(P)$ is unique). We thus proved that any subsequence of $(\mu^m_{\alpha,u}(P_\ell))$ admits a further subsequence converging to $\mu^m_{\alpha,u}(P)$. This establishes that $(\mu^m_{\alpha,u}(P_\ell))$ converges to $\mu^m_{\alpha,u}(P)$ as $\ell \to \infty$, as was to be proved.

(B) Second, consider the case $\alpha \in [p_m, 1]$. As $\ell \to \infty$, the map $M^{m,P_\ell}_{\alpha,u}$, as shown in (4.8.16), converges uniformly on compact sets of $\mathcal{S}^{d-1}_{-m}$ to the map $F_{\alpha,u}$ defined by

$$F_{\alpha,u}(\mu) := p_m \mathrm{E}[\|Z - \pi_m(\mu)\| - \|Z\|] - \alpha u' \pi_m(\mu);$$

recall that $Z = \pi_m(Y)$, where the $\mathcal{S}^{d-1}_{-m}$-valued random vector Y has distribution P_{-m}. If $\alpha = p_m$, then $F_{\alpha,u}(\mu) = p_m M^{m,P_{-m}}_{1,u}(\mu)$ for any $\mu \in \mathcal{S}^{d-1}_{-m}$. Note that the distribution of Z is not concentrated on a line. Indeed, would it be concentrated on a line (of $T_m\mathcal{S}^{d-1}$), then Lemma 4.8.4 would entail that P_{-m} is concentrated on a circle containing $-m$, which, in turn, would imply (Corollary 4.8.3) that P_{-m} is concentrated on a great circle containing m. Since this is not the case by assumption, the distribution of Z is indeed not concentrated on a line, so that Lemma 5 in [83] entails that the map $G = F_{\alpha,u} \circ \pi_m^{-1}$ has no minimum on $T_m\mathcal{S}^{d-1}$, which implies that $F_{\alpha,u}$ admits no minimum on $\mathcal{S}^{d-1}_{-m}$. If $\alpha > p_m$, then

$$F_{\alpha,u}(\mu) = \|\pi_m(\mu)\| \left\{ p_m \mathrm{E}\left[\frac{\|Z - \pi_m(\mu)\| - \|Z\|}{\|\pi_m(\mu)\|} \right] - \alpha u' \frac{\pi_m(\mu)}{\|\pi_m(\mu)\|} \right\}$$

for any $\mu \in \mathcal{S}^{d-1}_{-m} \setminus \{m\}$. Let (μ_n) be the sequence in $\mathcal{S}^{d-1}_{-m} \setminus \{m\}$ such that $\pi_m(\mu_n) = nu$. A routine application of Lebesgue's dominated convergence theorem then yields that

$$\mathrm{E}\left[\frac{\|Z - \pi_m(\mu_n)\| - \|Z\|}{\|\pi_m(\mu_n)\|} \right] \to 1,$$

hence that

$$\frac{F_{\alpha,u}(\mu_n)}{n} = \frac{F_{\alpha,u}(\mu_n)}{\|\pi_m(\mu_n)\|} \to p_m - \alpha < 0.$$

Thus, $F_{\alpha,u}(\mu_n)$ diverges to $-\infty$, which shows that, also for $\alpha > p_m$, the map $F_{\alpha,u}$ does not admit a minimum on $\mathcal{S}^{d-1}_{-m}$.

Fix then $\alpha \geq p_m$ and a unit vector u in $T_m\mathcal{S}^{d-1}$, and consider the corresponding quantile sequence $(q_\ell := \mu^m_{\alpha,u}(P_\ell))$ (recall that these quantiles are unique under the assumptions of the theorem). Let us assume that $(\|\pi_m(q_\ell)\|)$ does not converge to infinity as ℓ does. Up to extracting a subsequence, we may assume that $(\pi_m(q_\ell))$ is bounded, hence, upon extraction of a further subsequence, that $(\pi_m(q_\ell))$ converges in $T_m\mathcal{S}^{d-1}$, to z_0 say. For any $\mu \in \mathcal{S}^{d-1}_{-m}$, we then have, for ℓ large enough, that

$$F_{\alpha,u}(q_\ell) - c_{\ell,2\|z_0\|} \leq M^{m,P_\ell}_{\alpha,u}(q_\ell) \leq M^{m,P_\ell}_{\alpha,u}(\mu) \leq F_{\alpha,u}(\mu) + c_{\ell,\|\mu\|}.$$

Taking limits as $\ell \to \infty$, we obtain that $F_{\alpha,u}(\pi_m^{-1}(z_0)) \leq F_{\alpha,u}(\mu)$ for any $\mu \in \mathcal{S}^{d-1}_{-m}$, which contradicts the fact that $F_{\alpha,u}$ has no minimum over $\mathcal{S}^{d-1}_{-m}$. We conclude that $(\|\pi_m(q_\ell)\|)$ diverges to infinity, hence that (q_ℓ) converges to $-m = \mu^m_{\alpha,u}(P)$, as was to be proved. ■

Proofs for Section 4.3

Proof of Theorem 4.3.1. Since existence and uniqueness trivially hold if $\alpha \in [p_m, 1]$, we restrict below to $\alpha \in [0, p_m)$. (i) For such a value of α, a quantile $\mu^m_{\alpha,u}$ exists if and only if $\mu \mapsto M^{m,P}_{\alpha,u}(\mu)$ has a global minimizer over $\mathcal{S}^{d-1}_{-m}$, which is the case if and only if

$$z \mapsto \int_{\mathcal{S}^{d-1}_{-m}} \big\{ \|\pi_m(x) - z\| - \|\pi_m(x)\| - \alpha u' z / p_m \big\} \, dP_{-m}(x)$$

$$= \int_{\mathbb{R}^d} \big\{ \|y - z\| - \|y\| - \alpha u' z / p_m \big\} \, d(\pi_m \# P_{-m})(y) \qquad (4.8.25)$$

has a global minimizer over $T_m\mathcal{S}^{d-1}$, where $\pi_m \# P_{-m}$, the pushforward probability measure of P_{-m} by π_m, is seen as a probability measure over $\mathbb{R}^d$ that is concentrated on $T_m\mathcal{S}^{d-1}$. Now, since $\alpha/p_m \in [0,1)$, the objective function (4.8.25) admits at least one global minimizer in $\mathbb{R}^d$, that is a (17) geometric quantile of order α/p_m in direction u for $\pi_m \# P_{-m}$; see, e.g., Theorem 1(i) in [83]. In view of Lemma 4.8.1, any such geometric quantile actually belongs to $T_m\mathcal{S}^{d-1}$, hence provides a global minimizer of (4.8.25) in $T_m\mathcal{S}^{d-1}$. (ii) Uniqueness of $\mu^m_{\alpha,u}$ then holds if and only if $\pi_m \# P_{-m}$ has a unique geometric quantile of order α/p_m in direction u. In view of Theorem 1(ii) in [83], this is the case in particular if $\pi_m \# P_{-m}$ is not concentrated on any line of $\mathbb{R}^d$, or equivalently, on any line of $T_m\mathcal{S}^{d-1}$. The result then follows from Lemma 4.8.4 and Corollary 4.8.3. (iii) Using the same reasoning as in (ii), the result follows from Theorem 1(iii)–(iv) in [83]. ■

The proof of of Theorem 4.3.2 requires the following result.

Lemma 4.8.6. *Fix $P \in \mathcal{P}_{d-1}$, $\mu \in \mathcal{S}^{d-1}$ and a unit vector $v \in T_\mu\mathcal{S}^{d-1}$. Define the directional derivative of g_P in (4.2.2) at μ in direction v as*

$$\frac{\partial g_P}{\partial v}(\mu) = \lim_{t \overset{>}{\to} 0} \frac{g_P(\varphi(t)) - g_P(\mu)}{t},$$

where $\varphi : [0,\pi] \to \mathcal{S}^{d-1} : t \mapsto (\cos t)\mu + (\sin t)v$ is the geodesic path from $\varphi(0) = \mu$ to $\varphi(\pi) = -\mu$ such that $\dot\varphi(0) = v$. Then,

$$\frac{\partial g_P}{\partial v}(\mu) = -v' \mathrm{E}\left[\frac{(I_d - \mu\mu')X}{\|(I_d - \mu\mu')X\|} \xi_{X,\mu}\xi_{X,-\mu} \right] + P[\{\mu\}] - P[\{-\mu\}],$$

where X is a random vector with distribution P and where we let $\xi_{x,y} := \mathbb{I}[x \neq y]$.

Note that if m is a Fréchet median of P, then the directional derivative at m must be non-negative for any unit vector v in $T_m\mathcal{S}^{d-1}$, which implies that $P[\{m\}] - P[\{-m\}] \geq 0$, as was already mentioned in Section 4.2.2.

Proof of Lemma 4.8.6. Since $d(\varphi(t),\mu)=t$ for any $t\in[0,\pi]$, we have

$$d(\varphi(t),\mu)-d(\varphi(0),\mu)=t-0=t$$

and

$$d(\varphi(t),-\mu))-d(\varphi(0),-\mu)=(\pi-t)-\pi=-t,$$

which yields

$$\begin{aligned}\frac{g_P(\varphi(t))-g_P(\mu)}{t}&=\int_{\mathcal{S}^{d-1}}\frac{d(\varphi(t),x)-d(\varphi(0),x)}{t}\,dP(x)\\&=\int_{\mathcal{S}^{d-1}\setminus\{-\mu,\mu\}}\frac{d(\varphi(t),x)-d(\varphi(0),x)}{t}\,dP(x)+P[\{\mu\}]-P[\{-\mu\}].\end{aligned}$$

Since the triangle inequality yields

$$\left|\frac{d(\varphi(t),x)-d(\varphi(0),x)}{t}\right|\leq\frac{d(\varphi(t),\varphi(0))}{t}=1$$

for any $x\in\mathcal{S}^{d-1}$ and $t\in(0,\pi]$, Lebesgue's dominated convergence theorem ensures that

$$\frac{\partial g_P}{\partial v}(\mu)=\int_{\mathcal{S}^{d-1}\setminus\{-\mu,\mu\}}\lim_{t\overset{>}{\to}0}\frac{d(\varphi(t),x)-d(\varphi(0),x)}{t}\,dP(x)+P[\{\mu\}]-P[\{-\mu\}],$$

provided that the limit in this integral exists for any $x\in\mathcal{S}^{d-1}\setminus\{-\mu,\mu\}$. Since a direct computation provides

$$\lim_{t\overset{>}{\to}0}\frac{d(\varphi(t),x)-d(\varphi(0),x)}{t}=-\frac{v'x}{\sqrt{1-(\mu'x)^2}}=-v'\frac{(I_d-\mu\mu')x}{\|(I_d-\mu\mu')x\|}$$

for any such x, the proof is complete. ∎

Proof of Theorem 4.3.2. Ad absurdum, assume that the Fréchet median m is not an m-geometric median for P. By definition, this means that $\pi_m(m)=0$ is not a geometric median of $\pi_m(X)$, where X is a random vector with distribution P_{-m}. Since any geometric median of $\pi_m(X)$ belongs to $T_m\mathcal{S}^{d-1}$ (Lemma 4.8.1) and since the objective function $\mu\mapsto O(\mu)=O^P_{0,u}(\mu)=\mathrm{E}[\|\pi_m(X)-\mu\|-\|\pi_m(X)\|]$ (see (4.1.1)) defining geometric medians is convex, there must then exist a unit vector v in $T_m\mathcal{S}^{d-1}$ such that

$$\frac{\partial O}{\partial v}(0)<0$$

(would this partial derivative be nonnegative for any such v, then convexity would indeed guarantee that 0 is a geometric median of $\pi_m(X)$). From Proposition 3.5.1 in Chapter 3, this yields

$$\frac{\partial O}{\partial v}(0)=P[\pi_m(X)=0]+v'\mathrm{E}\bigg[\frac{0-\pi_m(X)}{\|0-\pi_m(X)\|}\xi_{\pi_m(X),0}\bigg]<0$$

Since $\pi_m(X)=0$ if and only if $X=m$, this rewrites

$$P[\{m\}]-v'\mathrm{E}\bigg[\frac{\pi_m(X)}{\|\pi_m(X)\|}\xi_{X,m}\bigg]<0.$$

Using the expression of π_m and the fact that, by construction, $\xi_{X,-m} = 1$ almost surely, this takes the form

$$P[\{m\}] - v'\mathrm{E}\Big[\frac{(I - mm')X}{\|(I - mm')X\|}\xi_{X,m}\xi_{X,-m}\Big] < 0. \tag{4.8.26}$$

From Lemma 4.8.6, we thus have that

$$\frac{\partial g_P}{\partial v}(m) = P[\{m\}] - v'\mathrm{E}\Big[\frac{(I - mm')X}{\|(I - mm')X\|}\xi_{X,m}\xi_{X,-m}\Big] - P[\{-m\}] < 0,$$

which implies that m is not a minimizer of $\mu \mapsto g_P(\mu) = \mathrm{E}[d(X, \mu)]$. This entails that m is not a Fréchet median of P, a contradiction. ■

Proof of Theorem 4.3.3. (i) Recall that $\alpha_n < p_m$ for any n. Since P is not concentrated on a circle containing $-m$, each quantile $\mu^m_{\alpha_n,u_n}$ then exists and is unique (Theorem 4.3.1(ii)), and it coincides with the image by π_m^{-1} of the geometric quantile, $q_{\alpha_n/p_m,u_n}$ say, of order α_n/p_m and direction u_n for the pushforward probability measure $\pi_m\#P_{-m}$. Since this pushforward measure is not concentatred on a line of $\mathbb{R}^d$ (would it be, then, by Lemma 4.8.4, P would be concentrated on a circle containing $-m$, hence, by Corollary 4.8.3, on a great circle containing m), it follows from Theorem 2 in [83] that $\|q_{\alpha_n/p_m,u_n}\| \to \infty$ as $n \to \infty$. It follows that $\mu^m_{\alpha_n,u_n} = \pi_m^{-1}(q_{\alpha_n/p_m,u_n})$ converges to $-m$ as $n \to \infty$.

(ii) Since $\mu^m_{\alpha_n,u_n} \neq -m$ for $\alpha_n \in [0, p_m)$, there are a unique $t_m \in [0, \pi)$ and a unique unit vector $v_n \in T_m\mathcal{S}^{d-1}$ such that $\mu^m_{\alpha_n,u_n} = (\cos t_n)m + (\sin t_n)v_n$. This implies that

$$q_{\alpha_n/p_m,u_n} = \pi_m(\mu^m_{\alpha_n,u_n}) = \pi_m((\cos t_n)m + (\sin t_n)v_n) = \frac{(\sin t_n)}{1 + \cos t_n}v_n$$

for any n. Part (i) of the result entails that $(t_n) \to \pi$ as $n \to \infty$. Therefore, for n large enough, we obtain that

$$v_n = \frac{q_{\alpha_n/p_m,u_n}}{\|q_{\alpha_n/p_m,u_n}\|} \to u,$$

where the convergence follows from Theorem 3 in [83]. ■

Proof of Theorem 4.3.5. Fix a sequence (z_n) in $\mathcal{B}^\infty_{m,p_m}$ converging to some z (in this proof, all convergences in $\mathcal{B}^\infty_{m,p_m}$ are with respect to the metric δ_{m,p_m}). We consider two cases. (a) Assume first that $z = u^\infty_{m,p_m}$. Since $Q(z) = -m$ by definition, we need to show that $Q(z_n)$ converges to $-m$. Of course, it is sufficient to show that the subsequence $(Q(z_{n_k}))$ obtained from $(Q(z_n))$ by discarding the terms for which $z_n = z$ converges to $Q(z)$. Writing $z_{n_k} = \alpha_{n_k}u_{n_k}$, where $\alpha_{n_k} \in [0, p_m)$ and u_{n_k} is a unit vector in $T_m\mathcal{S}^{d-1}$, we have that $\alpha_{n_k} \to p_m$ as $k \to \infty$ (otherwise, (z_{n_k}) does not converge to z, which is a contradiction). Theorem 4.3.3(i) then entails that $Q(z_{n_k}) \to -m$ as $k \to \infty$, as was to be shown. (b) Assume then that $z \in \mathcal{B}^\infty_{m,p_m} \setminus \{u^\infty_{m,p_m}\} = \mathcal{B}_{m,p_m}$, hence can be written as $z = \alpha u$, where $\alpha \in [0, p_m)$ and u is a unit vector in $T_m\mathcal{S}^{d-1}$. For any $n \geq N$ with N large enough, $z_n \in \mathcal{B}_{m,p_m}$, hence can also be written as $z_n = \alpha_n u_n$. For any $n \geq N$, observe that $Q(\alpha_n u_n)$ is the image by π_m^{-1} of the geometric quantile, q_n

say, of order α_n/p_m in direction u_n for $\pi_m\#P_{-m}$ in $\mathbb{R}^d$. Since P is not concentrated on a circle containing $-m$ (Corollary 4.8.3), then $\pi_m\#P_{-m}$ is not concentrated on a line of $\mathbb{R}^d$ (Lemma 4.8.4), so that all geometric quantiles of $\pi_m\#P_{-m}$ are unique. Therefore, by Proposition 3.6.2 in Chapter 3, the sequence (q_n) converges as $n \to \infty$ to the unique geometric quantile, $q_{\alpha/p_m,u}$ say, of order α/p_m in direction u for $\pi_m\#P_{-m}$. Continuity of π_m^{-1} then yields that $Q(z_n) = Q(\alpha_n u_n) = \pi_m^{-1}(q_n) \to \pi_m^{-1}(q_{\alpha/p_m,u})$ as $n \to \infty$. This shows the result since, by definition, $\pi_m^{-1}(q_{\alpha/p_m,u}) = Q(\alpha u)(= Q(z))$. ■

Proofs for Section 4.4

Lemma 4.8.7. *For any $m \in \mathcal{S}^{d-1}$, the map $\pi_m : \mathcal{S}^{d-1}_{-m} \to T_m\mathcal{S}^{d-1}$ in (4.2.3) is smooth. Moreover, the differential $d\pi_m(\mu) : T_\mu\mathcal{S}^{d-1} \to T_m\mathcal{S}^{d-1}$ of π_m at $\mu(\in \mathcal{S}^{d-1}_{-m})$ is defined by*

$$d\pi_m(\mu)v = \frac{4}{\|m+\mu\|^4}\{(1+m'\mu)I_d - (m+\mu)m'\}v$$

for any $v \in T_\mu\mathcal{S}^{d-1}_{-m}$, and it is invertible.

Proof of Lemma 4.8.7. The map π_m is clearly continuously differentiable on $\mathcal{S}^{d-1}_{-m}$. In order to identify the differential of π_m at $\mu(\in \mathcal{S}^{d-1}_{-m})$, it is therefore enough to choose geodesic paths along which to differentiate. Let then $v \in T_\mu\mathcal{S}^{d-1}$ and $\varphi : [0,\pi] \to \mathcal{S}^{d-1} : t \mapsto (\cos t)\mu + (\sin t)v$ be a geodesic path such that $\varphi(0) = \mu$ and $\dot\varphi(0) = v$. It is easy to see that the derivative of

$$t \mapsto \pi_m(\varphi(t)) = \frac{(\cos t)(\mu - (m'\mu)m) + (\sin t)(v - (m'v)m)}{1 + (\cos t)(m'\mu) + (\sin t)(m'v)}$$

at $t = 0$ is given by

$$\frac{1}{(1+m'\mu)^2}\{(I_d - mm') + ((m'\mu)I_d - \mu m')\}v.$$

Letting $w := m + \mu$, we may rewrite

$$\begin{aligned} A &:= \frac{1}{(1+m'\mu)^2}\{(I_d - mm') + ((m'\mu)I_d - \mu m')\} \\ &= \frac{4}{\|m+\mu\|^4}\big\{(m+\mu)'mI_d - (m+\mu)m'\big\} = \frac{4}{\|w\|^4}(w'mI_d - wm'), \end{aligned}$$

which provides the differential defined in the statement of the theorem. For any $v \in \mathbb{R}^d$, observe that $(w'mI_d - wm')v = (m'w)v - (m'v)w$ is zero if and only if v and w are colinear. It follows that the kernel of A is $\mathrm{Ker}(A) = \mathrm{span}(\{m+\mu\})$, hence has dimension 1. The image of A has thus dimension $d-1$. Since $m'(Av) = 0$ for any $v \in \mathbb{R}^d$, this image is contained in $T_m\mathcal{S}^{d-1}$, which has dimension $d-1$. Therefore, the image of A is $T_m\mathcal{S}^{d-1}$. Further note that, since $\mu \in \mathcal{S}^{d-1}_{-m}$, $\mathrm{Ker}(A) \cap T_\mu\mathcal{S}^{d-1} = \{0\}$. This implies that $d\pi_m(\mu)$ is a one-to-one mapping from $T_\mu\mathcal{S}^{d-1}$ to $T_m\mathcal{S}^{d-1}$. ■

Proof of Theorem 4.4.1. For any $t > 0$ small enough to have $\varphi(t) \in \mathcal{S}^{d-1}_{-m}$, let us write

$$M^{m,P}_{\alpha,u}(\varphi(t)) = \mathrm{E}[\{\|\pi_m(X) - \pi_m(\varphi(t))\| - \|\pi_m(X)\|\}\xi_{X,\mu}]$$
$$- \alpha u'\pi_m(\varphi(t))/p_m + \{\|\pi_m(\mu) - \pi_m(\varphi(t))\| - \|\pi_m(\mu)\|\}P_{-m}[\{\mu\}]; \qquad (4.8.27)$$

recall that $\xi_{x,y} := \mathbb{I}[x \neq y]$. Observe that the map $t \mapsto \pi_m(\varphi(t))$ is differentiable at $t = 0$ with

$$\frac{d}{dt}\pi_m(\varphi(t))\Big|_{t=0} = d\pi_m(\varphi(0))\dot{\varphi}(0) = d\pi_m(\mu)v.$$

This implies that

$$\lim_{t \overset{>}{\to} 0} \frac{u'\pi_m(\varphi(t)) - u'\pi_m(\mu)}{t} = u'd\pi_m(\mu)v$$

and that

$$\lim_{t \overset{>}{\to} 0} \frac{\|\pi_m(\mu) - \pi_m(\varphi(t))\|}{t} = \|d\pi_m(\mu)v\|. \qquad (4.8.28)$$

Let us now show that Lebesgue's dominated convergence theorem applies to the expectation term in (4.8.27). For any $x \in \mathcal{S}^{d-1}_{-m} \setminus \{\mu\}$, the chain rule yields

$$\lim_{t \overset{>}{\to} 0} \frac{\|\pi_m(x) - \pi_m(\varphi(t))\| - \|\pi_m(x) - \pi_m(\mu)\|}{t} = \bigg(\frac{\pi_m(\mu) - \pi_m(x)}{\|\pi_m(\mu) - \pi_m(x)\|}\bigg)'\big(d\pi_m(\mu)v\big).$$

Observe now that, for any $x \in \mathcal{S}^{d-1}_{-m} \setminus \{\mu\}$,

$$\bigg|\frac{\|\pi_m(x) - \pi_m(\varphi(t))\| - \|\pi_m(x) - \pi_m(\mu)\|}{t}\bigg| \leq \frac{\|\pi_m(\varphi(t)) - \pi_m(\mu)\|}{t},$$

which, in view of (4.8.28), is bounded in $(0, a)$ for some $a > 0$. Lebesgue's dominated convergence theorem thus yields that the expectation term in (4.8.27) is differentiable at $t = 0$ and that the corresponding derivative is

$$\mathrm{E}\bigg[\frac{\pi_m(\mu) - \pi_m(X)}{\|\pi_m(\mu) - \pi_m(X)\|}\xi_{X,\mu}\bigg]'\big(d\pi_m(\mu)v\big).$$

We conclude that

$$\lim_{t \overset{>}{\to} 0} \frac{M^{m,P}_{\alpha,u}(\varphi(t)) - M^{m,P}_{\alpha,u}(\mu)}{t}$$

exists and is given by the expression provided in the statement of the theorem. ■

Lemma 4.8.8. *Let $P \in \mathcal{P}_{d-1}$ and let $m \in$ be a Fréchet median of P. Fix $\alpha \in [0, p_m)$ and a unit vector $u \in T_m\mathcal{S}^{d-1}$. Then, $\mu_*(\in \mathcal{S}^{d-1}_{-m})$ is an m-geometric quantile of order α in direction u for P if and only if*

$$\frac{\partial M^{m,P}_{\alpha,u}}{\partial v}(\mu_*) \geq 0 \qquad (4.8.29)$$

for any unit vector v in $T_{\mu_*}\mathcal{S}^{d-1}$.

Proof of Lemma 4.8.8. If μ_* is an m-geometric quantile of order α in direction u for P, then it minimizes the objective function $\mu \mapsto M^{m,P}_{\alpha,u}(\mu)$ over $\mathcal{S}^{d-1}_{-m}$, which implies that (4.8.29) holds for any unit vector v in $T_{\mu_*}\mathcal{S}^{d-1}$. Assume then that (4.8.29) holds for any unit vector v in $T_{\mu_*}\mathcal{S}^{d-1}$. In view of Theorem 4.4.1, this implies that

$$\left(\frac{d\pi_m(\mu_*)v}{\|d\pi_m(\mu_*)v\|}\right)'\left\{p_m\mathrm{E}\left[\frac{\pi_m(\mu_*)-\pi_m(X)}{\|\pi_m(\mu_*)-\pi_m(X)\|}\xi_{X,\mu_*}\right]-\alpha u\right\}+P[\{\mu_*\}]\geq 0$$

for any unit vector v in $T_{\mu_*}\mathcal{S}^{d-1}$, where X is an $\mathcal{S}^{d-1}_{-m}$-valued random vector with distribution P_{-m}; invertibility of $d\pi_m(\mu_*)$ (Lemma 4.8.7) ensures that we may divide by $\|d\pi_m(\mu_*)v\|$. Using invertibility of $d\pi_m(\mu_*)$ again, this entails that

$$w'\left\{p_m\mathrm{E}\left[\frac{\pi_m(\mu_*)-\pi_m(X)}{\|\pi_m(\mu_*)-\pi_m(X)\|}\xi_{X,\mu_*}\right]-\alpha u\right\}+P[\{\mu_*\}]\geq 0$$

for any unit vector $w \in T_m\mathcal{S}^{d-1}$, or equivalently, that

$$w'\mathrm{E}\left[\left(\frac{\pi_m(\mu_*)-\pi_m(X)}{\|\pi_m(\mu_*)-\pi_m(X)\|}-\frac{\alpha}{p_m}u\right)\xi_{X,\mu_*}\right]+(1-\alpha u'w)P_{-m}[\{\mu_*\}]\geq 0$$

for any unit vector $w \in T_m\mathcal{S}^{d-1}$. From Proposition 3.5.1 from Chapter 3 (and the comments thereafter), this shows that

$$\frac{\partial O^{\pi_m\#P_{-m}}_{\alpha/p_m,u}}{\partial w}(\pi_m(\mu_*))\geq 0$$

for any unit vector $w \in T_m\mathcal{S}^{d-1}$. Convexity of $O^{\pi_m\#P_{-m}}_{\alpha/p_m,u}$ and Lemma 4.8.1 then entail that $\pi_m(\mu_*)$ is a (Euclidean) geometric quantile of order α/p_m in direction u for $\pi_m\#P_{-m}$. Definition 4.2.2 thus implies that μ^* is an m-geometric quantile of order α in direction u for P. ∎

Proof of Theorem 4.4.2. Fix $\alpha \in [0, p_m)$ and a unit vector u in $T_m\mathcal{S}^{d-1}$. From Lemma 4.8.8, μ is an m-quantile of order α in direction u for P if and only if

$$\frac{\partial M^{m,P}_{\alpha,u}}{\partial v}(\mu)\geq 0$$

for any unit vector v in $T_\mu\mathcal{S}^{d-1}$. Since $d\pi_m(\mu)$ is a one-to-one linear map from $T_\mu\mathcal{S}^{d-1}$ to $T_m\mathcal{S}^{d-1}$ (Lemma 4.8.7), Theorem 4.4.1 entails that this is the case if and only if

$$\frac{1}{p_m}v'\left\{p_m\mathrm{E}\left[\frac{\pi_m(\mu)-\pi_m(X)}{\|\pi_m(\mu)-\pi_m(X)\|}\xi_{X,\mu}\right]-\alpha u\right\}+\frac{1}{p_m}P[\{\mu\}]\geq 0$$

for any unit vector v in $T_m\mathcal{S}^{d-1}$. Therefore, μ is an m-quantile of order α in direction u for P if and only if

$$-\left\|p_m\mathrm{E}\left[\frac{\pi_m(\mu)-\pi_m(X)}{\|\pi_m(\mu)-\pi_m(X)\|}\xi_{X,\mu}\right]-\alpha u\right\|+P[\{\mu\}]\geq 0,$$

which establishes the result. ∎

Proof of Theorem 4.4.4. Let (μ_n) be a sequence in $\mathcal{S}^{d-1}$ converging to some μ, that obviously belong to $\mathcal{S}^{d-1} = \mathcal{S}^{d-1}_{-m} \cup \{-m\}$. We consider two cases: (a) $\mu \in \mathcal{S}^{d-1}_{-m}$. Then there exists N such that $\mu_n \in \mathcal{S}^{d-1}_{-m}$ for any $n \geq N$. For such values of n, the rank $R(\mu_n)$ is thus given by (4.4.7) and takes its value in $\mathcal{B}_{m,p_m}$; see the paragraph below Definition 4.4.3. Using continuity of π_m, a routine application of Lebesgue's dominated convergence theorem then yields that $R(\mu_n) \to R(\mu)$ as $n \to \infty$. (b) $\mu = -m$. We thus need to show that $R(\mu_n) \to R(\mu) = u^{\infty}_{m,p_m}$. Since $R(-m) = u^{\infty}_{m,p_m}$ by definition, it is sufficient to prove that the subsequence $(R(\mu_{n_k}))$ obtained from $(R(\mu_n))$ by discarding the terms for which $\mu_n = -m$ converges to u^{∞}_{m,p_m}. Since $\mu_{n_k} \in \mathcal{S}^{d-1}_{-m}$ for any k and $\mu_{n_k} \to -m$, we have that $\|\pi_m(\mu_{n_k})\| \to \infty$ and $R(\mu_{n_k})$ is still given by (4.4.7). Consider then an arbitrary subsequence $(\|R(\mu_{n_{k_\ell}})\|)$ of $(\|R(\mu_{n_k})\|)$. From compactness of the unit sphere in $T_m\mathcal{S}^{d-1}$, we can extract a subsequence of $(\pi_m(\mu_{n_{k_\ell}})/\|\pi_m(\mu_{n_{k_\ell}})\|)$ that will converge to a unit vector of $T_m\mathcal{S}^{d-1}$, v say. A routine application of Lebesgue's dominated convergence theorem then shows that the corresponding subsequence of $(R(\mu_{n_{k_\ell}}))$ converges to $p_m v$. Thus, any subsequence of $(\|R(\mu_{n_k})\|)$ admits a further subsequence converging to p_m, which establishes that $\|R(\mu_{n_k})\| \to p_m$ as $k \to \infty$. We conclude that

$$\delta_{m,p_m}(R(\mu_{n_k}), u^{\infty}_{m,p_m}) = p_m - \|R(\mu_{n_k})\| \to 0$$

as $k \to \infty$, which shows that $R(\mu_{n_k}) \to u^{\infty}_{m,p_m} = R(\mu)$ as $k \to \infty$. ■

Proof of Theorem 4.4.5. (i) Since P is not concentrated on a great circle containing m, the quantile function $Q = Q^m_P : \mathcal{B}^{\infty}_{m,p_m} \to \mathcal{S}^{d-1}$ and the rank function $R = R^m_P : \mathcal{S}^{d-1} \to \mathcal{B}^{\infty}_{m,p_m}$ are well-defined. Fix then $\mu \in \mathcal{S}^{d-1}_{-m}$. Since $\|R(\mu)\| < p_m$ (as explained below Definition 4.4.3), there exist $\alpha \in [0, p_m)$ and a unit vector u in $T_m\mathcal{S}^{d-1}$ such that

$$R(\mu) = p_m \mathrm{E}\left[\frac{\pi_m(\mu) - \pi_m(X)}{\|\pi_m(\mu) - \pi_m(X)\|}\xi_{X,\mu}\right] = \alpha u, \tag{4.8.30}$$

where X is an $\mathcal{S}^{d-1}_{-m}$-valued random vector with distribution P_{-m}. Theorem 4.4.2 thus yields that $\mu = Q(\alpha u)$, which shows that the image of Q contains $\mathcal{S}^{d-1}_{-m}$. This establishes the result since, by definition, $-m = Q(u)$ for any unit vector u in $T_m\mathcal{S}^{d-1}$.
(ii) Since P is non-atomic, $p_m = 1$. Fix then $\mu \in \mathcal{S}^{d-1}_{-m}$, $\alpha \in [0, 1)$, and a unit vector $u \in T_m\mathcal{S}^{d-1}$. Since P is non-atomic, Theorem 4.4.2 yields that $\mu = Q(\alpha u)$ *if and only if* (4.8.30) holds. This establishes that the restriction of Q to $\mathcal{B}_{m,1}$ and the restriction of R to $\mathcal{S}^{d-1}_{-m}$ are inverse maps of one another. Now, since $\|R(\mu)\| < 1$ for any $\mu \in \mathcal{S}^{d-1}_{-m}$, we have that $R(\mu) = u^{\infty}_{m,1}$ if and only if $\mu = -m$. Similarly, $Q(z) = -m$ if and only if $z = u^{\infty}_{m,1}$. This establishes that $Q = R^{-1}$. Since Theorems 4.3.5 and 4.4.4 further ensure continuity of Q and R, respectively, the result is proved. ■

Proof of Theorem 4.4.6. (i) Let us first show that $p_{1,m} = p_{2,m}$, with $p_{\ell,m} := 1 - P_\ell[\{-m\}]$. Let v be an arbitrary unit vector in $T_m\mathcal{S}^{d-1}$. For any $\lambda \in \mathbb{R}$, note that $\mu_\lambda := \pi_m^{-1}(\lambda v) \neq -m$, so that $R^m_{P_\ell}(\mu_\lambda) \in \mathcal{B}_{m,p_{\ell,m}}$ (recall that only $-m$ gets the rank $u^{\infty}_{m,p_{\ell,m}}$). For $\ell = 1, 2$, consider then the functions

$$g_\ell : \mathbb{R} \to \mathcal{B}_{m,p_{\ell,m}}$$
$$\lambda \mapsto \|R^m_{P_\ell}(\mu_\lambda)\| = p_{\ell,m}\left\|\mathrm{E}\left[\frac{\lambda v - \pi_m(X_\ell)}{\|\lambda v - \pi_m(X_\ell)\|}\xi_{X_\ell,\mu_\lambda}\right]\right\|,$$

where X_ℓ is an $\mathcal{S}^{d-1}_{-m}$-valued random vector with distribution $(P_\ell)_{-m}$. A routine application of Lebesgue's dominated convergence theorem yields that $g_\ell(\lambda) \to p_{\ell,m}$ as $\lambda \to \infty$, for $\ell = 1, 2$. Since $R^m_{P_1} = R^m_{P_2}$ on $\mathcal{S}^{d-1}$, we have that $g_1 = g_2$ on $\mathbb{R}$, which then yields $p_{1,m} = p_{2,m}$.

Let $v_1, \ldots, v_{d-1}$ form an orthonormal basis of $T_m\mathcal{S}^{d-1}$ and denote as e_j the jth vector of the canonical basis of $\mathbb{R}^{d-1}$. Let Ψ be the linear mapping defined by

$$\Psi : T_m\mathcal{S}^{d-1} \to \mathbb{R}^{d-1} : z = \sum_{j=1}^{d-1}\lambda_j v_j \mapsto \sum_{j=1}^{d-1}\lambda_j e_j.$$

Note that Ψ is an isometry since, for any $z := \sum_{j=1}^{d-1}\lambda_j v_j \in T_m\mathcal{S}^{d-1}$, we have

$$\|\Psi(z)\|^2 = \sum_{j=1}^{d-1}|\lambda_j|^2 = \|z\|^2.$$

Since $\pi_m(\mu)$ and $\pi_m(X_\ell)$, $\ell = 1, 2$, take their values in $T_m\mathcal{S}^{d-1}$, we thus have

$$\Psi(R^m_{P_\ell}(\mu)) = p_m\mathrm{E}\left[\frac{\Psi(\pi_m(\mu)) - \Psi(\pi_m(X_\ell))}{\|\Psi(\pi_m(\mu)) - \Psi(\pi_m(X_\ell))\|}\xi_{X_\ell,\mu}\right], \quad \ell = 1, 2,$$

for any $\mu \in \mathcal{S}^{d-1}_{-m}$, where p_m stands for the common value of $p_{1,m}$ and $p_{2,m}$. Noting that $X_\ell = \mu$ if and only if $\Psi(\pi_m(X_\ell)) = \Psi(\pi_m(\mu))$, this yields, still for any $\mu \in \mathcal{S}^{d-1}_{-m}$,

$$\Psi(R^m_{P_\ell}(\mu)) = p_m\mathrm{E}\left[\frac{\Psi(\pi_m(\mu)) - Y_\ell}{\|\Psi(\pi_m(\mu)) - Y_\ell\|}\xi_{Y_\ell,\Psi(\pi_m(\mu))}\right], \quad \ell = 1, 2,$$

where we let $Y_\ell := \Psi(\pi_m(X_\ell))$, $\ell = 1, 2$. Since $R^m_{P_1} = R^m_{P_2}$ on $\mathcal{S}^{d-1}_{-m}$ and since Ψ is a surjective mapping, we deduce that

$$\mathrm{E}\left[\frac{y - Y_1}{\|y - Y_1\|}\xi_{Y_1,y}\right] = \mathrm{E}\left[\frac{y - Y_2}{\|y - Y_2\|}\xi_{Y_2,y}\right]$$

for any $y \in \mathbb{R}^{d-1}$. Proposition 2.5 from [50] then entails that Y_1 and Y_2 are equal in distribution. Since $\Psi \circ \pi_m$ is a one-to-one mapping, this implies that X_1 and X_2 are equal in distribution, that is, that $(P_1)_{-m} = (P_2)_{-m}$. Jointly with the fact that $P_1[\{-m\}] = 1 - p_{1,m} = 1 - p_{2,m} = P_2[\{-m\}]$, this establishes that $P_1 = P_2$.
(ii) Fix $\ell \in \{1, 2\}$. Since P_ℓ is non-atomic and is not concentrated on a great circle containing m, Theorem 4.4.5 yields that $Q^m_{P_\ell}$ is a one-to-one mapping with inverse $R^m_{P_\ell}$. It follows that $R^m_{P_1} = R^m_{P_2}$, so that Part (i) of the result entails that $P_1 = P_2$. ■

Proofs for Section 4.5

Since the result is needed in the proof of Theorem 4.5.2, we first prove Theorem 4.5.3.

Proof of Theorem 4.5.3. Assume that μ is an m-quantile of order α in direction u for P. We consider two cases. (a) $\alpha \in [p_m, 1]$. Then, $\mu = -m$ (see Definition 4.2.2). Since

$$p_m(P_O) := 1 - P_O[\{-m\}] = 1 - P[\{O(-m)\}] = 1 - P[\{-m\}] = p_m(P),$$

we have $\alpha \in [p_m(P_O), 1]$, so that Definition 4.2.2 implies that $\mu = -m$ is an m-geometric quantile of order α in any direction for P_O. In particular, $O\mu = -m$ is an m-geometric quantile of order α in direction Ou for P_O. (b) $\alpha \in [0, p_m)$. Then, $\mu \in \mathcal{S}^{d-1}_{-m}$ and

$$M^{m,P}_{\alpha,u}(\mu) \le M^{m,P}_{\alpha,u}(x) \quad \text{for any } x \in \mathcal{S}^{d-1}_{-m}. \tag{4.8.31}$$

Since $Om = O'm = m$, we have $\pi_m(Ox) = O\pi_m(x)$ for any $x \in \mathcal{S}^{d-1}_{-m}$, which, after straightforward computations, yields that

$$M^{m,P}_{\alpha,u}(x) = M^{m,P_O}_{\alpha,Ou}(Ox) \quad \text{for any } x \in \mathcal{S}^{d-1}_{-m}.$$

This allows us to rewrite (4.8.31) as

$$M^{m,P_O}_{\alpha,Ou}(O\mu) \le M^{m,P_O}_{\alpha,Ou}(Ox) \quad \text{for any } x \in \mathcal{S}^{d-1}_{-m}.$$

Since $x \mapsto Ox$ defines a one-to-one transformation of $\mathcal{S}^{d-1}_{-m}$, we have that

$$M^{m,P_O}_{\alpha,Ou}(O\mu) \le M^{m,P_O}_{\alpha,Ou}(x) \quad \text{for any } x \in \mathcal{S}^{d-1}_{-m},$$

which states that $O\mu$ is an m-quantile of order α in direction Ou for P_O.

Let us now assume that $O\mu$ is an m-quantile of order α in direction Ou for P_O. By considering O' instead of O in the first part of the proof, we obtain that $\mu = O'(O\mu)$ is an m-quantile of order α in direction $u = O'(Ou)$ for $P = P_{O'O}$. ■

Proof of Theorem 4.5.2. (i) Fix an arbitrary unit vector u in $T_m\mathcal{S}^{d-1}$. By Theorem 4.3.2, we have that m is an m-quantile of order $\alpha = 0$ in direction u for P. It remains to show that it is unique. If P is not concentrated on a great circle containing m, then uniqueness follows from Theorem 4.3.1(ii). Assume thus that P is concentrated on a great circle containing m, hence, from rotational symmetry, concentrated on $\{-m, m\}$. By Definition 4.2.2, any m-quantile of order $\alpha = 0$ in direction u for P must belong to $\mathcal{S}^{d-1}_{-m}$. Assume then that $\mu \in \mathcal{S}^{d-1}_{-m} \setminus \{m\}$ is an m-quantile of order $\alpha = 0$ in direction u for P. From Theorem 4.4.2, we must then have $P[\{m\}] \le P[\{\mu\}] = 0$. It follows that $P[\{-m\}] = 1$, which contradicts the fact that m is a Fréchet median of P. Thus, m is the unique m-quantile of order $\alpha = 0$ in direction u for P.

(ii) We consider two cases. (a) $\alpha \in [p_m, 1]$. Then, the only quantile of order α in direction u for P is $-m$ (Definition 4.2.2). In particular, it is unique and it belongs to the meridian $\{(\cos t)m + (\sin t)u : t \in [0, \pi]\}$. (b) $\alpha \in [0, p_m)$. The existence of a quantile of order α in direction u for P is guaranteed by Theorem 4.3.1(i). Let us prove uniqueness. If P is not concentrated on a great circle containing m, uniqueness follows from Theorem 4.3.1(ii). Assume therefore that P is concentrated on a great circle containing m, hence, from rotational symmetry, on $\{-m, m\}$. Theorem 4.4.2 then entails that $\mu \in \mathcal{S}^{d-1}_{-m}$ is an m-quantile of order α in direction u for P if and only if

$$\left\| \frac{\pi_m(\mu)}{\|\pi_m(\mu)\|} \xi_{\mu,m} P[\{m\}] - \alpha u \right\| \leq P[\{\mu\}]. \tag{4.8.32}$$

Assume that this condition is fulfilled at some $\mu \in \mathcal{S}^{d-1}_{-m} \setminus \{m\}$. Since P is concentrated on $\{-m, m\}$, we then have

$$\left\| \frac{\pi_m(\mu)}{\|\pi_m(\mu)\|} P[\{m\}] - \alpha u \right\| = 0.$$

In particular, we have $\alpha = P[\{m\}] = p_m$, a contradiction since $\alpha \in [0, p_m)$. On the other hand, Condition (4.8.32) is fulfilled at $\mu = m$ since it then rewrites $\alpha = \|\alpha u\| \leq P[\{m\}] = p_m$. This implies that, when P is not concentrated on a great circle containing m, the m-quantile of order α in direction u exists, is unique and equal to m, irrespective of $\alpha \in [0, p_m)$ and of the unit vector u in $T_m\mathcal{S}^{d-1}$. Let us show that $\mu^m_{\alpha,u} = \mu^m_{\alpha,u}(P)$ belongs to the meridian $\{(\cos t)m + (\sin t)u : t \in [0, \pi]\}$. Fix an arbitrary unit vector v that is orthogonal to both m and u. Then, $O = I_d - 2vv'$ is a $d \times d$ matrix O such that $Om = m$ and $Ou = u$, so that Theorem 4.5.3 provides

$$\mu^m_{\alpha,u}(P) = O\mu^m_{\alpha,u}(P_O) = O\mu^m_{\alpha,u}(P) = \mu^m_{\alpha,u}(P) - 2(v'\mu^m_{\alpha,u}(P))v,$$

hence $v'\mu^m_{\alpha,u}(P) = 0$. Since this holds for any unit vector v orthogonal to m and u, it follows that $\mu^m_{\alpha,u}$ belongs to the vector space spanned by m and u, hence to one of the meridians $M_u := \{(\cos t)m + (\sin t)u : t \in [0, \pi]\}$ and $M_{-u} := \{(\cos t)m + (\sin t)(-u) : t \in [0, \pi]\}$. It remains to show that $\mu^m_{\alpha,u}$ does not belong to $M_{-u} \setminus M_u$. Ad absurdum, assume that it does, hence is of the form $\mu^m_{\alpha,u} = (\cos t_*)m - (\sin t_*)u$ for some $t_* \in (0, \pi)$. Then, for $\tilde{\mu} := (\cos t_*)m + (\sin t_*)u$, rotational symmetry of P_* yields

$$\begin{aligned}
M^{m,P}_{\alpha,u}(\tilde{\mu}) &= \left(\int_{\mathcal{S}^{d-1}_{-m}} \{ \|\pi_m(x) - \pi_m(\tilde{\mu})\| - \|\pi_m(x)\| \} \, dP_{-m}(x) \right) - \alpha u' \pi_m(\tilde{\mu})/p_m \\
&= \left(\int_{\mathcal{S}^{d-1}_{-m}} \{ \|\pi_m(x) - \pi_m(\mu^m_{\alpha,u})\| - \|\pi_m(x)\| \} \, dP_{-m}(x) \right) - \alpha u' \pi_m(\tilde{\mu})/p_m \\
&= M^{m,P}_{\alpha,u}(\mu^m_{\alpha,u}) + \alpha(u'\pi_m(\mu^m_{\alpha,u}) - u'\pi_m(\tilde{\mu}))/p_m \\
&< M^{m,P}_{\alpha,u}(\mu^m_{\alpha,u}),
\end{aligned}$$

which contradicts the fact that $\mu^m_{\alpha,u}$ is the m-geometric quantile of order α in direction u for P. We conclude that $\mu^m_{\alpha,u} \in M_u = \{(\cos t)m + (\sin t)u : t \in [0, \pi]\}$.

(iii) Fix an arbitrary unit vector u in $T_m\mathcal{S}^{d-1}$. If P is concentrated on a great circle containing m, then we showed in Part (ii) of the proof that $\mu^m_{\alpha,u} = m$ for any $\alpha \in [0, p_m)$. Since $\mu^m_{\alpha,u} = -m$ for any $\alpha \in [p_m, 1]$, we have that $d(\mu_{\alpha,u}, m) = 1$ for any $\alpha \in [0, p_m)$ and $d(\mu^m_{\alpha,u}, m) = \pi$ for any $\alpha \in [p_m, 1]$. In particular, the map $\alpha \mapsto d(\mu_{\alpha,u}, m)$ is monotone non-decreasing over $[0,1]$. We may thus assume that P is not concentrated on a great circle containing m. Since rotational symmetry of P about m yields

$$\mathrm{E}\left[\frac{\pi_m(X)}{\|\pi_m(X)\|}\xi_{X,m}\right]=0,$$

Theorem 4.4.6 implies that m is an m-quantile of order α in direction u for P if and only if $\alpha \leq P[\{m\}]$. It follows that $\mu^m_{\alpha,u}=m$ for any $\alpha\in[0,P[\{m\}]]$, that $\mu^m_{\alpha,u}\in\{(\cos t)m+(\sin t)u : t\in(0,\pi)\}$ for any $\alpha\in(P[\{m\}],p_m)$, and $\mu^m_{\alpha,u}=-m$ for any $\alpha\in[p_m,1]$ (note that we indeed have $P[\{m\}]\leq p_m=P[\mathcal{S}^{d-1}_{-m}]$). It remains to show the monotonicity of $\alpha\mapsto d(\mu^m_{\alpha,u},m)$ over $(P[\{m\}],p_m)$. For any $\alpha\in(P[\{m\}],p_m)$, Theorem 4.4.6 implies that $\mu=\mu^m_{\alpha,u}$ if and only if $R^m_P(\mu)=\alpha u$, since $P[\{\mu^m_{\alpha,u}\}]=0$. In particular, if $\mu^m_{\alpha,u}=\mu^m_{\beta,u}$ for some $\alpha,\beta\in(P[\{m\}],p_m)$, then $\alpha=\beta$. This entails that $\alpha\mapsto\mu^m_{\alpha,u}$ is injective over $(P[\{m\}],p_m)$. For any $\alpha\in(P[\{m\}],p_m)$, Part (ii) of the result allows us to write $\mu^m_{\alpha,u}=(\cos t_\alpha)m+(\sin t_\alpha)u$, for some $t_\alpha\in(0,\pi)$. The map $\alpha\mapsto t_\alpha$ is injective over $(P[\{m\}],p_m)$. Indeed, if $t_\alpha=t_\beta$ for some $\alpha,\beta\in(P[\{m\}],p_m)$, then $\mu^m_{\alpha,u}=\mu^m_{\beta,u}$, which entails that $\alpha=\beta$. Since P is not concentrated on a great circle containing m, Theorem 4.3.5 ensures that $\alpha\mapsto\mu^m_{\alpha,u}$ is continuous over $[0,1]$, hence that

$$\alpha\mapsto t_\alpha=d(\mu^m_{\alpha,u},m)=\arccos(m'\mu^m_{\alpha,u})$$

is continuous over $[0,1]$. Since $\alpha\mapsto d(\mu^m_{\alpha,u},m)$ is continuous over $[0,1]$ and injective over $(P[\{m\}],p_m)$, it is monotone strictly increasing over $(P[\{m\}],p_m)$, hence monotone non-decreasing over $[0,1]$ (it is constant over $[0,P[\{m\}]]$ and over $[p_m,1]$).

(iv) Since P is rotationally symmetric about m, the fact that P is not concentrated on $\{-m,m\}$ implies that P is not concentrated on a great circle containing m. In this case, recall that $C^m_P(\alpha)$ is defined, for any $\alpha\in[1-p_m,1]$, as the collection of all locations $\mu\in\mathcal{S}^{d-1}$ such that $D^m_P(\mu)=\alpha$, i.e. such that $\|R^m_P(\mu)\|_{m,p_m}=1-\alpha$. We then have $C^m_P(\alpha)=\{-m\}$ when $\alpha=1-p_m$, and $C^m_P(\alpha)=\{m\}$ when $\alpha\in[1-P[\{m\},1]$; for such values of α, the claim on $C^m_P(\alpha)$ trivially holds with $c_\alpha=-1$ and $c_\alpha=1$, respectively. We may therefore restrict to case $\alpha\in(1-p_m,1-P[\{m\}])$. For any $\mu\in\mathcal{S}^{d-1}_{-m}$, any $\beta\in(P[\{m\}],p_m)$ and any unit vector $v\in T_m\mathcal{S}^{d-1}$, we showed in Part (iii) of the proof that $R^m_P(\mu)=\beta v$ if and only if $\mu=\mu^m_{\beta,v}$, the unique quantile of order β in direction v for P (see Theorem 4.3.1). Since $R^m_P(\mu)\in T_m\mathcal{S}^{d-1}$, it follows that, for any $\alpha\in(1-p_m,1-P[\{m\}])$, $\|R^m_P(\mu)\|_{m,p_m}=1-\alpha$ if and only if $\mu=\mu^m_{1-\alpha,u}$ for some unit vector $u\in T_m\mathcal{S}^{d-1}$. In particular, we have that

$$C^m_P(\alpha)=\{\mu^m_{1-\alpha,u} : u\in T_m\mathcal{S}^{d-1},\|u\|=1\}$$

for any $\alpha\in(1-p_m,1-P[\{m\}])$. For any $\beta\in(P[\{m\}],p_m)$ and any unit vector $u\in T_m\mathcal{S}^{d-1}$, let us write

$$\mu^m_{\beta,u}=(\cos t_{\beta,u})m+(\sin t_{\beta,u})u$$

for some $t_{\beta,u}\in[0,\pi]$. In Parts (ii)–(iii), we considered $t_{\beta,u}$ as depending only on β since u was fixed. We now consider dependence in both β and u. Let us show that, for any $\beta\in[0,1]$, the map $u\mapsto t_{\beta,u}$ is constant over the unit sphere of $T_m\mathcal{S}^{d-1}$. Fix $\beta\in(P[\{m\}],p_m)$ and let u,v be unit vectors in $T_m\mathcal{S}^{d-1}$. Let O be a $d\times d$ orthogonal matrix such that $Om = m$ and $Ou = v$. Together with the uniqueness from Part (iii) of the result, Theorem 4.5.3 entails that $O\mu^m_{\beta,u} = \mu^m_{\beta,v}$, which rewrites

$$(\cos t_{\beta,u})m + (\sin t_{\beta,u})v = (\cos t_{\beta,v})m + (\sin t_{\beta,v})v.$$

This yields $t_{\beta,u} = t_{\beta,v}$. We therefore proved that, for any $\beta \in (P[\{m\}], p_m)$, the map $u \mapsto t_{\beta,u}$ is constant over the unit sphere of $T_m\mathcal{S}^{d-1}$, which allows us to write $t_{\beta,u} = t_\beta$ for any $\beta \in (P[\{m\}], p_m)$ and any unit vector $u \in T_m\mathcal{S}^{d-1}$. It follows that

$$C^m_P(\alpha) = \{(\cos t_{1-\alpha})m + (\sin t_{1-\alpha})u : u \in T_m\mathcal{S}^{d-1}, \|u\| = 1\}$$

for any $\alpha \in (1 - p_m, 1 - P[\{m\}])$. We therefore have that $C^m_P(\alpha) \subset \{\mu \in \mathcal{S}^{d-1} : \mu' m = \cos t_{1-\alpha}\}$. Let then $\mu \in \mathcal{S}^{d-1}$ be such that $\mu' m = \cos t_{1-\alpha}$. Then $\mu = (\cos t_{1-\alpha})m + z$ for some $z \in T_m\mathcal{S}^{d-1}$. Since $\|\mu\| = 1$, we have that $(\cos t_{1-\alpha})^2 + \|z\|^2 = 1$. In particular, we have $\|z\| = \sin t_{1-\alpha}$, and $z = (\sin t_{1-\alpha})u$ for some unit vector $u \in T_m\mathcal{S}^{d-1}$. This implies that $\mu \in C_P(\alpha)$. We conclude that $C^m_P(\alpha) = \{\mu \in \mathcal{S}^{d-1} : \mu' m = \cos t_{1-\alpha}\}$ for any $\alpha \in (1 - p_m, 1 - P[\{m\}])$. Since $\alpha \mapsto t_\alpha$ is continuous and monotone non-decreasing over $(1 - p_m, 1 - P[\{m\}])$, the map $\alpha \mapsto c_\alpha := \cos t_{1-\alpha}$ is continuous and monotone non-decreasing over $(1 - p_m, 1 - P[\{m\}])$. Since $c_\alpha = -1$ for $\alpha = 1 - p_m$ and $c_\alpha = 1$ for $\alpha \in [1 - P[\{m\}], 1]$, we conclude that $\alpha \mapsto c_\alpha$ is monotone non-decreasing over $[1 - p_m, 1]$. ■

Proofs for Section 4.6

Proof of Theorem 4.6.1. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be the probability space on which the random vectors $X_1, X_2, \ldots$ are defined. Let us first show that there exists $V \in \mathcal{A}$, with $\mathbb{P}[V] = 1$, such that for any $\mu \in \mathcal{S}^{d-1}$ and any $\omega \in V$,

$$\frac{1}{n}\sum_{i=1}^n d(\mu, X_i(\omega)) \to \mathbb{E}[d(\mu, X_1)] \tag{4.8.33}$$

as $n \to \infty$. To this end, let D be a countable and dense subset of $\mathcal{S}^{d-1}$. The strong law of large numbers entails that for any $\mu \in D$, there exists $V_\mu \in \mathcal{A}$ such that $\mathbb{P}[V_\mu] = 1$ and (4.8.33) holds for any $\omega \in V_\mu$. Of course, $V := \cap_{\mu\in D} V_\mu$ satisfies $\mathbb{P}[V] = 1$ and is such that (4.8.33) holds for any $\mu \in D$ and any $\omega \in V$. To extend the result from D to $\mathcal{S}^{d-1}$, fix $\mu \in \mathcal{S}^{d-1}$ and $\omega \in V$ arbitrarily. Pick then a sequence (μ_k) in D such that $d(\mu_k, \mu) \to 0$ as $k \to \infty$. The triangle inequality yields

$$\frac{1}{n}\sum_{i=1}^n d(\mu_k, X_i(\omega)) - d(\mu, \mu_k) \le \frac{1}{n}\sum_{i=1}^n d(\mu, X_i(\omega)) \le \frac{1}{n}\sum_{i=1}^n d(\mu_k, X_i(\omega)) + d(\mu, \mu_k)$$

for any $k \in \mathbb{N}$ and any positive integer n. Since (4.8.33) holds for any $\mu \in D$ and any $\omega \in V$, taking the lim inf and lim sup as $n \to \infty$ then yields

$$\begin{aligned}&\mathrm{E}[d(\mu_k, X_1)] - d(\mu, \mu_k)\\ &\quad\le \liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^n d(\mu, X_i(\omega)) \le \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n d(\mu, X_i(\omega)) \le \mathbb{E}[d(\mu_k, X_1)] + d(\mu, \mu_k)\end{aligned}$$

for any $k \in \mathbb{N}$. Lebesgue's dominated convergence theorem and continuity of the distance function then entail that

$$\frac{1}{n}\sum_{i=1}^{n} d(\mu, X_i(\omega)) \to \mathbb{E}[d(\mu, X_1)]$$

as $n \to \infty$, which proves that (4.8.33) holds for any $\mu \in \mathcal{S}^{d-1}$ and any $\omega \in V$.

Now, fix $\omega \in V$ and let $(\hat{m}_{n_k}(\omega))$ be a subsequence of $(\hat{m}_n(\omega))$. From compactness of $\mathcal{S}^{d-1}$, there exists a subsequence $(\hat{m}_{n_{k_\ell}}(\omega))$ that converges in $\mathcal{S}^{d-1}$ as $\ell \to \infty$, to μ_* say. By definition,

$$\frac{1}{n_{k_\ell}}\sum_{i=1}^{n_{k_\ell}} d(\hat{m}_{n_{k_\ell}}(\omega), X_i(\omega)) \leq \frac{1}{n_{k_\ell}}\sum_{i=1}^{n_{k_\ell}} d(\mu, X_i(\omega))$$

for any $\mu \in \mathcal{S}^{d-1}$, which, by taking the limit as $\ell \to \infty$, entails that

$$\mathbb{E}[d(\mu_*, X_1)] \leq \mathbb{E}[d(\mu, X_1)], \tag{4.8.34}$$

for any $\mu \in \mathcal{S}^{d-1}$; convergence in the lefthand side follows from the fact that the triangle inequality yields

$$\frac{1}{n_{k_\ell}}\sum_{i=1}^{n_{k_\ell}} d(\mu_*, X_i(\omega)) - d(\hat{m}_{n_{k_\ell}}(\omega), \mu_*)$$

$$\leq \frac{1}{n_{k_\ell}}\sum_{i=1}^{n_{k_\ell}} d(\hat{m}_{n_{k_\ell}}(\omega), X_i(\omega)) \leq \frac{1}{n_{k_\ell}}\sum_{i=1}^{n_{k_\ell}} d(\mu_*, X_i(\omega)) + d(\hat{m}_{n_{k_\ell}}(\omega), \mu_*).$$

Since P has a unique Fréchet median m by assumption, it follows from (4.8.34) that $\mu_* = m$. Therefore, for any $\omega \in V$, any subsequence of $(\hat{m}_n(\omega))$ admits a further subsequence converging to m, which implies that $(\hat{m}_n(\omega))$ itself converges to m. This establishes the result. ■

The proof of Theorem 4.6.2 requires Lemmas 4.8.9–4.8.11 below. The proof of the first lemma is left to the reader.

Lemma 4.8.9. *The inverse map $\pi_m^{-1} : T_m\mathcal{S}^{d-1} \to \mathcal{S}^{d-1}_{-m}$ of π_m satisfies*

$$\pi_m^{-1}(\theta) = \frac{2}{1+\|\theta\|^2}(\theta + m) - m = \frac{2}{1+\|\theta\|^2}\theta + \frac{1-\|\theta\|^2}{1+\|\theta\|^2}m, \tag{4.8.35}$$

for any $\theta \in T_m\mathcal{S}^{d-1}$. Furthermore, when extending π_m^{-1} to $\mathbb{R}^d$ via (4.8.35), the Jacobian matrix $J_\theta(\pi_m^{-1})$ of π_m^{-1} at θ is given by

$$J_\theta(\pi_m^{-1}) = \frac{2}{1+\|\theta\|^2}\left(I_d - \frac{2(\theta+m)\theta'}{1+\|\theta\|^2}\right) \tag{4.8.36}$$

for any $\theta \in \mathbb{R}^d$. For any $\theta \in T_m\mathcal{S}^{d-1}$,

$$J_\theta(\pi_m^{-1})(I_d - mm') = Q_\theta J_\theta(\pi_m^{-1})(I_d - mm'), \tag{4.8.37}$$

where we let $Q_\theta := I_d - \pi_m^{-1}(\theta)(\pi_m^{-1}(\theta))'$, so that, when restricted to $T_m\mathcal{S}^{d-1}$, the linear map whose matrix is $J_\theta(\pi_m^{-1})$ takes its values in $T_{\pi_m^{-1}(\theta)}\mathcal{S}^{d-1}$.

Lemma 4.8.10. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and assume that P admits a bounded density f with respect to the surface area measure on $\mathcal{S}^{d-1}$. Let X be an $\mathcal{S}^{d-1}$-valued random variable with distribution P. Then, for any $y \in \mathcal{S}^{d-1}$, the distribution of $y'X$ is absolutely continuous with respect to the Lebesgue measure on $[-1,1]$. Furthermore, denoting this density as $f^{y'X}$, there exists a positive constant C_d such that*

$$f^{y'X}(s) \leq C_d \|f\|_\infty (1-s^2)^{(d-3)/2}$$

for any $y \in \mathcal{S}^{d-1}$ and any $s \in [-1,1]$, with $\|f\|_\infty := \sup_{x\in\mathcal{S}^{d-1}} |f(x)|$.

Proof of Lemma 4.8.10. The map $\psi : (0,\pi)^{d-2} \times (0,2\pi) \to \mathcal{S}^{d-1}$ defined by

$$\begin{aligned}\psi(\varphi_1,\ldots,\varphi_{d-1}) := (&\cos\varphi_1, (\sin\varphi_1)(\cos\varphi_2),\ldots,\\ &(\sin\varphi_1)\ldots(\sin\varphi_{d-2})(\cos\varphi_{d-1}), (\sin\varphi_1)\ldots(\sin\varphi_{d-2})(\sin\varphi_{d-1}))\end{aligned} \tag{4.8.38}$$

provides a parametrization of $\mathcal{S}^{d-1} \setminus N$ for some subset N of $\mathcal{S}^{d-1}$ whose surface area measure $\sigma_{d-1}(N)$ is zero. Writing $\varphi := (\varphi_1,\ldots,\varphi_{d-1})$, the corresponding surface element is

$$g(\varphi) := \sqrt{\det\big((J_\varphi(\psi))' J_\varphi(\psi)\big)} = (\sin\varphi_1)^{d-2}(\sin\varphi_2)^{d-3}\ldots(\sin\varphi_{d-2}),$$

where $J_\varphi(\psi)$ is the $(d \times (d-1))$ Jacobian matrix of ψ at φ. Fix then $y \in \mathcal{S}^{d-1}$ and let $O = O_y$ be a $d \times d$ orthogonal matrix such that $Oy = (1,0,\ldots,0)'$. Fix further $a, b \in [-1,1]$ with $a < b$. Denoting as f^{OX} the density of OX, we have

$$\begin{aligned}P[a < y'X < b] &= P[a < (OX)_1 < b]\\ &= \int_{\{x: a < x_1 < b\}\cap\mathcal{S}^{d-1}} f^{OX}(x)\, d\sigma_{d-1}(x)\\ &= \int_{\arccos(b)}^{\arccos(a)} \int_{(0,\pi)^{d-3}\times(0,2\pi)} f^{OX}(\psi(\varphi)) g(\varphi)\, d\varphi.\end{aligned}$$

Letting $\varphi_1 := \arccos(x_1)$ then yields

$$\begin{aligned}&P[a < y'X < b]\\ &\quad= \int_a^b \frac{1}{\sqrt{1-x_1^2}} \int_{(0,\pi)^{d-3}\times(0,2\pi)} f^{OX}\big(\psi(\arccos(x_1),\varphi_2,\ldots,\varphi_{d-1})\big)\\ &\qquad\qquad \times g(\arccos(x_1),\varphi_2,\ldots,\varphi_{d-1})\, dx_1 d\varphi_2\ldots d\varphi_{d-1}\\ &\quad= \int_a^b \frac{1}{\sqrt{1-x_1^2}} \int_{(0,\pi)^{d-3}\times(0,2\pi)} f^{OX}\big(\psi(\arccos(x_1),\varphi_2,\ldots,\varphi_{d-1})\big)\\ &\qquad\qquad \times \Big(\sqrt{1-x_1^2}\Big)^{d-2} (\sin\varphi_2)^{d-3}\ldots(\sin\varphi_{d-2})\, dx_1 d\varphi_2\ldots d\varphi_{d-1}\\ &\quad= \int_a^b \Big(\sqrt{1-x_1^2}\Big)^{d-3} \int_{(0,\pi)^{d-3}\times(0,2\pi)} f^{OX}\big(\psi(\arccos(x_1),\varphi_2,\ldots,\varphi_{d-1})\big)\\ &\qquad\qquad \times (\sin\varphi_2)^{d-3}\ldots(\sin\varphi_{d-2})\, dx_1 d\varphi_2\ldots d\varphi_{d-1}.\end{aligned}$$

Since $f^{OX}(x) = f(O'x)$ for any $x \in \mathcal{S}^{d-1}$, this shows that $y'X$ is absolutely continuous with respect to the Lebesgue measure on $[-1,1]$ and admits the density

$$f^{y'X}(s) = (1-s^2)^{(d-3)/2} \int_{(0,\pi)^{d-3}\times(0,2\pi)} f\big(O'\psi(\arccos(s), \varphi_2, \ldots, \varphi_{d-1})\big) \times (\sin\varphi_2)^{d-3} \ldots (\sin\varphi_{d-2})\, d\varphi_2 \ldots d\varphi_{d-1}.$$

Clearly, we then have

$$|f^{y'X}(s)| \leq 2\pi^{d-2}\|f\|_\infty (1-s^2)^{(d-3)/2}$$

for any $y \in \mathcal{S}^{d-1}$ and any $s \in [-1,1]$. ■

Lemma 4.8.11. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and assume that P admits a bounded density f with respect to the surface area measure on $\mathcal{S}^{d-1}$. Let X be an $\mathcal{S}^{d-1}$-valued random vector with distribution P and let (Y_n) be a sequence of $\mathcal{S}^{d-1}$-valued random vectors that are each independent of X. Then, for any $\delta \in [0, d-2)$,*

$$\mathrm{E}\bigg[\frac{\xi_{Y_n,\pm X}}{(1-(Y_n'X)^2)^{(1+\delta)/2}}\bigg] = O(1)$$

as n diverges to infinity, where we recall that $\xi_{x,\pm y} := \mathbb{I}[x \notin \{-y, y\}]$ (this applies in particular when $(Y_n) = (y_n)$ is an arbitrary deterministic sequence in $\mathcal{S}^{d-1}$).

Proof of Lemma 4.8.11. Fix $\varepsilon \in (0,1)$. For any $x, y \in \mathcal{S}^{d-1}$ such that $y'x \neq \pm 1$, writing $1 = \mathbb{I}[y'x \leq -1+\varepsilon] + \mathbb{I}[-1+\varepsilon < y'x < 1-\varepsilon] + \mathbb{I}[y'x \geq 1-\varepsilon]$ provides

$$\begin{aligned}\frac{\xi_{y,\pm x}}{(1-(y'x)^2)^{(1+\delta)/2}} \leq{}& \frac{\xi_{-y,x}}{(2-\varepsilon)^{(1+\delta)/2}(1+y'x)^{(1+\delta)/2}}\mathbb{I}[y'x \leq -1+\varepsilon] \\ &+\frac{1}{(2-\varepsilon)^{1+\delta}}\mathbb{I}[-1+\varepsilon < y'x < 1-\varepsilon] \\ &+\frac{\xi_{y,x}}{(2-\varepsilon)^{(1+\delta)/2}(1-y'x)^{(1+\delta)/2}}\mathbb{I}[y'x \geq 1-\varepsilon].\end{aligned}$$

It is thus sufficient to examine

$$\mathrm{E}\bigg[\frac{\xi_{-Y_n,X}}{(1+Y_n'X)^{(1+\delta)/2}}\mathbb{I}[Y_n'X \leq -1+\varepsilon]\bigg] \tag{4.8.39}$$

and

$$\mathrm{E}\bigg[\frac{\xi_{Y_n,X}}{(1-Y_n'X)^{(1+\delta)/2}}\mathbb{I}[Y_n'X \geq 1-\varepsilon]\bigg]. \tag{4.8.40}$$

Conditioning with respect to Y_n, Lemma 4.8.10 yields (below, $f^{y'X}$ still is the density of $y'X$)

$$\mathrm{E}\Big[\frac{\xi_{-Y_n,X}}{(1+Y_n'X)^{(1+\delta)/2}}\mathbb{I}[Y_n'X \leq -1+\varepsilon]\Big]$$

$$= \int_{\mathcal{S}^{d-1}} \mathrm{E}\Big[\frac{\xi_{-y,x}}{(1+y'X)^{(1+\delta)/2}}\mathbb{I}[y'X \leq -1+\varepsilon]\Big]\, dP^{Y_n}(y)$$

$$= \int_{\mathcal{S}^{d-1}} \int_{-1}^{-1+\varepsilon} \frac{f^{y'X}(s)}{(1+s)^{(1+\delta)/2}}\, ds\, dP^{Y_n}(y)$$

$$\leq C_d \|f\|_\infty \int_{-1}^{-1+\varepsilon} \frac{((1-s)(1+s))^{(d-3)/2}}{(1+s)^{(1+\delta)/2}}\, ds$$

$$\leq 2^{(d-3)/2} C_d \|f\|_\infty \int_{-1}^{-1+\varepsilon} (1+s)^{(d-4-\delta)/2}\, ds,$$

which is finite since $d-4-\delta > -2$. This shows that (4.8.39) is $O(1)$. Since an entirely similar computation shows that (4.8.40) is $O(1)$, the result follows. ■

Proof of Theorem 4.6.2. For any $x \in \mathcal{S}^{d-1}$, consider the map

$$\theta \mapsto f_\theta(x) := \arccos(x'\pi_m^{-1}(\theta))$$

defined on $T_m\mathcal{S}^{d-1}$. It is easy to check that this map is differentiable on $T_m\mathcal{S}^{d-1} \setminus \{\pi_m(-x), \pi_m(x)\}$, with gradient

$$\nabla_\theta f_\theta(x) = -\frac{(J_\theta(\pi_m^{-1}))'x}{\sqrt{1-(x'\pi_m^{-1}(\theta))^2}},$$

where $J_\theta(\pi_m^{-1})$ was defined in Lemma 4.8.9. From (4.8.37), we have that, for any $h \in T_m\mathcal{S}^{d-1}$,

$$h'\nabla_\theta f_\theta(x) = -\frac{(J_\theta(\pi_m^{-1})h)'x}{\sqrt{1-(x'\pi_m^{-1}(\theta))^2}} = -\frac{(J_\theta(\pi_m^{-1})h)'Q_\theta x}{\|Q_\theta x\|},$$

with $Q_\theta := I_d - \pi_m^{-1}(\theta)(\pi_m^{-1}(\theta))'$. Since $\theta \mapsto f_\theta(x)$ is only defined on $T_m\mathcal{S}^{d-1}$, we thus have

$$\nabla_\theta f_\theta(x) = -\frac{(J_\theta(\pi_m^{-1}))'Q_\theta x}{\|Q_\theta x\|}.$$

In particular, if $x \in \mathcal{S}^{d-1} \setminus \{-m, m\}$, then the gradient of $\theta \mapsto f_\theta(x)$ at $\theta = 0$ exists and is given by

$$\eta(x) := (\nabla_\theta f_\theta(x))|_{\theta=0} = -\frac{2Q_0 x}{\|Q_0 x\|} = -\frac{2(I_d - mm')x}{\|(I_d - mm')x\|}.$$

In this case, it is easy to see that $\theta \mapsto f_\theta(x)$ is actually twice-differentiable at $\theta = 0$ with Hessian matrix

$$H(x) := (\nabla_\theta^2 f_\theta(x))|_{\theta=0} = \frac{4(m'x)}{\sqrt{1-(m'x)^2}}\Big(I_d - \frac{xx'}{1-(m'x)^2}\Big).$$

This Hessian matrix, when seen as a quadratic form on $T_m\mathcal{S}^{d-1}$, clearly rewrites

$$V(x) := \frac{4(m'x)}{\|Q_0x\|}\bigg(I_d - \frac{(Q_0x)(Q_0x)'}{\|Q_0x\|^2}\bigg), \tag{4.8.41}$$

by which we mean that $h'H(x)g = h'V(x)g$ for any $h, g \in T_m \in \mathcal{S}^{d-1}$.

Let us then show that there exists a positive constant C such that

$$|f_{\theta_2}(x) - f_{\theta_1}(x)| \leq C\|\theta_2 - \theta_1\| \tag{4.8.42}$$

for any $x \in \mathcal{S}^{d-1}$ and for any $\theta_1, \theta_2 \in T_m\mathcal{S}^{d-1}$. To this end, fix $x \in \mathcal{S}^{d-1}$ and let $\theta_1, \theta_2 \in T_m\mathcal{S}^{d-1}$ be such that $\{\theta_s := (1-s)\theta_1 + s\theta_2 : s \in [0,1]\} \subset T_m\mathcal{S}^{d-1} \setminus \{\pi_m(-x), \pi_m(x)\}$. Then,

$$\begin{aligned} f_{\theta_2}(x) - f_{\theta_1}(x) &= \int_0^1 (\theta_2 - \theta_1)'(\nabla_\theta f_\theta(x))\big|_{\theta=\theta_s}\, ds \\ &= -\int_0^1 (\theta_2 - \theta_1)' \frac{(J_{\theta_s}(\pi_m^{-1}))'Q_{\theta_s}x}{\|Q_{\theta_s}x\|}\, ds. \end{aligned}$$

This entails that

$$|f_{\theta_2}(x) - f_{\theta_1}(x)| \leq \|\theta_2 - \theta_1\| \int_0^1 \|J_{\theta_s}(\pi_m^{-1})\|_{\mathcal{L}}\, ds,$$

where $\|M\|_{\mathcal{L}} := \sup_{v \in \mathcal{S}^{d-1}} \|Mv\|$ is the operator norm of M. Since it is easy to show that $\sup_{\theta \in T_m\mathcal{S}^{d-1}} \|J_\theta(\pi_m^{-1})\|_{\mathcal{L}}$ is finite, we obtain that

$$|f_{\theta_2}(x) - f_{\theta_1}(x)| \leq C\|\theta_2 - \theta_1\| \tag{4.8.43}$$

for any $x \in \mathcal{S}^{d-1}$ and any $\theta_1, \theta_2 \in T_m\mathcal{S}^{d-1}$ such that $\{(1-s)\theta_1 + s\theta_2 : s \in [0,1]\} \subset T_m\mathcal{S}^{d-1} \setminus \{\pi_m(-x), \pi_m(x)\}$. From continuity of $\theta \mapsto f_\theta(x)$, (4.8.43) also holds for any $x \in \mathcal{S}^{d-1}$ and any $\theta_1, \theta_2 \in T_m\mathcal{S}^{d-1}$, which concludes the proof of (4.8.42).

We now want to show that $\theta \mapsto \mathbb{E}[f_\theta(X_1)]$ admits a second-order Taylor expansion at $\theta = 0$. Since $\theta \mapsto f_\theta(x)$ is twice-differentiable at $\theta = 0$ for any $x \in \mathcal{S}^{d-1} \setminus \{-m, m\}$, and since P is non-atomic, we have that

$$\lim_{\substack{\theta \to 0 \\ \theta \in T_m\mathcal{S}^{d-1}}} \frac{f_\theta(x) - f_0(x) - \theta'\eta(x) - \frac{1}{2}\theta'V(x)\theta}{\|\theta\|^2} = 0 \tag{4.8.44}$$

for P-almost every $x \in \mathcal{S}^{d-1}$. Let us now show that the collection

$$\bigg\{\frac{f_\theta(X_1) - f_0(X_1) - \theta'\eta(X_1) - \frac{1}{2}\theta'V(X_1)\theta}{\|\theta\|^2} : \theta \in T_m\mathcal{S}^{d-1},\, 0 < \|\theta\| < 1\bigg\}$$

is dominated by a P-integrable random variable. From (4.8.41),

$$\sup_{\substack{0<\|\theta\|<1 \\ \theta \in T_m\mathcal{S}^{d-1}}} \frac{|\theta'V(X_1)\theta|}{\|\theta\|^2} \leq \frac{4}{\|Q_0X_1\|} = \frac{4}{\sqrt{1-(m'X_1)^2}}$$

almost surely, where the upper-bound is P-integrable (Lemma 4.8.11), so that it is enough to show that the collection

$$\Big\{ L_\theta := \frac{f_\theta(X_1) - f_0(X_1) - \theta' \eta(X_1)}{\|\theta\|^2} : \theta \in T_m\mathcal{S}^{d-1},\, 0 < \|\theta\| < 1 \Big\}$$

is dominated by a P-integrable random variable. Fix then $\theta \in T_m\mathcal{S}^{d-1}$ with $0 < \|\theta\| < 1$. For any $x \in \mathcal{S}^{d-1}$, note that $\{s\theta : s \in [0,1]\} \subset T_m\mathcal{S}^{d-1} \setminus \{\pi_m(-x), \pi_m(x)\}$ if and only if $x \notin \{\pi_m^{-1}(s\theta) : s \in [0,1]\} \cup \{-\pi_m^{-1}(s\theta) : s \in [0,1]\}$. Since P admits a density, we have with P-probability zero that $X_1 \in \{\pi_m^{-1}(s\theta) : s \in [0,1]\} \cup \{-\pi_m^{-1}(s\theta) : s \in [0,1]\}$. This entails that, with P-probability one,

$$f_\theta(X_1) - f_0(X_1) = -\int_0^1 \frac{\theta'(J_{s\theta}(\pi_m^{-1}))' Q_{s\theta} X_1}{\|Q_{s\theta} X_1\|}\, ds,$$

hence that

$$L_\theta = \frac{1}{\|\theta\|} \int_0^1 \Big\{ \frac{2\theta' Q_0 X_1}{\|\theta\| \|Q_0 X_1\|} - \frac{\theta'(J_{s\theta}(\pi_m^{-1}))' Q_{s\theta} X_1}{\|\theta\| \|Q_{s\theta} X_1\|} \Big\}\, ds.$$

Write the integrand in this last expression as

$$\begin{aligned}
&\frac{2\theta' Q_0 X_1}{\|\theta\| \|Q_0 X_1\|} - \frac{\theta'(J_{s\theta}(\pi_m^{-1}))' Q_{s\theta} X_1}{\|\theta\| \|Q_{s\theta} X_1\|} \\
&= \frac{\theta'}{\|\theta\|} \big(2I_d - (J_{s\theta}(\pi_m^{-1}))'\big) \frac{Q_0 X_1}{\|Q_0 X_1\|} + \frac{\theta'(J_{s\theta}(\pi_m^{-1}))'}{\|\theta\|} \Big(\frac{Q_0 X_1}{\|Q_0 X_1\|} - \frac{Q_{s\theta} X_1}{\|Q_{s\theta} X_1\|} \Big).
\end{aligned}$$

On the one hand, we have that

$$\begin{aligned}
2I_d - (J_{s\theta}(\pi_m^{-1}))' &= 2I_d \Big(1 - \frac{1}{1 + s^2\|\theta\|^2}\Big) + \frac{4\theta(\theta + m)'}{(1 + s^2\|\theta\|^2)^2} \\
&= \frac{2s^2\|\theta\|^2}{1 + s^2\|\theta\|^2} I_d + \frac{4\theta(\theta + m)'}{(1 + s^2\|\theta\|^2)^2},
\end{aligned}$$

so that there exists a positive constant C such that

$$\|2I_d - (J_{s\theta}(\pi_m^{-1}))'\|_{\mathcal{L}} \leq C\|\theta\|$$

for any $s \in [0,1]$. On the other hand,

$$\Big\| \frac{Q_0 X_1}{\|Q_0 X_1\|} - \frac{Q_{s\theta} X_1}{\|Q_{s\theta} X_1\|} \Big\| \leq \frac{2\|Q_0 X_1 - Q_{s\theta} X_1\|}{\|Q_0 X_1\|} \leq \frac{4\|\pi_m^{-1}(s\theta) - m\|}{\|Q_0 X_1\|}.$$

Since π_m^{-1} is differentiable at any θ with Jacobian matrix $J_\theta(\pi_m^{-1})$, there exist $\lambda_i \in (0, s)$, $i = 1, \ldots, d$, such that, for any i,

$$(\pi_m^{-1}(s\theta) - m)_i = (\pi_m^{-1}(s\theta))_i - (\pi_m^{-1}(0))_i = s(J_{\lambda_i\theta}(\pi_m^{-1})\theta)_i,$$

which ensures that there exists some positive constant C such that

$$\|\pi_m^{-1}(s\theta) - m\| \leq C\|\theta\|$$

for any $s \in [0,1]$. We conclude that there exists a positive constant C such that, irrespective of $\theta \in T_m\mathcal{S}^{d-1}$ with $0 < \|\theta\| < 1$,

$$|L_\theta| \le C\Big(1 + \frac{1}{\|Q_0X_1\|}\Big) = C + \frac{C}{\sqrt{1-(m'X_1)^2}} \tag{4.8.45}$$

P-almost surely. Since the upper-bound in (4.8.45) is P-integrable (Lemma 4.8.11), Lebesgue's dominated convergence theorem applies and yields (recall (4.8.44))

$$\begin{aligned}
&\mathbb{E}[f_\theta(X_1) - f_0(X_1)] \\
&\quad = \theta'\mathbb{E}[\eta(X_1)\xi_{X_1,\pm m}] + \frac{1}{2}\theta'\mathrm{E}[V(X_1)\xi_{X_1,\pm m}]\theta + o(\|\theta\|^2) \\
&\quad = \theta'\mathbb{E}[\eta(X_1)\xi_{X_1,\pm m}] + \frac{1}{2}\theta'(4K)\theta + o(\|\theta\|^2)
\end{aligned}$$

as $\theta \to 0$ in $T_m\mathcal{S}^{d-1}$, where K is the matrix defined in the statement of the theorem. Let $\hat{\theta}_n := \pi_m(\hat{m}_n)$ and $\theta_0 := \pi_m(m) = 0$. By continuity of π_m, Theorem 4.6.1 entails that $\hat{\theta}_n \to \theta_0$ P-almost surely. By definition, we further have that $\hat{\theta}_n$ and θ_0 respectively minimize

$$\theta \mapsto \frac{1}{n}\sum_{i=1}^n f_\theta(X_i) \qquad \text{and} \qquad \theta \mapsto \mathrm{E}[f_\theta(X_1)]$$

over $T_m\mathcal{S}^{d-1}$. Recalling the Lipschitz result in (4.8.42) and the fact that K is invertible by assumption, Theorem 5.23 from [101] then entails that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -(4K)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \eta(X_i)\xi_{X_i,\pm m} + o_{\mathrm{P}}(1)$$

as $n \to \infty$. The Delta method then yields that

$$\begin{aligned}
\sqrt{n}(\hat{m}_n - m) &= \sqrt{n}(\pi_m^{-1}(\hat{\theta}_n) - \pi_m^{-1}(\theta_0)) \\
&= J_{\theta_0}(\pi_m^{-1})\sqrt{n}(\hat{\theta}_n - \theta_0) + o_{\mathrm{P}}(1) = 2\sqrt{n}(\hat{\theta}_n - \theta_0) + o_{\mathrm{P}}(1) \qquad (4.8.46) \\
&= K^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(-\frac{\eta(X_i)}{2}\Big)\xi_{X_i,\pm m} + o_{\mathrm{P}}(1) \\
&= K^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Q_0X_i}{\|Q_0X_i\|}\xi_{X_i,\pm m} + o_{\mathrm{P}}(1).
\end{aligned}$$

Finally, since $\hat{\theta}_n = \pi_m(\hat{m}_n)$, $\theta_0 = \pi_m(m)$, and $(I_d - mm')\pi_m(\mu) = \pi_m(\mu)$ for any $\mu \in \mathcal{S}^{d-1}_{-m}$, applying twice the asymptotic equivalence in (4.8.46) provides

$$\begin{aligned}
\sqrt{n}(\hat{m}_n - m) &= 2\sqrt{n}(\pi_m(\hat{m}_n) - \pi_m(m)) + o_{\mathrm{P}}(1) \\
&= (I_d - mm')\{2\sqrt{n}(\pi_m(\hat{m}_n) - \pi_m(m))\} + o_{\mathrm{P}}(1) \\
&= (I_d - mm')\sqrt{n}(\hat{m}_n - m) + o_{\mathrm{P}}(1),
\end{aligned}$$

which establishes the result. ■

Lemma 4.8.12. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and assume that P admits a density f with respect to the surface area measure σ_{d-1} on $\mathcal{S}^{d-1}$. Let X be an $\mathcal{S}^{d-1}$-valued random vector with distribution P. Then, for any $y \in \mathcal{S}^{d-1}$, the distribution of $\pi_y(X)$ is absolutely continuous with respect to $\mathcal{H}_{d-1}$, the $(d-1)$-dimensional Hausdorff measure in $\mathbb{R}^d$ and the corresponding density is defined by*

$$f^{\pi_y(X)}(z) = \Big(\frac{2\|z\|}{1+\|z\|^2}\Big)^{d-2} f(\pi_y^{-1}(z))$$

for any $z \in T_y\mathcal{S}^{d-1}$.

Proof of Lemma 4.8.12. Fix $y \in \mathcal{S}^{d-1}$. Let A be a Borel subset of $\pi_y(\mathcal{S}^{d-1}_{-y}) = T_y\mathcal{S}^{d-1}$. Since P is non-atomic and π_y is a homeomorphism between $\mathcal{S}^{d-1}_{-y}$ and $T_y\mathcal{S}^{d-1}$,

$$P[\pi_y(X) \in A] = \int_A f(\pi_y^{-1}(z))\, d\sigma^{\pi_y}_{d-1}(z),$$

where $\sigma^{\pi_y}_{d-1}$ denotes the push-forward measure of σ_{d-1} by π_y, that is the measure defined by $\sigma^{\pi_y}_{d-1}(B) := \sigma_{d-1}(\pi_y^{-1}(B))$ for any Borel subset B of $T_y\mathcal{S}^{d-1}$. In the sequel, we will naturally consider $T_y\mathcal{S}^{d-1}$ as the product space $[0,\infty) \times \mathcal{S}^{d-2}$ (here, $\mathcal{S}^{d-2}$ is identified to $\mathcal{S}^{d-1} \cap T_y\mathcal{S}^{d-1}$). Clearly, we also identify $\sigma^{\pi_y}_{d-1}$ to a (spherically symmetric) measure defined on this product space. For any Borel subset I of $[0,\infty)$ and Borel subset V of $\mathcal{S}^{d-2}$,

$$\pi_y^{-1}(I \times V) = \bigcup_{r \in I} \pi_y^{-1}(\{r\} \times V) = \bigcup_{r \in I} \Big(\frac{2r}{1+r^2}V + \frac{1-r^2}{1+r^2}y\Big).$$

Since these sets are Borel subsets of $(d-2)$-dimensional disjoint spheres, we have

$$\sigma^{\pi_y}_{d-1}(I \times V) = \int_I \mathcal{H}_{d-2}\Big(\frac{2r}{1+r^2}V + \frac{1-r^2}{1+r^2}y\Big)\, dr.$$

Since $\mathcal{H}_{d-2}$ is translation invariant on $\mathbb{R}^d$ and homogeneous with degree $d-2$, this yields

$$\begin{aligned}\sigma^{\pi_y}_{d-1}(I \times V) &= \int_I \mathcal{H}_{d-2}\Big(\frac{2r}{1+r^2}V\Big)\, dr \\ &= \int_I \Big(\frac{2r}{1+r^2}\Big)^{d-2} \mathcal{H}_{d-2}(V)\, dr = \int_{I\times V} \Big(\frac{2r}{1+r^2}\Big)^{d-2} dr\, d\mathcal{H}_{d-2}.\end{aligned}$$

Since $dr \otimes d\mathcal{H}_{d-2} = d\mathcal{H}_{d-1}$ on $T_y\mathcal{S}^{d-1}$, we conclude that $\sigma^{\pi_y}_{d-1}$ is absolutely continuous with respect to $\mathcal{H}_{d-1}$ with density given by

$$z \mapsto \Big(\frac{2\|z\|}{1+\|z\|^2}\Big)^{d-2}.$$

Therefore,

$$P[\pi_y(X) \in A] = \int_A f(\pi_y^{-1}(z))\Big(\frac{2\|z\|}{1+\|z\|^2}\Big)^{d-2} d\mathcal{H}_{d-1}(z),$$

which establishes the result. ■

Lemma 4.8.13. *Fix $P \in \mathcal{P}_{d-1}$, with $d \geq 3$, and assume that P admits a bounded density f with respect to the surface area measure on $\mathcal{S}^{d-1}$. Let X be an $\mathcal{S}^{d-1}$-valued random vector with distribution P and let (Y_n) be a sequence of $\mathcal{S}^{d-1}$-valued random vectors that are each independent of X. Then, for any $q \in \mathcal{S}^{d-1}$ and any $\delta \in [0, d-2)$,*

$$\mathrm{E}\bigg[\frac{\xi_{\pi_{Y_n}(X),q}}{\|\pi_{Y_n}(X) - q\|^{1+\delta}}\bigg] = O(1)$$

as n diverges to infinity (this applies in particular when $(Y_n) = (y_n)$ is an arbitrary deterministic sequence in $\mathcal{S}^{d-1}$).

Proof of Lemma 4.8.13. Since Y_n and X are independent,

$$\begin{aligned}
\mathrm{E}\bigg[\frac{\xi_{\pi_{Y_n}(X),q}}{\|\pi_{Y_n}(X) - q\|^{1+\delta}}\bigg] &\leq 1 + \mathrm{E}\bigg[\frac{\xi_{\pi_{Y_n}(X),q}}{\|\pi_{Y_n}(X) - q\|^{1+\delta}}\, \mathbb{I}[\|\pi_{Y_n}(X) - q\| \leq 1]\bigg] \\
&= 1 + \int_{\mathcal{S}^{d-1}} \mathrm{E}\bigg[\frac{\xi_{\pi_y(X),q}}{\|\pi_y(X) - q\|^{1+\delta}}\, \mathbb{I}[\|\pi_y(X) - q\| \leq 1]\bigg]\, dP^{Y_n}(y) \\
&= 1 + \int_{\mathcal{S}^{d-1}} \int_{T_y\mathcal{S}^{d-1} \cap B(q,1)} \frac{\xi_{z,q}}{\|z - q\|^{1+\delta}} f^{\pi_y(X)}(z)\, d\mathcal{H}_{d-1}(z)\, dP^{Y_n}(y),
\end{aligned}$$

where $B(z, r)$ is the closed ball centered at z with radius r. Let $q_y := (I_d - qq')y$ be the orthogonal projection of q onto $T_y\mathcal{S}^{d-1}$. Since $\|z - q_y\| \leq \|z - q\|$ and $\xi_{z,q} = \xi_{z,q_y}$ for any $z \in T_y\mathcal{S}^{d-1}$, Lemma 4.8.12 and the boundedness of f entail that there exists a positive constant C_d such that

$$\begin{aligned}
&\int_{T_y\mathcal{S}^{d-1} \cap B(q,1)} \frac{\xi_{z,q}}{\|z - q\|^{1+\delta}} f^{\pi_y(X)}(z)\, d\mathcal{H}_{d-1}(z) \\
&\qquad \leq C_d \|f\|_\infty \int_{T_y\mathcal{S}^{d-1} \cap B(q_y,1)} \frac{\xi_{z,q_y}}{\|z - q_y\|^{1+\delta}}\, d\mathcal{H}_{d-1}(z) \\
&\qquad = C_d \|f\|_\infty \int_{T_y\mathcal{S}^{d-1} \cap B(0,1)} \frac{\xi_{z,0}}{\|z\|^{1+\delta}}\, d\mathcal{H}_{d-1}(z).
\end{aligned}$$

Using hyperspherical coordinates in $T_y\mathcal{S}^{d-1}$ and denoting as σ the surface area measure on $\partial(B(0,1) \cap T_y\mathcal{S}^{d-1})$, which is the $(d-2)$-dimensional unit sphere in $T_y\mathcal{S}^{d-1}$, we have

$$\begin{aligned}
\int_{T_y\mathcal{S}^{d-1} \cap B(0,1)} \frac{\xi_{z,0}}{\|z\|^{1+\delta}}\, d\mathcal{H}_{d-1}(z) &= \int_0^1 r^{d-2} \int_{\partial(T_y\mathcal{S}^{d-1} \cap B(0,1))} \frac{\xi_{z,0}}{r^{1+\delta}}\, d\sigma(v)\, dr \\
&= \omega_{d-2} \int_0^1 r^{d-3-\delta} dr = \frac{\omega_{d-2}}{d-2-\delta},
\end{aligned}$$

where ω_{d-2} is the surface area measure of a $(d-2)$-dimensional unit sphere (by assumption, $d - 3 - \delta > -1$, which ensures integrability above). Therefore,

$$\int_{T_y\mathcal{S}^{d-1} \cap B(q,1)} \frac{\xi_{z,q}}{\|z - q\|^{1+\delta}} f^{\pi_y(X)}(z)\, d\mathcal{H}_{d-1}(z) \leq \frac{C_d \|f\|_\infty \omega_{d-2}}{d-2-\delta}$$

for any $y \in \mathcal{S}^{d-1}$, which entails that

$$\mathrm{E}\bigg[\bigg(\frac{\xi_{\pi_{Y_n}(X),q}}{\|\pi_{Y_n}(X) - q\|}\bigg)^{1+\delta}\bigg] \leq 1 + \frac{C_d \|f\|_\infty \omega_{d-2}}{d-2-\delta}$$

for any n. This concludes the proof. ∎

Lemma 4.8.14. *Let $\delta \in [0,1]$. Then, there exists a positive constant C such that*

$$(i) \qquad \left| \|x-\theta\| - \|x\| + \frac{1}{\|x\|}\theta' x \right| \leq C \min\left(\frac{\|\theta\|^2}{\|x\|}, \|\theta\| \right)$$

and

$$(ii) \qquad \left| \|x-\theta\| - \|x\| + \frac{1}{\|x\|}\theta' x - \frac{1}{2\|x\|}\theta' \left(I_d - \frac{xx'}{\|x\|^2} \right)\theta \right|$$
$$\leq C \min\left(\frac{\|\theta\|^2}{\|x\|}, \frac{\|\theta\|^3}{\|x\|^2} \right) \leq C \frac{\|\theta\|^{2+\delta}}{\|x\|^{1+\delta}}$$

for any $\theta \in \mathbb{R}^d$ and any $x \in \mathbb{R}^d \setminus \{0\}$.

Proof of Lemma 4.8.14. Part (i) of the result and the first inequality in Part (ii) of the result are Parts (ii) and (iv), respectively, in Lemma 19 from [2]. The second inequality in Part (ii) of the result follows from the fact that if $\delta \in [0,1]$, then $\min(t, t^2) \leq t^{1+\delta}$ for any $t \geq 0$. ∎

Lemma 4.8.15. *Let $\delta \in [0,1]$. Then, there exists a positive constant C such that*

$$\left\| \frac{x-\theta}{\|x-\theta\|} - \frac{x}{\|x\|} + \frac{1}{\|x\|}\left(I_d - \frac{xx'}{\|x\|^2} \right)\theta \right\| \leq C\left(\frac{\|\theta\|^{1+\delta}}{\|x\|^{1+\delta}} + \frac{\|\theta\|^{1+\delta}}{\|x-\theta\|^{1+\delta}} \right)$$

for any $\theta \in \mathbb{R}^d$ and any $x \in \mathbb{R}^d \setminus \{0, \theta\}$.

Proof of Lemma 4.8.15. Let $\Upsilon := I_d - xx'/\|x\|^2$ be the matrix of the orthogonal projection onto the orthogonal complement of $x \setminus \{0\}$ in $\mathbb{R}^d$. Writing $I_d = (xx'/\|x\|^2) + \Upsilon$ and using the fact that Υ is a symmetric and idempotent matrix, we have

$$\begin{aligned} &\left\| \frac{x-\theta}{\|x-\theta\|} - \frac{x}{\|x\|} + \frac{1}{\|x\|}\Upsilon\theta \right\|^2 \\ &\quad = \left(\left(\frac{x-\theta}{\|x-\theta\|} - \frac{x}{\|x\|} + \frac{1}{\|x\|}\Upsilon\theta \right)' \frac{x}{\|x\|} \right)^2 + \left\| \Upsilon\left(\frac{x-\theta}{\|x-\theta\|} - \frac{x}{\|x\|} + \frac{1}{\|x\|}\Upsilon\theta \right) \right\|^2 \\ &\quad = T_1 + T_2, \end{aligned}$$

say. Now,

$$\begin{aligned} T_1 &= \left(\left(\frac{x-\theta}{\|x-\theta\|} - \frac{x}{\|x\|} \right)' \frac{x}{\|x\|} \right)^2 = \left(1 - \frac{\|x\|^2 - x'\theta}{\|x\|\|x-\theta\|} \right)^2 \\ &= \frac{1}{\|x-\theta\|^2} \left(\|x-\theta\| - \|x\| + \frac{x'\theta}{\|x\|} \right)^2 \end{aligned}$$

and

$$T_2 = \left\| \Upsilon\left(\frac{-\theta}{\|x-\theta\|} + \frac{\theta}{\|x\|} \right) \right\|^2 = \left| \frac{1}{\|x-\theta\|} - \frac{1}{\|x\|} \right|^2 \|\Upsilon\theta\|^2.$$

Therefore, Lemma 4.8.14(i) yields that, for some positive constant C,

$$\begin{aligned}
&\left\| \frac{x-\theta}{\|x-\theta\|} - \frac{x}{\|x\|} + \frac{1}{\|x\|}\Upsilon\theta \right\| \\
&\qquad \leq \frac{1}{\|x-\theta\|}\left| \|x-\theta\| - \|x\| + \frac{x'\theta}{\|x\|} \right| + \left| \frac{1}{\|x-\theta\|} - \frac{1}{\|x\|} \right| \|\Upsilon\theta\| \\
&\qquad \leq \frac{C}{\|x-\theta\|} \min\left(\frac{\|\theta\|^2}{\|x\|}, \|\theta\| \right) + \frac{\|\theta\|^2}{\|x-\theta\|\|x\|} \\
&\qquad \leq \frac{(C+1)\|\theta\|^2}{\|x-\theta\|\|x\|}.
\end{aligned}$$

Since Lemma 4.8.5 in Chapter 3 readily yields

$$\left\| \frac{x-\theta}{\|x-\theta\|} - \frac{x}{\|x\|} + \frac{1}{\|x\|}\Upsilon\theta \right\| \leq \left\| \frac{x-\theta}{\|x-\theta\|} - \frac{x}{\|x\|} \right\| + \frac{\|\Upsilon\theta\|}{\|x\|} \leq 3\frac{\|\theta\|}{\|x\|},$$

it follows that

$$\left\| \frac{x-\theta}{\|x-\theta\|} - \frac{x}{\|x\|} + \frac{1}{\|x\|}\Upsilon\theta \right\| \leq (C+3)\min\left(\frac{\|\theta\|}{\|x\|}, \frac{\|\theta\|^2}{\|x-\theta\|\|x\|} \right) =: (C+3)\Delta,$$

say. If $\|x\| \leq \|x-\theta\|$, then

$$\Delta \leq \min\left(\frac{\|\theta\|}{\|x\|}, \frac{\|\theta\|^2}{\|x\|^2} \right) \leq \frac{\|\theta\|^{1+\delta}}{\|x\|^{1+\delta}}$$

for any $\delta \in [0,1]$, whereas if $\|x\| > \|x-\theta\|$, then

$$\Delta \leq \min\left(\frac{\|\theta\|}{\|x-\theta\|}, \frac{\|\theta\|^2}{\|x-\theta\|^2} \right) \leq \frac{\|\theta\|^{1+\delta}}{\|x-\theta\|^{1+\delta}}$$

for any $\delta \in [0,1]$. This entails that

$$\begin{aligned}
\Delta &\leq \frac{\|\theta\|^{1+\delta}}{\|x\|^{1+\delta}}\mathbb{I}[\|x\| \leq \|x-\theta\|] + \frac{\|\theta\|^{1+\delta}}{\|x-\theta\|^{1+\delta}}\mathbb{I}[\|x\| > \|x-\theta\|] \\
&\leq \frac{\|\theta\|^{1+\delta}}{\|x\|^{1+\delta}} + \frac{\|\theta\|^{1+\delta}}{\|x-\theta\|^{1+\delta}}
\end{aligned}$$

for any $\delta \in [0,1]$. ■

We can now prove Theorem 4.6.3.

Proof of Theorem 4.6.3. Since the vector field $\mu \mapsto U(\mu)$ is differentiable at m, the root-n consistency of $\hat{m}_n$ (Theorem 4.6.2) and the Delta method (Theorem 3.1 from 101) yield

$$\sqrt{n}(u_n - u) = \sqrt{n}(U(\hat{m}_n) - U(m)) = J_m(U)\sqrt{n}(\hat{m}_n - m) + o_{\mathrm{P}}(1).$$

Let $\hat{\mu}_n := \hat{\mu}^{\hat{m}_n}_{\alpha,u_n} =: \pi^{-1}_{\hat{m}_n}(\hat{q}_n)$ and $\mu := \mu^m_{\alpha,u} = \pi^{-1}_m(q)$ (note that q was already defined in the statement of Theorem 4.6.3). By definition,

$$\hat{q}_n = \text{argmin}_{z\in\mathbb{R}^d} O^{\pi_{\hat{m}_n}\#P'_n}_{\alpha,u_n}(z) \quad \text{and} \quad q = \text{argmin}_{z\in\mathbb{R}^d} O^{\pi_m\#P}_{\alpha,u}(z),$$

where P'_n denotes the empirical probability measure associated with X_i, $i = \lfloor cn \rfloor + 1, \ldots, n$. Write then

$$\begin{aligned} \sqrt{n}(\hat{\mu}_n - \mu) &= \sqrt{n}\big(\pi^{-1}_{\hat{m}_n}(\hat{q}_n) - \pi^{-1}_m(q)\big) \\ &= \sqrt{n}\big(\pi^{-1}_{\hat{m}_n}(\hat{q}_n) - \pi^{-1}_m(\hat{q}_n)\big) + \sqrt{n}\big(\pi^{-1}_m(\hat{q}_n) - \pi^{-1}_m(q)\big) \\ &=: W_{1n} + W_{2n}, \end{aligned}$$

say. Assume that $\sqrt{n}(\hat{q}_n - q)$ converges in distribution (this will be established below). The Delta method then yields that $W_{2n} = J_q(\pi^{-1}_m)\sqrt{n}(\hat{q}_n - q) + o_{\mathrm{P}}(1)$. For W_{1n}, Lemma 4.8.9 provides

$$\begin{aligned} \pi^{-1}_{\hat{m}_n}(\hat{q}_n) - \pi^{-1}_m(\hat{q}_n) &= \left(\frac{2}{1+\|\hat{q}_n\|^2}(\hat{q}_n + \hat{m}_n) - \hat{m}_n\right) - \left(\frac{2}{1+\|\hat{q}_n\|^2}(\hat{q}_n + m) - m\right) \\ &= \frac{1-\|\hat{q}_n\|^2}{1+\|\hat{q}_n\|^2}(\hat{m}_n - m). \end{aligned}$$

Since $\sqrt{n}(\hat{q}_n - q)$ converges in distribution, $\hat{q}_n$ converges to q in probability, which yields (recall that $q = \pi_m(\mu)$)

$$\begin{aligned} W_{1n} &= \sqrt{n}\big(\pi^{-1}_{\hat{m}_n}(\hat{q}_n) - \pi^{-1}_m(\hat{q}_n)\big) = \frac{1-\|q\|^2}{1+\|q\|^2}\sqrt{n}(\hat{m}_n - m) + o_{\mathrm{P}}(1) \\ &= \frac{1-\left(\frac{1-m'\mu}{1+m'\mu}\right)}{1+\left(\frac{1-m'\mu}{1+m'\mu}\right)}\sqrt{n}(\hat{m}_n - m) + o_{\mathrm{P}}(1) = (m'\mu)\sqrt{n}(\hat{m}_n - m) + o_{\mathrm{P}}(1). \end{aligned}$$

Thus, we obtain that

$$\sqrt{n}(\hat{\mu}_n - \mu) = (m'\mu)\sqrt{n}(\hat{m}_n - m) + J_q(\pi^{-1}_m)\sqrt{n}(\hat{q}_n - q) + o_{\mathrm{P}}(1). \tag{4.8.47}$$

We now derive the asymptotic behaviour of $\sqrt{n}(\hat{q}_n - q)$, which we will do by applying Theorem 3 from [2]. To do so, the main challenge is to check that Condition (iii) in that theorem holds in the present situation, that is, writing $k_n := n - \lfloor cn \rfloor$ for any n, to check that, for any $h \in \mathbb{R}^d$,

$$k_n\left\{O^{\pi_{\hat{m}_n}\#P'_n}_{\alpha,u_n}\left(q + \frac{h}{\sqrt{k_n}}\right) - O^{\pi_{\hat{m}_n}\#P'_n}_{\alpha,u_n}(q)\right\} - h'\eta_n - h'V_n h = o_{\mathrm{P}}(1),$$

for some sequence of random d-vectors η_n that is $O_{\mathrm{P}}(1)$ and some sequence of invertible symmetric matrices (V_n) with all eigenvalues eventually in a fixed compact subset of $(0, \infty)$. By proceeding as in the proof of Lemma 3.9.23 from Chapter 3, one readily obtains that

$$k_n\left\{O^{\pi_m\#P'_n}_{\alpha,u}\left(q + \frac{h}{\sqrt{k_n}}\right) - O^{\pi_m\#P'_n}_{\alpha,u}(q)\right\} = h'\zeta_n + \frac{1}{2}h'Vh + o_{\mathrm{P}}(1), \tag{4.8.48}$$

where we let

$$\zeta_n := -\frac{1}{\sqrt{k_n}}\sum_{i=\lfloor cn \rfloor+1}^{n}\left(\frac{\pi_m(X_i) - q}{\|\pi_m(X_i) - q\|}\xi_{\pi_m(X_i),q} + \alpha u\right)$$

and where V is the matrix defined in the statement of Theorem 4.6.3. Consequently, it is enough to study the asymptotic behaviour of

$$D_n := k_n\Big\{O^{\pi_{\hat{m}_n}\#P'_n}_{\alpha,u_n}\Big(q+\frac{h}{\sqrt{k_n}}\Big)-O^{\pi_{\hat{m}_n}\#P'_n}_{\alpha,u_n}(q)\Big\}-k_n\Big\{O^{\pi_m\#P'_n}_{\alpha,u}\Big(q+\frac{h}{\sqrt{k_n}}\Big)-O^{\pi_m\#P'_n}_{\alpha,u}(q)\Big\}.$$

Writing $h_n := h/\sqrt{k_n}$, decompose D_n into

$$\begin{aligned}
D_n = & \sum_{i=\lfloor cn\rfloor+1}^{n} \big\{(\|\pi_{\hat{m}_n}(X_i) - q - h_n\| - \alpha u_n'(q+h_n)) - (\|\pi_{\hat{m}_n}(X_i) - q\| - \alpha u_n' q)\big\} \\
& \quad - \sum_{i=\lfloor cn\rfloor+1}^{n} \big\{(\|\pi_m(X_i) - q - h_n\| - \alpha u'(q+h_n)) - (\|\pi_m(X_i) - q\| - \alpha u' q)\big\} \\
& = -\alpha\sqrt{k_n}h'(u_n - u) \\
& \quad + \sum_{i=\lfloor cn\rfloor+1}^{n} (\|\pi_{\hat{m}_n}(X_i) - q - h_n\| - \|\pi_{\hat{m}_n}(X_i) - q\|)\xi_{\pi_{\hat{m}_n}(X_i),q} \\
& \quad - \sum_{i=\lfloor cn\rfloor+1}^{n} (\|\pi_m(X_i) - q - h_n\| - \|\pi_m(X_i) - q\|)\xi_{\pi_m(X_i),q} + o_{\mathrm{P}}(1) \\
& =: -\alpha\sqrt{k_n}h'(u_n - u) + T_{1n} - T_{2n} + o_{\mathrm{P}}(1),
\end{aligned}$$

where we used the assumption that P admits a density on the unit sphere. Now, fix $\delta \in (0,1)$ arbitrarily. Lemma 4.8.14(ii) entails that

$$\begin{aligned}
T_{1n} = & -\frac{1}{\sqrt{k_n}} \sum_{i=\lfloor cn\rfloor+1}^{n} \frac{h'(\pi_{\hat{m}_n}(X_i) - q)}{\|\pi_{\hat{m}_n}(X_i) - q\|}\xi_{\pi_{\hat{m}_n}(X_i),q} \\
& + \frac{1}{2k_n} \sum_{i=\lfloor cn\rfloor+1}^{n} \frac{\xi_{\pi_{\hat{m}_n}(X_i),q}}{\|\pi_{\hat{m}_n}(X_i) - q\|} h'\Big(I_d - \frac{(\pi_{\hat{m}_n}(X_i) - q)(\pi_{\hat{m}_n}(X_i) - q)'}{\|\pi_{\hat{m}_n}(X_i) - q\|^2}\Big)h + R_{1n},
\end{aligned}$$

with

$$|R_{1n}| \le C\frac{\|h\|^{2+\delta}}{k_n^{1+\delta/2}} \sum_{i=\lfloor cn\rfloor+1}^{n} \frac{\xi_{\pi_{\hat{m}_n}(X_i),q}}{\|\pi_{\hat{m}_n}(X_i) - q\|^{1+\delta}}.$$

Similarly,

$$\begin{aligned}
T_{2n} = & -\frac{1}{\sqrt{k_n}} \sum_{i=\lfloor cn\rfloor+1}^{n} \frac{h'(\pi_m(X_i) - q)}{\|\pi_m(X_i) - q\|}\xi_{\pi_m(X_i),q} \\
& + \frac{1}{2k_n} \sum_{i=\lfloor cn\rfloor+1}^{n} \frac{\xi_{\pi_m(X_i),q}}{\|\pi_m(X_i) - q\|} h'\Big(I_d - \frac{(\pi_m(X_i) - q)(\pi_m(X_i) - q)'}{\|\pi_m(X_i) - q\|^2}\Big)h + R_{2n},
\end{aligned}$$

with

$$|R_{2n}| \le C\frac{\|h\|^{2+\delta}}{k_n^{1+\delta/2}} \sum_{i=\lfloor cn\rfloor+1}^{n} \frac{\xi_{\pi_m(X_i),q}}{\|\pi_m(X_i) - q\|^{1+\delta}}.$$

Thanks to the sample splitting scheme, Lemma 4.8.13 implies that

$$\mathrm{E}[|R_{1n}|] \leq C\frac{\|h\|^{2+\delta}}{k_n^{\delta/2}}\mathrm{E}\bigg[\frac{\xi_{\pi_{\hat{m}_n}(X_1),q}}{\|\pi_{\hat{m}_n}(X_1)-q\|^{1+\delta}}\bigg] = o(1),$$

so that $R_{1n} = o_{\mathrm{P}}(1)$. The same argument yields that $R_{2n} = o_{\mathrm{P}}(1)$. Now, observe that, letting X have distribution P and be independent of $\hat{m}_n$, we have that, for any n,

$$\begin{aligned}
\mathrm{E}\bigg[\bigg|\frac{1}{2k_n}\sum_{i=\lfloor cn\rfloor+1}^{n}\frac{\xi_{\pi_{\hat{m}_n}(X_i),q}}{\|\pi_{\hat{m}_n}(X_i)-q\|}h'\bigg(I_d-\frac{(\pi_{\hat{m}_n}(X_i)-q)(\pi_{\hat{m}_n}(X_i)-q)'}{\|\pi_{\hat{m}_n}(X_i)-q\|^2}\bigg)h&\\
-\frac{1}{2k_n}\sum_{i=\lfloor cn\rfloor+1}^{n}\frac{\xi_{\pi_m(X_i),q}}{\|\pi_m(X_i)-q\|}h'\bigg(I_d-\frac{(\pi_m(X_i)-q)(\pi_m(X_i)-q)'}{\|\pi_m(X_i)-q\|^2}\bigg)h\bigg|\bigg]&\\
\leq \frac{1}{2}\mathrm{E}\bigg[\bigg|\frac{\xi_{\pi_{\hat{m}_n}(X),q}}{\|\pi_{\hat{m}_n}(X)-q\|}h'\bigg(I_d-\frac{(\pi_{\hat{m}_n}(X)-q)(\pi_{\hat{m}_n}(X)-q)'}{\|\pi_{\hat{m}_n}(X)-q\|^2}\bigg)h&\\
-\frac{\xi_{\pi_m(X),q}}{\|\pi_m(X)-q\|}h'\bigg(I_d-\frac{(\pi_m(X)-q)(\pi_m(X)-q)'}{\|\pi_m(X)-q\|^2}\bigg)h\bigg|\bigg]&\\
=: \frac{1}{2}\mathrm{E}[|g_q(\pi_{\hat{m}_n}(X))-g_q(\pi_m(X))|],&
\end{aligned}$$

say. Since $\hat{m}_n \to m$ almost surely (Theorem 4.6.1), we have that $\pi_{\hat{m}_n}(X) \to \pi_m(X)$ almost surely. From continuity of g_q on $\mathbb{R}^d \setminus \{q\}$ (note that, since X admits a density, $\pi_m(X) \in \mathbb{R}^d \setminus \{q\}$ almost surely), this entails that $g_q(\pi_{\hat{m}_n}(X)) \to g_q(\pi_m(X))$ almost surely, hence also in probability. Now, Lemma 4.8.13 shows that

$$\mathrm{E}[|g_q(\pi_{\hat{m}_n}(X))|^{1+\delta}] \leq \|h\|^{2(1+\delta)}\mathrm{E}\bigg[\frac{\xi_{\pi_{\hat{m}_n}(X),q}}{\|\pi_{\hat{m}_n}(X)-q\|^{1+\delta}}\bigg] = O(1),$$

which implies that $(g_q(\pi_{\hat{m}_n}(X)))$ is uniformly integrable. Therefore, Vitali's theorem entails that $\mathrm{E}[|g_q(\pi_{\hat{m}_n}(X))-g_q(\pi_m(X))|] = o(1)$. We conclude that

$$\begin{aligned}
D_n &= -\alpha\sqrt{k_n}h'(u_n-u)\\
&\quad -\frac{1}{\sqrt{k_n}}h'\sum_{i=\lfloor cn\rfloor+1}^{n}\bigg\{\frac{\pi_{\hat{m}_n}(X_i)-q}{\|\pi_{\hat{m}_n}(X_i)-q\|}\xi_{\pi_{\hat{m}_n}(X_i),q}-\frac{\pi_m(X_i)-q}{\|\pi_m(X_i)-q\|}\xi_{\pi_m(X_i),q}\bigg\}+o_{\mathrm{P}}(1)\\
&= -\alpha\sqrt{k_n}h'(u_n-u)\\
&\quad -\frac{1}{\sqrt{k_n}}h'\sum_{i=\lfloor cn\rfloor+1}^{n}\bigg\{\frac{\pi_{\hat{m}_n}(X_i)-q}{\|\pi_{\hat{m}_n}(X_i)-q\|}-\frac{\pi_m(X_i)-q}{\|\pi_m(X_i)-q\|}\bigg\}\\
&\qquad\qquad\times\xi_{\pi_{\hat{m}_n}(X_i),q}\xi_{\pi_m(X_i),q}\xi_{X_i,-\hat{m}_n}\xi_{X_i,-m}+o_{\mathrm{P}}(1),
\end{aligned}$$

where we used again the absolute continuity assumption. Writing

$$\frac{\pi_y(X_i)-q}{\|\pi_y(X_i)-q\|}\xi_{\pi_y(X_i),q}\xi_{X_i,-y} = \frac{(1+y'X_i)(\pi_y(X_i)-q)}{\|(1+y'X_i)(\pi_y(X_i)-q)\|}\xi_{\pi_y(X_i),q}\xi_{X_i,-y},$$

for both $y = \hat{m}_n$ and $y = m$, Lemma 4.8.15 yields

$$-\frac{1}{\sqrt{k_n}} h' \sum_{i=\lfloor cn \rfloor+1}^{n} \left\{ \frac{\pi_{\hat{m}_n}(X_i) - q}{\|\pi_{\hat{m}_n}(X_i) - q\|} - \frac{\pi_m(X_i) - q}{\|\pi_m(X_i) - q\|} \right\} \xi_{\pi_{\hat{m}_n}(X_i),q} \xi_{\pi_m(X_i),q} \xi_{X_i,-m} \xi_{X_i,-\hat{m}_n}$$

$$= \frac{1}{\sqrt{k_n}} \sum_{i=\lfloor cn \rfloor+1}^{n} \frac{\xi_{\pi_{\hat{m}_n}(X_i),q} \xi_{\pi_m(X_i),q} \xi_{X_i,-\hat{m}_n} \xi_{X_i,-m}}{(1 + m'X_i)\|\pi_m(X_i) - q\|}$$

$$\times h' \bigg(I_d - \frac{(\pi_m(X_i) - q)(\pi_m(X_i) - q)'}{\|\pi_m(X_i) - q\|^2} \bigg) \theta_{ni} + S_n,$$

where

$$\begin{aligned} \theta_{ni} &:= (1 + m'X_i)(\pi_m(X_i) - q) - (1 + \hat{m}_n'X_i)(\pi_{\hat{m}_n}(X_i) - q) \\ &= (\hat{m}_n'X_i)\hat{m}_n + (\hat{m}_n'X_i)q - (m'X_i)m - (m'X_i)q \\ &= \big((\hat{m}_n - m)'X_i\big)\hat{m}_n + (m'X_i)(\hat{m}_n - m) + \big((\hat{m}_n - m)'X_i\big)q \\ &= \big\{(m'X_i)I_d + (\hat{m}_n + q)X_i'\big\}(\hat{m}_n - m) \end{aligned}$$

and where

$$\begin{aligned} |S_n| \leq \ & \frac{C}{\sqrt{k_n}} \sum_{i=\lfloor cn \rfloor+1}^{n} \frac{\xi_{\pi_m(X_i),q} \xi_{X_i,-m} \|\theta_{ni}\|^{1+\delta}}{(1 + m'X_i)^{1+\delta} \|\pi_m(X_i) - q\|^{1+\delta}} \\ & + \frac{C}{\sqrt{k_n}} \sum_{i=\lfloor cn \rfloor+1}^{n} \frac{\xi_{\pi_{\hat{m}_n}(X_i),q} \xi_{X_i,-\hat{m}_n} \|\theta_{ni}\|^{1+\delta}}{(1 + \hat{m}_n'X_i)^{1+\delta} \|\pi_{\hat{m}_n}(X_i) - q\|^{1+\delta}} =: C(S_{1n} + S_{2n}), \end{aligned}$$

say. For some positive constant $\tilde{C}$, we have

$$\begin{aligned} S_{2n} \ &\leq \ \tilde{C} \|\sqrt{k_n}(\hat{m}_n - m)\|^{1+\delta} \bigg(\frac{1}{k_n^{1+\frac{\delta}{2}}} \sum_{i=\lfloor cn \rfloor+1}^{n} \frac{\xi_{\pi_{\hat{m}_n}(X_i),q} \xi_{X_i,-\hat{m}_n}}{(1 + \hat{m}_n'X_i)^{1+\delta} \|\pi_{\hat{m}_n}(X_i) - q\|^{1+\delta}} \bigg) \\ &= \ O_{\mathrm{P}}(1) \bigg(\frac{1}{k_n^{1+\frac{\delta}{2}}} \sum_{i=\lfloor cn \rfloor+1}^{n} \frac{\xi_{\pi_{\hat{m}_n}(X_i),q} \xi_{X_i,-\hat{m}_n}}{(1 + \hat{m}_n'X_i)^{1+\delta} \|\pi_{\hat{m}_n}(X_i) - q\|^{1+\delta}} \bigg), \end{aligned} \quad (4.8.49)$$

where we used the fact that k_n is of order n as $n \to \infty$. Let us show that, still with X having distribution P and being independent of $\hat{m}_n$, we have

$$\mathrm{E}\bigg[\frac{\xi_{\pi_{\hat{m}_n}(X),q} \xi_{X,-\hat{m}_n}}{(1 + \hat{m}_n'X)^{1+\delta} \|\pi_{\hat{m}_n}(X) - q\|^{1+\delta}} \bigg] = O(1).$$

Let $\varepsilon > 0$ be such that $\|\pi_y(x)\| > 2\|q\|$ for any $x, y \in \mathcal{S}^{d-1}$ with $y'x \leq -1 + \varepsilon$. We have

$$\begin{aligned} & \frac{\xi_{\pi_{\hat{m}_n}(X),q} \xi_{X,-\hat{m}_n}}{(1 + \hat{m}_n'X)^{1+\delta} \|\pi_{\hat{m}_n}(X) - q\|^{1+\delta}} \\ & \qquad \leq \frac{\xi_{\pi_{\hat{m}_n}(X),q} \xi_{X,-\hat{m}_n}}{(1 + \hat{m}_n'X)^{1+\delta} \|\pi_{\hat{m}_n}(X) - q\|^{1+\delta}} \mathbb{I}[\hat{m}_n'X \leq -1 + \varepsilon] \\ & \qquad \quad + \frac{\xi_{\pi_{\hat{m}_n}(X),q}}{\varepsilon^{1+\delta} \|\pi_{\hat{m}_n}(X) - q\|^{1+\delta}} \mathbb{I}[\hat{m}_n'X > -1 + \varepsilon]. \end{aligned} \quad (4.8.50)$$

When $\hat{m}_n' X \leq -1+\varepsilon$, we have $\|\pi_{\hat{m}_n}(X)\| > 2\|q\|$, hence also

$$0 < \frac{\|\pi_{\hat{m}_n}(X)\|}{\|\pi_{\hat{m}_n}(X)\| - \|q\|} \leq 2,$$

which yields

$$\begin{aligned}
&\frac{\xi_{\pi_{\hat{m}_n}(X),q}\xi_{X,-\hat{m}_n}}{(1+\hat{m}_n'X)^{1+\delta}\|\pi_{\hat{m}_n}(X)-q\|^{1+\delta}}\mathbb{I}[\hat{m}_n'X \leq -1+\varepsilon] \\
&\qquad \leq \frac{\xi_{X,-\hat{m}_n}}{(1+\hat{m}_n'X)^{1+\delta}\|\pi_{\hat{m}_n}(X)\|^{1+\delta}}\left(\frac{\|\pi_{\hat{m}_n}(X)\|}{\|\pi_{\hat{m}_n}(X)\| - \|q\|}\right)^{1+\delta}\mathbb{I}[\hat{m}_n'X \leq -1+\varepsilon] \\
&\qquad = \frac{2^{1+\delta}\xi_{X,-\hat{m}_n}}{\|X-(\hat{m}_n'X)\hat{m}_n\|^{1+\delta}}\mathbb{I}[\hat{m}_n'X \leq -1+\varepsilon] \\
&\qquad = \frac{2^{1+\delta}\xi_{X,-\hat{m}_n}}{(1-(\hat{m}_n'X)^2)^{(1+\delta)/2}}\mathbb{I}[\hat{m}_n'X \leq -1+\varepsilon].
\end{aligned}$$

Plugging in (4.8.50), then using Lemmas 4.8.11 and 4.8.13, we obtain that

$$\begin{aligned}
&\mathrm{E}\left[\frac{\xi_{\pi_{\hat{m}_n}(X),q}\xi_{X,-\hat{m}_n}}{(1+\hat{m}_n'X)^{1+\delta}\|\pi_{\hat{m}_n}(X)-q\|^{1+\delta}}\right] \\
&\qquad\qquad \leq \mathrm{E}\left[\frac{2^{1+\delta}\xi_{X,\pm\hat{m}_n}}{(1-(\hat{m}_n'X)^2)^{(1+\delta)/2}}\right] + \mathrm{E}\left[\frac{\xi_{\pi_{\hat{m}_n}(X),q}}{\varepsilon^{1+\delta}\|\pi_{\hat{m}_n}(X)-q\|^{1+\delta}}\right] = O(1).
\end{aligned}$$

This shows that

$$\mathrm{E}\left[\frac{1}{k_n^{1+\frac{\delta}{2}}}\sum_{i=\lfloor cn\rfloor+1}^{n}\frac{\xi_{\pi_{\hat{m}_n}(X_i),q}\xi_{X_i,-\hat{m}_n}}{(1+\hat{m}_n'X_i)^{1+\delta}\|\pi_{\hat{m}_n}(X_i)-q\|^{1+\delta}}\right] = o(1),$$

hence that $S_{2n} = o_{\mathrm{P}}(1)$; see (4.8.49). The same argument shows that $S_{1n} = o_{\mathrm{P}}(1)$, so that $S_n = o_{\mathrm{P}}(1)$. It follows that

$$\begin{aligned}
D_n &= -\alpha\sqrt{k_n}h'(u_n-u) \\
&\quad + \frac{1}{\sqrt{k_n}}\sum_{i=\lfloor cn\rfloor+1}^{n}\frac{\xi_{\pi_m(X_i),q}\xi_{X_i,-m}}{(1+m'X_i)\|\pi_m(X_i)-q\|}h'\left(I_d - \frac{(\pi_m(X_i)-q)(\pi_m(X_i)-q)'}{\|\pi_m(X_i)-q\|^2}\right)\theta_{ni} + o_{\mathrm{P}}(1),
\end{aligned}$$

with $\theta_{ni} = \{(m'X_i)I_d + (\hat{m}_n+q)X_i'\}(\hat{m}_n - m)$. Of course, this rewrites

$$D_n = -\alpha\sqrt{k_n}h'(u_n-u) + h'L_n\sqrt{k_n}(\hat{m}_n - m) + o_{\mathrm{P}}(1),$$

with

$$\begin{aligned}
L_n &:= \frac{1}{k_n}\sum_{i=\lfloor cn\rfloor+1}^{n}\frac{\xi_{\pi_m(X_i),q}\xi_{X_i,-m}}{(1+m'X_i)\|\pi_m(X_i)-q\|} \\
&\qquad\qquad \times\left(I_d - \frac{(\pi_m(X_i)-q)(\pi_m(X_i)-q)'}{\|\pi_m(X_i)-q\|^2}\right)\{(m'X_i)I_d + (\hat{m}_n+q)X_i'\}.
\end{aligned}$$

Since

$$\mathrm{E}\Big[\frac{\xi_{\pi_m(X),q}\xi_{X,-m}}{(1+m'X)^{1+\delta}\|\pi_m(X)-q\|^{1+\delta}}\Big] < \infty$$

(this is proved when showing that $S_{1n} = o_{\mathrm{P}}(1)$ above), L_n converges in probability to the matrix L defined in the statement of the theorem. Thus,

$$D_n \quad = \quad -\alpha\sqrt{k_n}h'(u_n - u) + h'L\sqrt{k_n}(\hat{m}_n - m) + o_{\mathrm{P}}(1).$$

Jointly with (4.8.48), this yields

$$k_n\Big\{O^{\pi_{\hat{m}_n}\#P_n}_{\alpha,u_n}\Big(q + \frac{h}{\sqrt{k_n}}\Big) - O^{\pi_{\hat{m}_n}\#P_n}_{\alpha,u_n}(q)\Big\}$$

$$= h'\{\zeta_n + L\sqrt{k_n}(\hat{m}_n - m) - \alpha\sqrt{k_n}(u_n - u)\} + \frac{1}{2}h'Vh + o_{\mathrm{P}}(1).$$

Applying Theorem 3 from [2] finally provides

$$\sqrt{k_n}(\hat{q}_n - q) = -V^{-1}\Big(\zeta_n + L\sqrt{k_n}(\hat{m}_n - m) - \alpha\sqrt{k_n}(u_n - u)\Big) + o_{\mathrm{P}}(1).$$

which yields

$$\sqrt{n}(\hat{q}_n - q) \quad = \quad -V^{-1}\Big((1-c)^{-1/2}\zeta_n + L\sqrt{n}(\hat{m}_n - m) - \alpha\sqrt{n}(u_n - u)\Big) + o_{\mathrm{P}}(1).$$

Thus, plugging in (4.8.47), we obtain

$$\begin{aligned}\sqrt{n}(\hat{\mu}_n - \mu) &= (m'\mu)\sqrt{n}(\hat{m}_n - m) + J_q(\pi_m^{-1})\sqrt{n}(\hat{q}_n - q) + o_{\mathrm{P}}(1)\\ &= \big\{(m'\mu)I_d - J_q(\pi_m^{-1})V^{-1}L\big\}\sqrt{n}(\hat{m}_n - m) - (1-c)^{-1/2}J_q(\pi_m^{-1})V^{-1}\zeta_n\\ &\qquad +\alpha J_q(\pi_m^{-1})V^{-1}\sqrt{n}(u_n - u) + o_{\mathrm{P}}(1)\\ &= \big\{(m'\mu)I_d + J_q(\pi_m^{-1})V^{-1}(\alpha J_m(U) - L)\big\}\sqrt{n}(\hat{m}_n - m)\\ &\qquad -(1-c)^{-1/2}J_q(\pi_m^{-1})V^{-1}\zeta_n + o_{\mathrm{P}}(1),\end{aligned}$$

which establishes the result. ■

Proofs for Section 4.7

The proof of Theorem 4.7.1 requires the following rank analog of Theorem 4.5.3.

Lemma 4.8.16. *Fix $P \in \mathcal{P}_{d-1}$ and let m be a Fréchet median of P. Assume that P admits a density on $\mathcal{S}^{d-1}$. Let O be a $d \times d$ orthogonal matrix such that $Om = m$ and denote as P_O the distribution of OX when X has distribution P. Then, $R^m_{P_O}(z) = OR^m_P(O'z)$ for any $z \in \mathcal{S}^{d-1}$.*

Proof of Lemma 4.8.16. Under the assumptions of the lemma, quantiles are uniquely defined (Theorem 4.3.1(ii)). For $\alpha \in [0,1]$ and unit vector u in $T_m\mathcal{S}^{d-1}$, Theorem 4.5.3 then guarantees that $Q^m_{P_O}(O\alpha u) = Q^m_{P_O}(\alpha Ou) = \mu^m_{\alpha,Ou}(P_O) = O\mu^m_{\alpha,u}(P) =$

$OQ_P^m(\alpha u)$, i.e., that the mappings $z \mapsto Q_{P_O}^m(Oz)$ and $z \mapsto OQ_P^m(z)$ do coincide. The corresponding inverse mappings, which, in view of Theorem 4.4.5(ii), are $z \mapsto O' R_{P_O}^m(z)$ and $z \mapsto R_P^m(O'z)$, respectively, therefore also coincide, which establishes the result. ■

Proof of Theorem 4.7.1. We use again the notation $z_{\varphi,u}^m = (\cos\varphi)m + (\sin\varphi)u$. If P is rotationally symmetric with respect to m, then, as explained in Section 4.7, $R_P^m(z_{\varphi,u}^m) = \lambda_\varphi u$ for any $\varphi \in (0,\pi)$ and $u \in \mathcal{U}_m$, which implies that $T_P^m = 0$. Assume now that $T_m^P = 0$. Then,

$$R_P^m(z_{\varphi,u}^m) = \left(\int_{\mathcal{U}_m} \|R_P(z_{\varphi,v}^m)\| \, d\sigma_m(v) \right) u =: c_\varphi u$$

for any $(\varphi, u) \in \mathcal{A}$, where $\mathcal{A}$ is a subset of $(0,\pi) \times \mathcal{U}_m$ whose complement has measure zero for $\lambda \times \sigma_m$ (here, λ denotes the Lebesgue measure on $(0,\pi)$). Now, fix an arbitrary $d \times d$ orthogonal matrix O such that $Om = m$ and denote as P_O the distribution of OX when X has distribution P. Then,

$$R_{P_O}^m(z_{\varphi,u}^m) = OR_P^m(O' z_{\varphi,u}^m) = OR_P^m(z_{\varphi,O'u}^m) = O(c_\varphi O'u) = R_P^m(z_{\varphi,u}^m)$$

for any $(\varphi, u) \in \mathcal{A}$. From continuity (Theorem 4.4.4), it follows that $R_{P_O}^m(z) = R_P^m(z)$ for any $z \in \mathcal{S}^{d-1}$, which, from Theorem 4.4.6(i), implies that $P_O = P$. Since this holds for any $d \times d$ orthogonal matrix O such that $Om = m$, we conclude that P is rotationally symmetric about m. ■

Chapter 5

Conclusions and perspectives

In the first chapter, we explored the possibility of recovering a probability measure P over $\mathbb{R}^n$ via its geometric rank R_P only. We showed that this can always be done, via a (potentially fractional) linear PDE of order n. We thoroughly investigated the properties of this PDE, and discovered features about multivariate geometric ranks (some of which were really surprising). In the last two sections, we showed that convergence of geometric ranks characterizes convergence in distribution of the underlying probability measures, in the same way the univariate cdf does. We also proved that a Glivenko-Cantelli result holds for geometric ranks provided the limit distribution has no atoms. Letting

$$d_{\mathrm{G}}(P,Q) := \sup_{x\in\mathbb{R}^n} \|R_P(x) - R_Q(x)\|$$

for any probability measures P and Q, the results of the last two sections imply that d_{G} is a probability metric that characterizes weak convergence. This metric is finer than the popular Wasserstein distance d_{W}. Indeed, in addition to convergence in distribution, convergence in the Wasserstein distance requires the convergence of some moment. Furthermore, d_{G} is much simpler to compute than d_{W} since it is given in closed form. This calls for exploring how meaningful bounds on d_{G} can be derived. For instance, one expects to be able to bound d_{G} by d_{W}. This could lead to new and more practical ways to bound the distance between probability measures and, therefore, derive quantitative weak convergence results.

In the second chapter, we investigated the properties of the geometric ρ-quantiles in Definition 3.1.1. In the considered setup, this arguably settles the probabilistic study of these quantiles. However, our work naturally calls for an extension to more general setups, and for applications of these quantiles. As mentioned in the introduction, the geometric quantiles from [17] are flexible objects that can cope with more exotic types of data, such as functional data. This is associated with the fact that these quantiles are defined as minimizers of an objective function (see (3.1.1)) that involves norms and inner products only, hence that also makes sense in Hilbert spaces. This, however, is also the case for the objective function defining ρ-quantiles in (3.1.2), so that it would be natural to investigate the properties of ρ-quantiles for random variables taking values in Hilbert spaces and to compare their properties with those of the classical geometric quantiles; we refer to [13], [14], and [15] for results on the geometric median and geometric quantiles in infinite-dimensional spaces.

Another direction for future research is related to inferential applications. As mentioned in the introduction of the second chapter, the geometric quantiles from [17] and the companion geometric depth have been used in a quantile regression framework ([16], [20], [23]), and it would be interesting to consider ρ-quantiles in this setup. In particular, this would provide a geometric concept of multiple-output expectile regression, which would be quite natural since expectiles were originally introduced in [77] as an L_2-alternative to the traditional L_1-concept of quantile regression ([49]). Another natural venue for application of ρ-quantiles and ρ-depth is supervised classification. In the last decade, supervised classification based on depth - where a new observation is classified into the population with respect to which it is deepest - has met much success in the literature; see, e.g., [58], [85], and the references therein. In this framework, L_p-depths provide natural tools to implement this max-depth approach where p might be chosen through cross-validation. Such applications, or the application of ρ-quantiles in risk assessment, deserve a full-fledged paper, hence are left for future work.

The final chapter introduces a concept of quantiles for probability measures on unit spheres of arbitrary dimension d, as well as companion concepts of ranks and depth. The proposed objects show the excellent flexibility and computability properties of their geometric Euclidean antecedents. On the theoretical side, our investigation of these concepts is rather complete, although, as far as asymptotics is concerned, a stochastic process version of Theorem 4.6.3 could probably be obtained. It might thus be mainly on the methodological side that this work calls for follow-ups, and the perspectives for future research in this direction are very diverse. In particular, while we focused in Section 4.7 on testing for rotational symmetry, it should be clear that the proposed concepts may be useful in other inferential applications, too. For instance, it would be of interest to see how the proposed quantiles can be used to perform supervised classification on the sphere. Also, in the Euclidean case, geometric quantiles were used with much success to perform multiple-output quantile regression, and it is therefore natural to use our spherical quantiles to define quantile regression methods in cases where responses take values in the unit sphere.

Bibliography

[1] C. Agostinelli and M. Romanazzi. Nonparametric analysis of directional data based on data depth. *Environ. Ecol. Stat.*, 20:253–270, 2013.

[2] M. Arcones. Asymptotic theory for M-estimators over a convex kernel. *Econometric Theory*, 14:387–422, 1998.

[3] A. Azzalini and A. Capitanio. *The skew-normal and related families.* IMS Monograph series. Cambridge University Press, 2014.

[4] P. Billingsley. *Convergence of probability measures.* Wiley-Interscience, New York, 1999.

[5] S. Bochner. *Lectures on Fourier Integrals.* Princeton University Press, 1959.

[6] V. I. Bogachev. *Weak Convergence of Measures.* Amer. Math. Soc., 2018.

[7] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, Cambridge, 2004.

[8] J. Breckling and R. Chambers. M-quantiles. *Biometrika*, 75:761–771, 1988.

[9] B. M. Brown. Statistical uses of the spatial median. *J. Roy. Statist. Soc. Ser. B*, 45(1):25–30, 1983.

[10] M. A. Burr, E. Rafalin, and D. L. Souvaine. Simplicial depth: An improved definition, analysis, and efficiency for the finite sample case. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 72:195, 2006.

[11] L. Caffarelli and L. Silvestre. An extension problem related to the fractional Laplacian. *Communications in Partial Differential Equations*, 2007.

[12] H. Cardot, P. Cénac, and A. Godichon-Baggioni. Online estimation of the geometric median in Hilbert spaces: Nonasymptotic confidence balls. *Ann. Statist.*, 45:591–614, 2017.

[13] H. Cardot, P. Cénac, and P.-A. Zitt. Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli*, 19:18–43, 2013.

[14] G. Carlier, V. Chernozhukov, and A. Galichon. Vector quantile regression beyond the specified case. *J. Multivariate Anal.*, 161:96–102, 2017.

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

[16] B. Chakraborty. On multivariate quantile regression. *J. Statist. Plann. Inference*, 110:109–132, 2003.

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

[18] Z. Chen. Conditional L_p-quantiles and their application to testing of symmetry in non-parametric regression. *Statist. Probab. Lett.*, 29:107–115, 1996.

[19] A. Y. Cheng and M. Ouyang. On algorithms for simplicial depth. In *CCCG*, pages 53–56, 2001.

[20] Y. Cheng and J.G. De Gooijer. On the uth geometric conditional quantile. *J. Statist. Plann. Inference*, 137:1914–1930, 2007.

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

[22] J. Chowdhury and P. Chaudhuri. Nonparametric depth and quantile regression for functional data. *Bernoulli*, 25:395–423, 2019.

[23] J. Chowdhury and P. Chaudhuri. Nonparametric depth and quantile regression for functional data. *Bernoulli*, 25:395–423, 2019.

[24] X. Dai and S. Lopez-Pintado. Tukey's depth for object data. *J. Amer. Statist. Assoc.,*, page to appear, 2022.

[25] X. Dai and H.G. Müller. Principal component analysis for functional data on Riemannian manifolds and spheres. *Ann. Statist.*, 46:3334–3361, 2018.

[26] A. Daouia, S. Girard, and G. Stupfler. Estimation of tail risk based on extreme expectiles. *J. Roy. Statist. Soc. Ser. B*, 80:263–292, 2018.

[27] A. Daouia, S. Girard, and G. Stupfler. Extreme M-quantiles as risk measures: from L_1 to L_p optimization. *Bernoulli*, 25:264–309, 2019.

[28] E. del Barrio, A. Gonzàles-Sanz, and M. Hallin. A note on the regularity of optimal-transport-based center-outward distribution and quantile functions. *J. Multivariate Anal.*, 2020.

[29] D. L. Donoho and M. Gasko. Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.*, 20(4):1803–1827, 1992.

[30] L. Dümbgen. Limit theorems for the simplicial depth. *Statist. Probab. Lett.*, 14(2):119–128, 1992.

[31] L. Evans. *Partial Differential Equations*, volume 19. Amer. Math. Soc., 1998.

[32] T. Ferguson. *Mathematical Statistics : A decision theoretic approach.* Academic Press, 1967.

[33] X. Fernandez-Real and X. Ros-Oton. *Regularity theory for elliptic PDE.* Submitted, 2020.

[34] G. Francisci, A. Nieto-Reyes, and C. Agostinelli. Generalization of the simplicial depth: no vanishment outside the convex hull of the distribution support. *arXiv preprint arXiv:1909.02739*, 2019.

[35] M. R. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'I.H.P*, 10(4):215–310, 1948.

[36] E. Garcia-Portugues, D. Paindaveine, and T. Verdebout. On optimal tests for rotational symmetry against new classes of hyperspherical distributions. *J. Amer. Statist. Assoc.*, 115:1873–1887, 2020.

[37] L. Gardes, S. Girard, and G. Stupfler. Beyond tail median and conditional tail expectation: extreme risk estimation using tail L_p-optimization. *Scand. J. Statist.*, 47:922–949, 2020.

[38] P. Ghosal and B. Sen. Multivariate ranks and quantiles using optimal transport: Consistency, rates and nonparametric testing. *Annals of Statistics*, 50(2):1012–1037, 2022.

[39] A. K. Ghosh and P. Chaudhuri. On maximum depth and related classifiers. *Scand. J. Statist.*, 32:327–350, 2005.

[40] I. Gijbels and S. Nagy. On smoothness of Tukey depth contours. *Statistics*, 2016.

[41] S. Girard and G. Stupfler. Extreme geometric quantiles in a multivariate regular variation framework. *Extremes*, 18:629–663, 2015.

[42] S. Girard and G. Stupfler. Intriguing properties of extreme geometric quantiles. *REVSTAT*, 15:107–139, 2017.

[43] S. J. Haberman. Concavity and estimation. *The Annals of Statistics*, 1989.

[44] M. Hallin, E. del Barrio, J. C. Cuesta-Albertos, and C. Matrán. Distribution and quantile functions, ranks and signs in dimension d: a measure transportation approach. *Ann. Statist.*, 49:1139–1165, 2021.

[45] M. Hallin and G. Mordant. On the finite-sample performance of measure transportation-based multivariate rank tests. *Working Papers ECARES*, 2021.

[46] M. Hallin, D. Paindaveine, and M. Šiman. Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth (with discussion). *Ann. Statist.*, 38:635–669, 2010.

[47] K. Herrmann, M. Hofert, and M. Mailhot. Multivariate geometric expectiles. *Scand. Actuar. J.*, 2018:629–659, 2018.

[48] P. E. Jupp and A. Kume. Measures of goodness of fit obtained by almost-canonical transformations on Riemannian manifolds. *J. Multivariate Anal.*, 176:104579, 2020.

[49] R. Koenker and G. Bassett, Jr. Regression quantiles. *Econometrica*, 46:33–50, 1978.

[50] V. I. Koltchinski. M-estimation, convexity and quantiles. *Ann. Statist.*, 25:435–477, 1997.

[51] G. Koshevoy and K. Mosler. Zonoid trimming for multivariate distributions. *Ann. Statist.*, 25:1998–2017, 1997.

[52] C.-M. Kuan, J.-H. Yeh, and Y.-C. Hsu. Assessing value at risk with care, the conditional autoregressive expectile models. *J. Econometrics*, 150:261–270, 2009.

[53] N. S. Landkof. *Foundations of Modern Potential Theory.* Springer-Verlag, 1972.

[54] T. Leise and A. Cohen. Nonlinear oscillators at our fingertips. *Amer. Math. Monthly*, 114:14–28, 2007.

[55] C. Ley, C. Sabbah, and T. Verdebout. A new concept of quantiles for directional data and the angular Mahalanobis depth. *Electron. J. Stat.*, 8(1):795–816, 2014.

[56] C. Ley and T. Verdebout. *Modern Directional Statistics.* Chapman & Hall/CRC Interdisciplinary Statistics Series. CRC Press, Boca Raton, 2017.

[57] C. Ley and T. Verdebout. Skew-rotationally-symmetric distributions on the unit sphere and related efficient inferential procedures. *J. Multivariate Anal.*, 159:67–81, 2017.

[58] J. Li, J. Cuesta-Albertos, and R. Y. Liu. Dd-classifier: Nonparametric classification procedures based on dd-plots. *J. Amer. Statist. Assoc.*, 107:737–753, 2012.

[59] A. Lischke, G. Pang, M. Gulian, F. Song, C. Glusa, X. Zheng, Z. Mao, W. Cai, M. Meerschaert, M. Ainsworth, and G. Em Karniadakis. What is the fractional Laplacian? A comparative review with new results. *J. Comput. Phys.*, 2020.

[60] R. Y. Liu. On a notion of simplicial depth. *Proc. Nat. Acad. Sci. U.S.A.*, 85(6):1732–1734, 1988.

[61] R. Y. Liu. On a notion of data depth based on random simplices. *Ann. Statist.*, 18(1):405–414, 1990.

[62] R. Y. Liu. Data depth and multivariate rank tests. *L1-Statistics and Related Methods*, pages 279–294, 1992.

[63] R. Y. Liu and K. Singh. Ordering directional data: concepts of data depth on circles and spheres. *Ann. Statist.*, 20(3):1468–1484, 1992.

[64] R. Y. Liu and K. Singh. A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.*, 88(421):252–260, 1993.

[65] A. Magyar and David E. Tyler. The asymptotic efficiency of the spatial median for elliptically symmetric distributions. *Sankhyã*, 73:165–192, 2011.

[66] P. C. Mahalanobis. On the generalized distance in statistics. *Proceedings of the National Academy of Sciences of India*, 12:49–55, 1936.

[67] K. V. Mardia and P. E. Jupp. *Directional Statistics*. Wiley Series in Probability and Statistics. Wiley, Chichester, 2000.

[68] J.-C. Massé. Asymptotics for the Tukey median. *J. Multivariate Anal.*, 81:286–300, 2002.

[69] I.W. McKeague, S. López-Pintado, M. Hallin, and M. Šiman. Analyzing growth trajectories. *J. Developmental Origins of Health and Disease*, 2:322–329, 2011.

[70] K. Mosler. *Multivariate Dispersion, Central Regions, and Depth: The Lift Zonoid Approach*, volume 165. Springer Science & Business Media, 2012.

[71] J. Möttönen, K. Nordhausen, H. Oja, et al. Asymptotic theory of the spatial median. In *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková*, pages 182–193. Institute of Mathematical Statistics, 2010.

[72] J. Mottonen, H. Oja, and J. Tienari. On the efficiency of multivariate spatial sign and rank tests. *The Annals of Statistics*, 1997.

[73] N. D. Mukhopadhyaya and S. Chatterjee. High dimensional data analysis using multivariate generalized spatial quantiles. *J. Multivariate Anal.*, 102:768–780, 2011.

[74] S. Nagy. Monotonicity properties of spatial depth. *Statistics and Probability Letters*, 2017.

[75] S. Nagy. Halfspace depth does not characterize probability distributions. *Statistical Papers*, 2019.

[76] S. Nagy. The halfspace depth characterization problem. In Springer, editor, *Nonparametric statistics, 4th ISNPS*, 2020.

[77] W.K. Newey and J.L. Powell. Asymmetric least squares estimation and testing. *Econometrica*, 55:819–847, 1987.

[78] E. Nezza, G. Palatucci, and E. Valdinoci. Hitchhiker's guide to the fractional sobolev spaces. *Bulletin Des Sciences Mathematiques*, 2011.

[79] W. Niemiro. Asymptotics for M-estimators defined by convex minimization. *Ann. Statist.*, 20:1514–1533, 1992.

[80] H. Oja. Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.*, 1(6):327–332, 1983.

[81] H. Oja. *Multivariate Nonparametric Methods with R*. Springer, 2010.

[82] F. Olver, D. Lozier, R. Boisvert, and C. Clark, editors. *NIST Handbook of Mathematical Functions*. National Institute of Standards and Technology, and Cambridge University Press, 2010.

[83] D. Paindaveine and J. Virta. On the behavior of extreme d-dimensional spatial quantiles under minimal assumptions. In A. Daouia and A. Ruiz-Gazen, editors, *Advances in Contemporary Statistics and Econometrics*, pages 243–259, Cham., 2020. Springer.

[84] G. Pandolfo, Davy Paindaveine, and G. Porzio. Distance-based depths for directional data. *Canad. J. Statist.*, 46:593–609, 2018.

[85] O. Pokotylo, P. Mozharovskyi, and R. Dyckerhoff. Depth and depth-based classification with r-package ddalpha. *J. Statist. Softw.*, 91:1–46, 2019.

[86] R. Ranga Rao. Relations between weak and uniform convergence of measures with applications. *Indian Statistical Institue, Calcultta*, 1962.

[87] W. Rudin. *Functional analysis*. McGraw-Hill, Inc., second edition, 1991.

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

[89] R. Serfling. Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardization. *J. Nonparametr. Stat.*, 22:915–926, 2010.

[90] R. Serfling. Depth functions on general data spaces, I. Perspectives, with consideration of "density" and "local" depths. *Preliminary notes*, 2019.

[91] R. Serfling. Depth functions on general data spaces, II. Formulation and maximality, with consideration of the Tukey, projection, spatial, and "contour" depths. *Preliminary notes*, 2019.

[92] R. Serfling and U. Wijesuriya. Depth-based nonparametric description of functional data, with emphasis on use of spatial depth. *Comput. Statist. Data Anal.*, 105:24–45, 2017.

[93] R. Serfling and Y. Zuo. Discussion of "Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth", by M. Hallin, D. Paindaveine, and M. Šiman. *Ann. Statist.*, 38:676–684, 2010.

[94] L. Silvestre. Regularity of the obstacle problem for a fractional power of the Laplace operator. *Communications on Pure and Applied Mathematics*, 2007.

[95] P. Stinga. User's guide to the fractional Laplacian and the method of semigroups. *Fractional Differential Equations*, 2018.

[96] A. Struyf and P. Rousseeuw. Halfspace depth and regression depth characterize the empirical distribution. *Journal of Multivariate Analysis*, 69(1):135–153, 1999.

[97] J. Taylor. Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, 6:231–252, 2008.

[98] J. W. Tukey. Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974), Vol. 2*, pages 523–531. Canad. Math. Congress, Montreal, Que., 1975.

[99] D. Tyler. Statistical analysis for the angular central Gaussian distribution on the sphere. *Biometrika*, 74(3):579–589, 1987.

[100] A. Usseglio-Carleve. Estimation of conditional extreme risk measures from heavy-tailed elliptical random vectors. *Electron. J. Stat.*, 12:4057–4093, 2018.

[101] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Univ. Press, Cambridge, 1998.

[102] Y. Vardi and C.-H. Zhang. The multivariate L_1-median and associated data depth. *Proc. Natl. Acad. Sci. USA*, 97(4):1423–1426, 2000.

[103] Q. Wang, J. Zhu, W. Pan, J. Zhu, and H. Zhang. Nonparametric statistical inference via metric distribution function in metric spaces. *arXiv:2107.07317*, 2021.

[104] Y. Wei. An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts. *J. Amer. Statist. Assoc.*, 103:397–409, 2008.

[105] L. Yang. Some properties of Fréchet medians in Riemannian manifolds. *arXiv preprint arXiv:1110.3899v2*, 2011.

[106] W. Zhou and R. Serfling. Multivariate spatial U-quantiles: a Bahadur-Kiefer representation, a Theil–Sen estimator for multiple regression, and a robust dispersion estimator. *J. Statist. Plann. Inference*, 138:1660–1678, 2008.

[107] Y. Zuo. Projection-based depth functions and associated medians. *Ann. Statist.*, 31(5):1460–1490, 2003.

[108] Y. Zuo and R. Serfling. General notions of statistical depth function. *Ann. Statist.*, 28(2):461–482, 2000.