Multivariate quantiles and regularized optimal transport

Gauthier Thurin

École doctorale Mathématiques et informatique — Université de Bordeaux

Institut de Mathématiques de Bordeaux

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR
DE L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE MATHÉMATIQUES ET INFORMATIQUE

MATHÉMATIQUES APPLIQUÉES ET CALCUL SCIENTIFIQUE

Par **Gauthier THURIN**

Quantiles multivariés et transport optimal régularisé

Sous la direction de : **Jérémie BIGOT**
Co-directeur : **Bernard BERCU**

Soutenue le 15/11/2024

Membres du jury :

M. Eustasio DEL BARRIO	Professor	University of Valladolid	Rapporteur
M. Davy PAINDAVEINE	Professor	Université Libre de Bruxelles	Rapporteur
Mme. Elena DI BERNARDINO	Professor	Université Côte d'Azur	Examinatrice
M. Francois-Xavier VIALARD	Professor	Université Gustave Eiffel	Président
M. Jérémie BIGOT	Professor	Université de Bordeaux	Directeur
M. Bernard BERCU	Professor	Université de Bordeaux	Co-directeur

Membres invités :

Mme. Claire BOYER	Professor	Université Paris-Saclay	Invitée

Multivariate quantiles and regularized optimal transport

Gauthier Thurin

Advisors : Jérémie Bigot and Bernard Bercu

Quantiles multivariés et transport optimal régularisé

Résumé : L'objet d'intérêt principal de cette thèse est la fonction quantile de Monge-Kantorovich. On s'intéresse d'abord à la question cruciale de son estimation, qui revient à résoudre un problème de transport optimal. En particulier, on tente de tirer profit de la connaissance *a priori* de la loi de référence, une information additionnelle par rapport aux algorithmes usuels, qui nous permet de paramétrer les potentiels de transport par leur série de Fourier. Ce faisant, la régularisation entropique du transport optimal permet deux avantages : la construction d'un algorithme efficace et convergent pour résoudre la version semi-duale de notre problème, et l'obtention d'une fonction quantile empirique lisse et monotone. Ces considérations sont ensuite étendues à l'étude de données sphériques, en remplaçant les séries de Fourier par des harmoniques sphériques, et en généralisant la carte entropique à ce cadre non-euclidien. Le second objectif de cette thèse est de définir de nouvelles notions de superquantiles et d'expected shortfalls multivariés, pour compléter l'information fournie par les quantiles. Ces fonctions caractérisent la loi d'un vecteur aléatoire, ainsi que la convergence en loi, sous certaines hypothèses, et trouvent des applications directes en analyse de risque multivarié, pour étendre les mesures de risque classiques de Value-at-Risk et Conditional-Value-at-Risk.
Mots-clés : Quantiles régularisés, Transport entropique, Données sphériques, Superquantiles

---

Multivariate quantiles and regularized optimal transport

Abstract: This thesis is concerned with the study of the Monge-Kantorovich quantile function. We first address the crucial question of its estimation, which amounts to solve an optimal transport problem. In particular, we try to take advantage of the knowledge of the reference distribution, that represents additional information compared with the usual algorithms, and which allows us to parameterize the transport potentials by their Fourier series. Doing so, entropic regularization provides two advantages: to build an efficient and convergent algorithm for solving the semi-dual version of our problem, and to obtain a smooth and monotonic empirical quantile function. These considerations are then extended to the study of spherical data, by replacing the Fourier series with spherical harmonics, and by generalizing the entropic map to this non-Euclidean setting. The second main purpose of this thesis is to define new notions of multivariate superquantiles and expected shortfalls, to complement the information provided by the quantiles. These functions characterize the law of a random vector, as well as convergence in distribution under certain assumptions, and have direct applications in multivariate risk analysis, to extend the traditional risk measures of Value-at-Risk and Conditional-Value-at-Risk.
Keywords: Regularized quantiles, Entropic optimal transport, Spherical data, Superquantiles

---

Institut de mathématiques de Bordeaux
UMR 5251 Université de Bordeaux, 33000 Bordeaux, France.

Remerciements

En premier lieu, j'adresse ma gratitude à mes directeurs de thèse, Jérémie et Bernard. Merci de m'avoir fait confiance et pour vos conseils de ces trois dernières années, qui m'ont permis de mener à bien ce fastidieux projet. En fait, je vous suis aussi reconnaissant pour votre enseignement de master, ce qui remonte finalement aux cinq dernières années !

Merci aux membres du jury d'avoir accepté d'assister à ma soutenance, et aux rapporteurs pour leur lecture attentive de mon manuscript. Vos retours comptent énormément pour moi.

Puisque j'en suis à remercier des professeurs, je dois mentionner M.Normand, le meilleur enseignant que j'ai eu. Et tant qu'à remercier des adultes qui me connaissent depuis longtemps, merci à ma marraine, Manue, de venir assister à ma soutenance après tout ce qu'on a pu partager. Merci aussi à Monic et Roger-Pierre, pour leur gentillesse et leur hospitalité, à Saint-Sym ou au Jouquet.

À toutes les personnes que j'ai pu croiser à l'IMB, merci. Merci aux membres des missions environnement et parité, pour fournir un espace ouvert à la discussion sur la transition de la recherche. Surtout, merci à Christèle, Emmanuelle, Rémi, Nicolas, pour votre investissement pour la communauté et pour nos échanges bienveillants. Merci à tous ceux de l'équipe IOP pour votre sympathie. Merci aux plus jeunes membres de l'IMB, pour avoir égayé cet environnement de travail. Sans être exhaustif, merci à Pierre-Jean, Martin, Samuel, Jean, Anne-Edgar, Charu, Théo, Simon, Antoine, Magalie, Florian, Julien, Virginie, Alexis, Kylian, Agathe, Marco, Simone, Bianca, Julien, Marien, He, Ali, Eloan, Omar. Je garderai un très bon souvenir de vous toutes et tous, et je suis heureux d'avoir pu croiser votre chemin d'une façon ou d'une autre. J'ai eu la chance de partager ma direction de thèse avec plusieurs collègues formidables. Merci Lucile, sans doute la première personne à m'avoir accueilli, de surcroît dans ton bureau ! Merci à Paul, un moteur de la bonne ambiance de l'équipe ! Merci à Yiye et Issa pour les JDS auxquelles nous avons participé ensemble. Merci à Erell et Émeric, les prochains à qui je souhaite de belles années de thèse !

Aux amis de master, Mehdy, Benjamin, Pol, merci pour tout le temps passé ensemble, pour les projets, les week-ends, les soirées et le reste. Assurément, cette thèse n'existerait pas sans vous. Pol, nos discussions du lundi ont structuré tout ce que j'ai pu rechercher, en vain souvent ! Merci pour les conseils, le soutien, les pantalons, les blagues, les bières et les parties de poker. Pardon de ne pas avoir de personnalité et de copier le mullet. Notre amitié est plus que précieuse pour moi, u know it. Merci à Mariette de m'accueillir au RU de saint genès et pour les WE sur le bassin, le tour ostréicole et l'emballage des huîtres sont bien sûr les meilleurs souvenirs des 5 dernières années ! Merci à Harmonie et Benoît pour être d'aussi géniales personnes, et d'aussi bons maîtres du jeu pour tous les jeux de société qui peuvent exister (surtout celui que je ne vais pas nommer ici). Merci aux cousines de Camille : Amande, Annabel, Sacha (désolé je te range dans cette catégorie !) de

venir à ma soutenance, et pour tout ce qu'on peut partager. Je suis fier de tous vous compter parmi mes amis, j'ai beaucoup de chance de vous connaître. À celles et ceux de La Sauque, je vous suis reconnaissant d'avoir pu grandir à vos côtés, et d'avoir pu rester si proches. Merci à Maxime, Baudoin, Théo, Martin, Damien, pour tout ce qu'on a partagé, l'internat, des vacances, des anniversaires. Ma porte vous sera toujours ouverte, vous le savez. J'en profite pour te féliciter Max, tu seras un super papa ! Merci Gauthier, pour supporter mon retard et de m'accepter, quelques fois au moins, dans ton équipe au beer-pong. Je crois que je ne m'améliorerai jamais ... Je chéris tous les souvenirs qu'on a construit ensemble. Continue de nous faire rire, je suis le plus grand fan de tes one man show. Compte sur moi pour toutes les étapes de ta vie. Bravo encore pour ton futur mariage, 2025 est ton année, tu sais que je suis fier de toi ! On essaie de nager jeudi (lol). Merci Solène, avec qui je partage mon signe astrologique, et à ce titre moult aspects de ma personnalité ! La compétition avec Baud pour le titre de bestie n'est jamais finie ! Reste aussi solaire que tu l'es, ça éclaire tout ton entourage ! Merci Marine, pour les trajets jusque Pau, les voyages un peu partout, les cafés, les escape game, et ton investissement quand il s'agit d'anniversaires ! Surtout, merci pour ton soutien, d'être toujours de bon conseil. J'espère être un aussi bon ami pour toi que tu l'es pour moi, mais la tâche est ardue ! Vous faites partie de ma famille, ce qui fait une merveilleuse transition pour la fin de ces remerciements.

Merci à ma famille, aux THURIN et aux VRIGNY, à ceux qui sont toujours là pour moi, et à ceux qui ne le peuvent plus. Je ne suis que la somme d'un peu de vous tous, ce qui est sans doute ma plus grande fierté. Darwin a tort sur toute la ligne, chez moi l'homme descend des lionnes et des lions. Merci papa et maman pour m'avoir aimé même lorsque je ne le méritais pas vraiment, pour m'avoir montré ce qu'est le courage, le bien, et pour m'avoir donné mes deux petits frères ! Antoine, tu as toujours été le meilleur de nous deux. Bien que je sois l'aîné, tu m'inspires tous les jours. Fais attention avec les outils de travailleur, histoire de ne pas finir estropié ! Robin, ça compte pour toi aussi. Merci de préserver la passion pour les Super Picsou Géant. Garde bien ta force de caractère, c'est une qualité précieuse. Je suis fier de vous deux, vous êtes beaux comme des fleurs *clin d'oeil*.

Camille, merci pour tout. Il me faudrait une centaine de pages supplémentaires pour tout te dire, alors mieux vaut être incomplet. J'ai aimé chaque seconde des dix dernières années, et j'espère pouvoir en dire autant des dix prochaines. Si tu n'existais pas il faudrait t'inventer, pour moi, pour tes proches et pour l'humanité.

Sommaire

Résumé substanciel en français

Quantiles multivariés

La fonction quantile est un outil fondamental en statistiques, idéal pour décrire et caractériser une distribution de probabilité, utile depuis l'analyse descriptive jusqu'à des outils de régression, [100], ou des méthodes d'analyse robuste, [93]. La question de la définition des fonctions quantile et de répartition pourrait sembler triviale au lecteur familier avec les concepts univariés. Pourtant, ce qui est connu en dimension $d = 1$ ne se généralise pas d'une unique manière quand $d > 1$, en l'absence d'une relation d'ordre canonique. Pour ce qui est de la droite réelle $\mathbb{R}$, une variable aléatoire X de fonction de répartition $F(x) = \mathbb{P}(X \leq x)$ admet pour fonction quantile

$$Q(\alpha) = \inf\{x : F(x) \geq \alpha\},$$

pour $\alpha \in [0, 1]$. En d'autres termes, une observation x égale le quantile d'ordre α si la proportion de données inférieures à x égale α, ce qui demande, en pratique, d'ordonner les observations de gauche à droite. Les quantiles fournissent ainsi une information de centralité, illustrée par la boîte à moustaches, comme en Figure 1, qui donne une région centrale contenant 50% des observations.

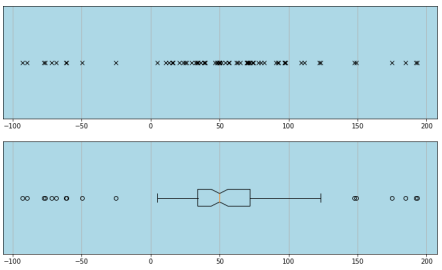


FIGURE 1 – Jeu de données réel et boîte à moustaches associée.

Toutefois, cette définition pour des données dans $\mathbb{R}$ ne s'étend pas directement à des distributions multivariées. Pour $d > 1$, l'espace $\mathbb{R}^d$ n'est pas muni d'une relation d'ordre canonique : il existe plusieurs manières satisfaisantes d'ordonner des vecteurs aléatoires, chacune avec ses atouts et limites. Pour une vue d'ensemble de différentes approches, des références instructives sont [87, 156]. Une fonction de

répartition d'un vecteur aléatoire $(X_1, \cdots, X_d)$ peut être définie par

$$F(x_1, \cdots, x_d) = \mathbb{P}(X_1 \leq x_1, \cdots, X_d \leq x_d).$$

Néanmoins, puisque $F : \mathbb{R}^d \to [0,1]$ n'est pas inversible, ceci n'est pas associé à une fonction quantile multivariée. À l'inverse, le concept de profondeur statistique [169] préfère un ordre du centre vers l'extérieur, jugé plus naturel pour un nuage de points en l'absence de contexte supplémentaire. La Figure 2 illustre la profondeur de Tukey [169] sur un échantillon Gaussien bivarié.

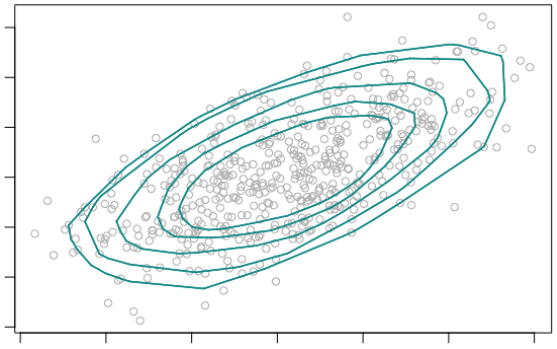


FIGURE 2 – Contours de même profondeur de Tukey, [169], dans $\mathbb{R}^2$.

Une liste de critères désirables pour une fonction quantile multivariée $\mathbf{Q}$ est proposée dans [156] sous forme de plusieurs questions, que nous reprenons ici.

(Q1) Existe-t-il une interprétation probabiliste analogue au cadre réel ?

(Q2) Est-il possible de formuler des statistiques descriptives de centralité, comme la médiane ou les moyennes tronquées ?

(Q3) Est-il possible de formuler des statistiques descriptives de dispersion ?

(Q4) Est-ce que $\mathbf{Q}$ satisfait des propriétés d'invariance convenables ?

(Q5) Peut-on construire des tests statistiques efficaces pour les hypothèses classiques ?

Les quantiles de Monge-Kantorovich, construits à partir du transport de mesures de probabillité [33, 87] répondent par l'affirmative à la plupart de ces pré-requis.

Tout d'abord, en dimension $d = 1$, il est bien connu que la fonction quantile de ν est solution du problème de transport optimal quadratique entre la loi uniforme sur $[0,1]$ et ν, see *e.g.* [138][Remark 2.30]. La définition des quantiles de Monge-Kantorovich en dimension $d \geq 2$ quelconque étend tout simplement cette propriété. Il s'agit, en fonction d'une loi de référence μ, de considérer la fonction $\mathbf{Q} : \mathbb{R}^d \to \mathbb{R}^d$ qui *transporte* μ vers ν de manière adéquate. Pour rappel, on dit d'une fonction $T : \mathbb{R}^d \to \mathbb{R}^d$ qu'elle transporte μ vers ν, si, pour tout ensemble mesurable B de $\mathbb{R}^d$, $\nu(B) = \mu\big(T^{-1}(B)\big)$, ce qu'on note $\nu = T_{\#}\mu$. En termes de vecteurs aléatoires, il suffit d'échantillonner U selon μ pour simuler $T(U)$ selon $\nu = T_{\#}\mu$.

La première brique dans la construction des quantiles de Monge-Kantorovich est le choix d'une distribution de référence μ idéale, pour laquelle les points sont ordonnés de manière naturelle. Par exemple, on peut considérer la loi uniforme sur la boule ou le cube unité, ou une loi normale centrée réduite. Des régions quantiles indexées par leur μ-probabilité peuvent alors être facilement définis, ce qui permet de répondre favorablement aux pré-requis énoncés ci-dessus.

Pour définir des quantiles pour une loi multivariée quelconque ν, il s'agit ensuite de *transporter* l'ordre naturel de μ vers ν par une fonction $\mathbf{Q} : \mathbb{R}^d \to \mathbb{R}^d$, qui joue ainsi le rôle d'une fonction quantile multivariée. Parmi toutes les manières de transporter μ vers ν, reste à en choisir une unique, en imposant que $\mathbf{Q}$ soit le gradient d'une fonction convexe. D'une part, être gradient d'une fonction convexe est une généralisation de la monotonie d'une fonction quantile univariée. D'autre part, la fonction quantile de Monge-Kantorovich est ainsi bien définie de manière unique pour n'importe quelle distribution ν, par un théorème de McCann, [121].

La Figure 3 illustre l'ordre induit par ces concepts, en prenant pour référence la loi sphérique uniforme $\mu = U_d$ définie par le produit $R\Phi$ entre les variables aléatoires indépendantes R et Φ, tirées uniformément sur $[0, 1]$ et sur la sphère unité, respectivement, [33, 87]. Pour une telle loi, les régions quantiles d'ordre $\alpha \in [0, 1]$ sont les boules de rayon α. En Figure 3, les rayons en rouge et les cercles en bleu sont transportés depuis la boule unité vers un jeu de données quelconque par la fonction quantile de Monge-Kantorovich. Puisque la définition de [33, 87] requiert que $Q_{\#}\mu = \nu$, l'ordre centre-extérieur s'adapte naturellement à la géométrie du support sous-jacent, même s'il n'est pas convexe.

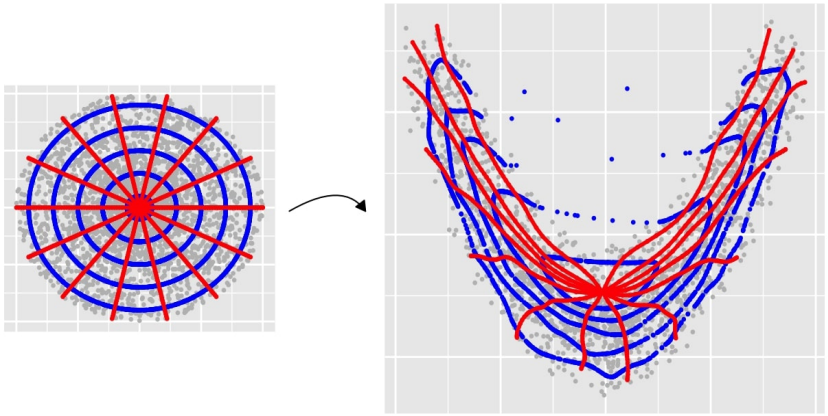


FIGURE 3 – (Gauche) quantiles d'une loi de référence et (droite) quantiles de Monge-Kantorovich d'une loi discrète ν obtenus par $\mathbf{Q}_{\#}\mu = \nu$.

La première question $(Q1)$ est relative aux régions quantiles de ν, qui doivent être indexées par leur ν-probabilité. En d'autres termes, une région quantile d'ordre 50% doit contenir la moitié des observations, comme pour la boîte à moustaches, en Figure 1. Les régions quantiles $\mathbb{C}_\alpha = \mathbf{Q}(\mathbb{B}(0, \alpha))$ vérifient bien ce pré-requis, puisque

$$\nu\big(\mathbf{Q}(\mathbb{B}(0, \alpha))\big) = U_d\big(\mathbb{B}(0, \alpha)\big) = \alpha.$$

Concernant $(Q2)$ et $(Q3)$, des statistiques descriptives peuvent être définies de manière intuitive. Une médiane multivariée est donnée par l'intersection de toutes les régions quantiles [87] et des moyennes tronquées se calculent aisément via $\mathbb{E}[X|X \in \mathbb{C}_\alpha] = \mathbb{E}[Q(U)|U \in \mathbb{B}(0,\alpha)]$, ce qui apparaît naturellement dans [60, 85], avec l'objectif de courbes de Lorenz multivariées. La même remarque est vraie pour des mesures de dispersion, notamment avec l'estimation des volumes des régions quantiles proposée dans [15]. Au regard de $(Q4)$, les seules invariances connues de $\mathbf{Q}$ concernent la translation et multiplication par une constante positive, [74][Lemma A.7], ainsi que les transformations orthogonales, [74][Lemma A.8], lorsque la loi de référence est sphérique. Finalement, depuis l'introduction des quantiles de Monge-Kantorovich, de nombreux tests statistiques ont été proposés, concernant l'adéquation, [50], l'indépendance, [50, 132, 159-161], la symmétrie [92], ou encore des modèles de regression, [82, 83, 86] ainsi que des bornes d'efficacité de Pitman [49]. Une propriété en particulier plaide en faveur de l'approche par transport optimal vis-à-vis de $(Q5)$: la distribution des rangs de Monge-Kantorovich est indépendante de celle des observations, [33, 50, 87].

Travaux de thèse

À la lumière de ses bonnes propriétés et de l'ensemble des travaux récents déjà évoqués [81], la fonction quantile de Monge-Kantorovich paraît prometteuse pour étendre dans $\mathbb{R}^d$ des méthodes d'inférence classiques sur la droite réelle. Du point de vue pratique, l'estimation de $\mathbf{Q}$ correspond à la résolution du problème de transport optimal de Monge entre la loi de référence μ et une mesure empirique $\widehat{\nu}_n$ associées à des données $Y_1, \cdots, Y_n$,

$$\min_{T:T_{\#}\mu=\widehat{\nu}_n} \mathbb{E}_{X\sim\mu}\left[\|X - T(X)\|^2\right].$$

Ce problème général a reçu beaucoup d'attention au cours des dernières années, et il existe différentes approches numériques pour le résoudre. Avant de pouvoir appliquer les quantiles de Monge-Kantorovich pour de l'inférence multivariée, la première question est donc le choix d'un estimateur.

Dans cette perspective, la première idée avancée dans cette thèse, dès le Chapitre 2, est de tirer profit de la connaissance *a priori* de la loi de référence μ. En effet, les approches existantes pour résoudre un problème de transport optimal considèrent des lois μ et ν quelconques, mais le contexte spécifiques des quantiles multivariés nous donne l'information supplémentaire de la loi de référence μ. En considérant la loi uniforme sur l'hypercube $\mu = \text{Unif}([0,1]^d)$, le potentiel $u : [0,1]^d \to \mathbb{R}$ tel que $Q(\cdot) = \nabla\left(\frac{\|\cdot\|^2}{2} - u(\cdot)\right)$ peut être assimilé à une fonction périodique et paramétré par ses coefficients de Fourier. De cette façon, il devient possible de formuler un algorithme du gradient stochastique pour résoudre le problème de transport optimal *continu* entre $\mu = \text{Unif}([0,1]^d)$ et ν. Ici, le terme *continu* signifie qu'il n'est pas requis que la loi ν soit discrète, ce qui est communément dé-

signé par les cas *discret* ou *semi-discret*. Ainsi, à la limite de ses itérations, notre algorithme converge vers la solution du transport entre μ et une loi ν arbitraire, discrète ou continue. Afin d'obtenir des garanties de convergence, le problème de transport est régularisé, [138], ce qui convexifie notre problème d'optimisation et nous permet d'appliquer des schémas de preuve issus de l'optimisation stochastique. Par ailleurs, pour de nombreuses applications, il convient de régulariser la fonction quantile empirique puisque cette dernière est restreinte à prendre ses valeurs dans $\{Y_1, \cdots, Y_n\}$. Précisément, notre approche qui repose sur la carte entropique, [140, 153], fournit naturellement un estimateur lisse et monotone de $\mathbf{Q}$, au sens d'être le gradient d'une fonction convexe.

Pour aller plus loin, la définition des quantiles de Monge-Kantorovich a récemment été étendue pour des données non-euclidiennes, [84], ce qui constitue le point de départ du Chapitre 3. Puisque le besoin d'estimateurs efficaces et lisses est le même que dans $\mathbb{R}^d$, nous avons généralisé l'estimation de quantiles régularisés au contexte de données sphériques. Sur la sphère $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d, \|x\| = 1\}$, les potentiels peuvent être paramétrés par des harmoniques sphériques, analogues aux séries de Fourier, pour considérer, encore une fois, un algorithme de résolution du problème continu entre une loi uniforme et une loi ν arbitraire. Cette estimation régularisée permet des applications en analyse descriptive, pour comparer des distributions de probabilité selon leur dispersion. En outre, une notion de profondeur statistique permet de catégoriser une observation selon sa centralité vis-à-vis de plusieurs nuages de points, [113]. Des applications pour de la classification sont donc également considérées dans le Chapitre 3, par le biais d'une version directionnelle de la profondeur statistique de Monge-Kantorovich, que l'on définit et dont on étudie les propriétés au regard des traditionnels axiomes de Liu-Zuo-Serfling. Sur les tâches d'inférence considérées, les quantiles régularisés de Monge-Kantorovich présentent des avantages qualitatifs et quantitatifs par rapport aux notions de quantiles existantes sur la sphère $\mathbb{S}^2$, ce qui motive d'autant plus leur étude.

Un dernier axe de recherche est de définir des versions multivariées de superquantiles et d'expected shortfalls à partir des quantiles de Monge-Kantorovich, un problème ouvert selon [87]. En dimension $d = 1$, ces fonctions sont définies à partir de la fonction quantile Q, par

$$S(\alpha) = \frac{1}{1-\alpha}\int_\alpha^1 Q(t)dt \qquad \text{et} \qquad E(\alpha) = \frac{1}{\alpha}\int_0^\alpha Q(t)dt.$$

Les définitions du Chapitre 4 sont aussi des versions intégrales de la fonction quantile de Monge-Kantorovich $\mathbf{Q}$, définies de manière similaire par

$$S(u) = \frac{1}{1-\|u\|}\int_{\|u\|}^1 \mathbf{Q}\Big(t\frac{u}{\|u\|}\Big)dt \qquad \text{et} \qquad E(u) = \frac{1}{\|u\|}\int_0^{\|u\|} \mathbf{Q}\Big(t\frac{u}{\|u\|}\Big)dt.$$

Ce faisant, elles préservent l'interprétation univariée d'une observation typique au delà, ou en dessous, d'un quantile. Nous montrerons que ces fonctions caractérisent

les vecteurs aléatoires et leur convergence en loi sous certaines hypothèses, ce qui souligne leur importance. Sur la droite réelle, leur principale application concerne l'analyse de risque, bien que certains auteurs plaident pour un usage plus général, [105, 145]. En effet, en dimension $d = 1$, les mesures de risque les plus communément utilisées sont certainement la Value-at-Risk et la Conditional-Value-at-Risk, définies simplement comme un quantile ou un superquantile. Il convient alors d'explorer l'extension multivariée de ces notions par le biais de nos nouvelles définitions, ce qui est fait par la suite, avec des mesures de risques qui préservent l'intuition univariée, et dont le bon comportement est vérifié en pratique sur des données réelles et simulées.

Contributions

Cette thèse a donné lieu à la rédaction des articles suivants.

Bernard, Bercu, Jérémie Bigot, and Gauthier Thurin.
Stochastic optimal transport in Banach Spaces for regularized estimation of multivariate quantiles. *arXiv :2302.00982* (2023)

Bernard, Bercu, Jérémie Bigot, and Gauthier Thurin.
Monge-Kantorovich superquantiles and expected shortfalls with applications to multivariate risk measurements. *arXiv :2307.01584*. To appear in Electronic Journal of Statistics, vol 18 (2024).

Bernard, Bercu, Jérémie Bigot, and Gauthier Thurin.
Regularized estimation of Monge-Kantorovich quantiles for spherical data. *arXiv :2407.02085* (2024).

1 Introduction

With the advent of machine learning, [89], the statistical world has been deeply transformed, giving rise to a plethora of techniques to *picture data*, in the terminology of Tukey, [169]. Somehow surprisingly, there is still no consensus for a concept of quantile and distribution function beyond the real line $\mathbb{R}$, even if these preoccupations lie at the foundations of univariate statistics, [43, 158]. A recent proposal builds on optimal transportation of measures, for which this first chapter provides an introduction. Doing so, we motivate the remaining of the thesis, from the issue of finding regularized estimators to the extension of the latter to a directional setting, together with new concepts derived from MK quantiles, to describe peripheral or central areas of point clouds.

Contents

1.1 Background

Quantiles are a fundamental tool in statistics, from the very basis of descriptive analysis to more developed inference, such as quantile regression [100] or robust analysis [93]. The definition of distribution and quantile function might not seem an issue for the reader familiar with univariate concepts. Nevertheless, the latter do not readily extend to the multivariate setting, where a canonical order relation is lacking, [33, 87, 156]. Despite profuse efforts in this direction, the quest for a consensus about multivariate quantile and distribution functions is still ongoing. The present thesis is particularly interested in the recent proposal that builds upon the theory of optimal transportation (OT) [33, 87] and that has been shown to benefit from most of sought-after properties. In some sense, the latter concepts are in line with the modern formalism of probability and statistical theory [101] since they are rooted in the basic tools of measure theory. While it is well-known that the distinction between discrete and continuous distributions is avoided with the formalism of measures [42] the main objective of OT-based concepts is to overcome another arbitrary distinction between univariate and multivariate distributions.

1.1.1 The old problem of ordering multivariate data

The issue of defining proper concepts of multivariate distribution and quantile functions is of major importance, as evidenced by the fundamental position of univariate counterparts [43, 158]. The main objective is to *order* points within a data set in a meaningful way, with regions indexed by their probability content $\alpha \in [0,1]$. Indeed, for a univariate distribution ν with cumulative distribution function $F(x) = \mathbb{P}_\nu(X \leq x)$, recall that the quantile $Q(\alpha)$ of level $\alpha \in [0,1]$ is given by

$$Q(\alpha) = \inf\{x : F(x) \geq \alpha\}. \tag{1.1}$$

In other words, an observation x equals the quantile of order α if the proportion of data at the left of x is equal to α. Thus, in practice, quantiles are given by ranking the observations from left to right. One may think of the univariate boxplot, illustrated in Figure 1.1, that gives a central quantile region containing 50% of the observations.

Apart for univariate data, these concepts are well understood for elliptical distributions [33] where quantile contours shall coincide with density contours. Indeed, with elliptic symmetry, it is quite natural to *order* points from the center to the outward with respect to the covariance structure. The outlyingness of a point shall thus be consistent with Mahalanobis distance, see *e.g.* the discussion in [80] that relates the latter with Monge-Kantorovich quantiles. However, data of interest in practice is far from being limited to univariate or elliptical samples. The traditional

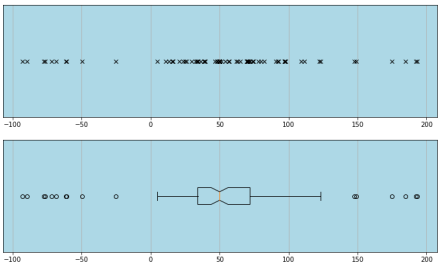


FIGURE 1.1 – Samples on the real line and their associated boxplot.

multivariate distribution function for a random vector $(X_1, \cdots, X_d)$ is given by

$$F(x_1, \cdots, x_d) = \mathbb{P}(X_1 \leq x_1, \cdots, X_d \leq x_d), \tag{1.2}$$

but this is not readily invertible, and thus it does not come with a suitable quantile concept. Such cumulative distribution function leverages an adequate feature of a distribution function, that is $F(x) = \mathbb{P}(X \in B_x)$, for some set B_x. However, the choice of $B_x =]-\infty, x_1] \times \cdots \times]-\infty, x_d]$ is somehow arbitrary. Of course, it is inherited by the univariate case where $\leq$ orders the whole space canonically and where the choice $]-\infty, x]$ is not restrictive. Yet, in $\mathbb{R}^d$, there is no reason for this left-to-right ordering to be prevalent in view of intuitive notions of quantile regions. These considerations gave rise to the concept of *statistical depth*, that aims to overcome the lack of a canonical ordering in $\mathbb{R}^d$ with a data-dependent ordering. To some extent, statistical depth and multivariate quantiles are the two sides of the same coin. The first of the kind, the halfspace depth, was introduced by Tukey in [169], with descriptive analysis and geometric intuition in mind, and an application to multivariate boxplots, as illustrated in Figure 1.2. This halfspace depth characterizes the underlying distribution when it is discrete [45] or under elliptical symmetry [103] but not in general [128]. Under elliptical symmetry, contours of same halfspace depth also share the same value for the density functions [177] which yields a highly satisfactory ordering of points from such a distribution.

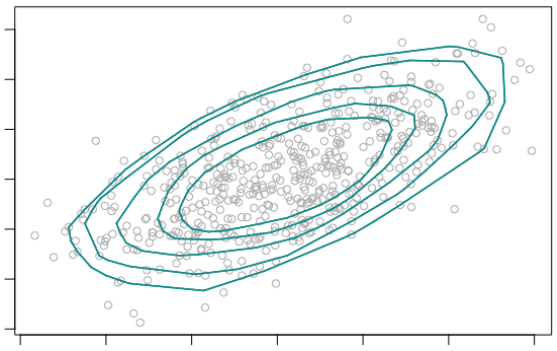


FIGURE 1.2 – Contours of same halfspace depth on a Gaussian sample in $\mathbb{R}^2$.

The issue of defining adequate statistical depths has gained too much attention for us to provide an exhaustive survey. Rather, we refer to [10, 112, 127, 157], see also [75, 94] for applications in classification or clustering. There, quoting [33],

> "*if the ultimate purpose of statistical depth is to provide, for each distribution P, a P-related ordering of $\mathbb{R}^d$ producing adequate concepts of quantile and distribution functions, ranks and signs, the relevance of a given depth function should be evaluated in terms of the relevance of the resulting ordering, and the quantiles, ranks and signs it produces.*"

To that extent, we shall see that the concepts based on optimal transport show promising features, with meaningful concepts of distribution and quantile functions, ranks and signs.

1.1.2 Optimal transport in a nutshell

This part is dedicated to provide a short introduction to optimal transport adapted to our needs. For further precisions, we refer the reader to the classical books [172, 173]. An important notion for our purposes is the one of push-forward measure, that allows the very definition of a probability $\mathbb{P}$ as a measure on some unspecified measure space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}(\Omega) = 1$, [101]. Informally, a push-forward measure $\nu = T_{\#}\mu$ is defined by *transporting* a measure μ using a measurable function T. For any measurable set B, this definition writes

$$T_{\#}\mu(B) = \mu\big(T^{-1}(B)\big). \tag{1.3}$$

This yields a change-of-variables formula. For any integrable function g,

$$\int_B g d(T_{\#}\mu) = \int_{T^{-1}(B)} g \circ T d\mu. \tag{1.4}$$

In terms of random variables, a sample from $\nu = T_{\#}\mu$ can be generated by $T(U)$ if U is a random variable drawn from μ. Throughout the following, we will also refer to T as a push-forward map, or mapping, from μ to ν.

Among all push-forward mappings from one distribution μ towards another ν, optimal transportation seeks for the one that minimizes the average cost of its displacements. Denote by $\mathcal{X}, \mathcal{Y}$ the respective supports of the measures μ and ν. Consider a cost function c on $\mathcal{X} \times \mathcal{Y}$. Then, the Monge problem writes as follows.

Monge optimal transport problem :

$$\textit{Minimize } I[T] = \int_{\mathcal{X}} c\left(x, T(x)\right) d\mu(x) \textit{ on the set of } T \textit{ such that } \nu = T\#\mu \quad (P_0)$$

A solution of the above problem is often called a Monge map. In addition, the Kantorovich relaxation allows that one amount of mass x can be sent towards several locations with a given probability. This is made possible by considering

joint probabilities instead of mappings $T : \mathbb{R}^d \to \mathbb{R}^d$. Let $\Pi(\mu, \nu)$ be the set of all joint probability measures on $\mathcal{X} \times \mathcal{Y}$ whose marginals are μ, ν. The Kantorovich relaxation of OT writes as follows.

Kantorovich optimal transport problem :

$$\textit{Minimize } I[\pi] = \int_{\times \mathcal{Y}} c(x, y) d\pi(x, y) \textit{ subject to } \pi \in \Pi(\mu, \nu). \tag{P_0'}$$

When the underlying space is endowed with a distance d, optimal transport comes with the Wasserstein distance between probability measures, [173][Chapter 6]. Crucially, the problem (P_0') admits a dual formulation. Define the set Φ_c of measurable functions $(\varphi, \phi) \in L^1(d\mu) \times L^1(d\nu)$ such that, almost everywhere, $\varphi(x) + \phi(y) \leq c(x, y)$. Then, ($P_0'$) is equivalent to the following [173][Theorem 5.10],

Dual Kantorovich problem :

$$\sup_{(\varphi,\phi)\in\Phi_c} \int \varphi(x) d\mu(x) + \int \phi(y) d\nu(y). \tag{D_0}$$

When the optimum of (D_0) is reached, one necessarily has that φ and ϕ are c-transform one to another, [173][Chapter 5], meaning that

$$\varphi(x) = \inf_{y\in\mathcal{Y}} \{c(x, y) - \phi(y)\} \qquad \text{and} \qquad \phi(y) = \inf_{x\in\mathcal{X}} \{c(x, y) - \varphi(x)\}. \tag{1.5}$$

Thus, one can fix either $\varphi = \phi^c$ or $\phi = \varphi^c$ in the dual problem (D_0), giving rise to the following semi-dual formulation,

Semi-dual Kantorovich problem :

$$\sup_{\varphi\in L^1(d\mu)} \int \varphi(x) d\mu(x) + \int \varphi^c(y) d\nu(y). \tag{S_0}$$

These equivalent formulations (P_0'),(D_0) and (S_0) each allow different solvers to estimate the solution of Kantorovich problem, [138]. Under a number of situations, Monge and Kantorovich problems are in fact equivalent, [173]. This is the case for the quadratic cost $c(x, y) = \frac{1}{2}\|x - y\|^2$, if μ is continuous and if $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, by the result known in the literature as Brenier's theorem [27] and independently shown in the probabilistic literature by Cuesta and Matrán in [44]. Without finite second-order moments, even if the Monge-Kantorovich problem with quadratic cost does not make sense, a theorem of McCann states the existence of a monotonic mapping between μ and ν, [121]. The following, taken from [87], summarizes these fundamental results.

Theorem 1.1 (Brenier-McCann)**.** *Let $\mathbb{P}_1$ and $\mathbb{P}_2$ denote two distributions respectively supported on $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^d$. Suppose that $\mathbb{P}_1$ is absolutely continuous with respect to the Lebesgue measure. Then,*

(i) there exists a lower semi-continuous convex $\psi : \mathbb{R}^d \to \mathbb{R}$ such that $\nabla\psi_{\#}\mu = \nu$,

(ii) if ψ' is another such function, $\nabla\psi$ and $\nabla\psi'$ coincide $\mathbb{P}_1$-a.s.,

(iii) if $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{P}_2(\mathbb{R}^d)$, any such $\nabla\psi$ is a Monge map pushing $\mathbb{P}_1$ towards $\mathbb{P}_2$.

Besides, the latter convex function ψ involved in the μ-*a.e.* unique Monge map $\nabla\psi$ can be retrieved via the solution φ of (S_0), through $\psi(x) = \frac{1}{2}\|x\|^2 - \varphi(x)$, [27]. In other words, a Monge map shall verify

$$T(x) = x - \nabla\varphi(x). \tag{1.6}$$

In the one-dimensional setting, it is well-known that computing the classical quantile function of ν amounts to solve the optimal transport problem between ν and μ the uniform distribution over $[0, 1]$, [138][Remark 2.30]. By stating this as a definition when $d \geq 2$, the authors of [33] have extended univariate quantiles to define the Monge-Kantorovich quantile function. In the next part, we shall present the main definitions about Monge-Kantorovich quantile and distribution functions, taken from [33, 87]. Doing so, we motivate and introduce the contributions of the present thesis.

1.2 Multivariate statistics from optimal transport

1.2.1 Monge-Kantorovich quantiles

The first step is to depart from a reference distribution μ that is endowed with a natural ordering, so that a relevant notion of quantiles can be derived from it. For instance, one can choose the uniform distribution over some compact and convex set, such as the unit ball $\mathbb{B}(0, 1)$ or the unit square $[0, 1]^d$. Another natural choice is the *spherical uniform* distribution, denoted by $\mu = U_d$. The latter is given by the product $R\Phi$ between two independent random variables R and Φ, being drawn respectively from a uniform distribution on $[0, 1]$ and on the unit sphere. Samples from U_d are distributed from the origin to the outward within the unit ball, so that the balls of radius $\alpha \in [0, 1]$ have probability α while being nested, as α grows. With this in mind, the hyperspheres of radius α are relevant quantile contours with respect to U_d. Thus, the spherical uniform distribution U_d plays a specific role for the interpretation of quantiles. In addition, for U sampled from U_d, the radius $\|U\|$ and the unit vector $U/\|U\|$ are appropriate notions of multivariate ranks and signs, respectively, and their independence is also a convenient property.

Then, the basic idea of the MK quantile function of ν is to *transport* adequately the ordering of the reference μ towards ν.

Definition 1.2. *The MK quantile function of a multivariate distribution ν, with respect to a reference distribution μ, is the a.e. unique push-forward map $\mathbf{Q}_\#\mu = \nu$ such that $\mathbf{Q} = \nabla\psi$ for some convex potential $\psi : \mathbb{R}^d \to \mathbb{R}$.*

Moreover, the MK distribution function of ν is defined by $\mathbf{F} = \nabla\psi^*$ where ψ^* is the Fenchel-Legendre transform of ψ,

$$\psi^*(x) = \sup_{u \in \mathbb{B}(0,1)} \{\langle x, u\rangle - \psi(u)\}. \tag{1.7}$$

Being the gradient of a convex function is a multivariate kind of monotonicity, desirable for a quantile function, that is equivalent to

$$\forall x, y \in \mathbb{R}^d, \qquad \langle \mathbf{Q}(x) - \mathbf{Q}(y), x - y\rangle \geq 0.$$

It follows from Theorem 1.1 that, as soon as μ is absolutely continuous, such a Monge map $\mathbf{Q}$ exists and is unique. Moreover, if μ and ν have finite moments of order two, $\mathbf{Q}$ is characterized as the solution of the Monge problem of optimal transport. Regarding the MK distribution function $\mathbf{F}$, our definition follows from [74]. It is defined for an arbitrary ν, and it satisfies $\mathbf{F} = \mathbf{Q}^{-1}$ when $\mathbf{Q}$ is invertible, which holds under continuity assumptions on ν, as discussed in [74, 87].

When the reference measure on $\mathbb{R}^d$ is the spherical uniform distribution $\mu = U_d$, $\mathbf{Q}$ is also referred to as the *center-outward* quantile function of ν [87]. In view of $\mathbf{Q}_\#\mu = \nu$, the related ordering is adapted to the underlying geometry, as illustrated in Figure 1.3, where the radius in red and circles in blue are transported from the unit ball to a banana-shaped distribution thanks to a regularized estimator of $\mathbf{Q}$, to be introduced in Chapter 2. This relevant ordering clearly catches the geometry of the support of the target distribution ν, and it comes with quantile regions indexed by a probability level $\alpha \in [0, 1]$, by use of the change of variables formula for push-forward maps (1.4). More details are given in [33, 87].

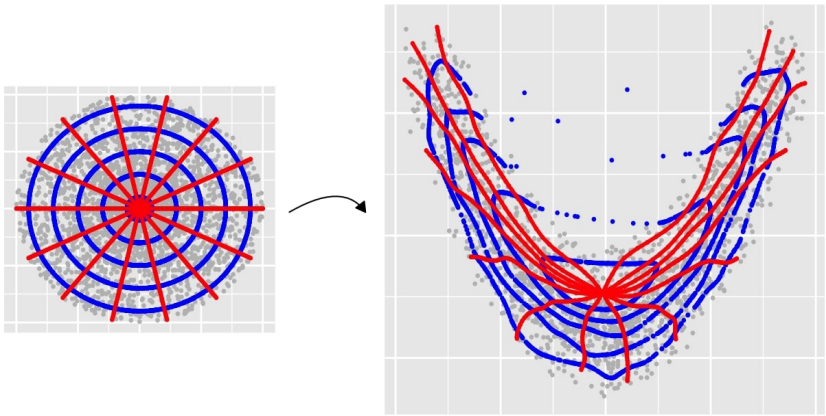


FIGURE 1.3 – (Left) center-outward quantiles of the spherical uniform distribution $\mu = U_d$, and (right) center-outward quantiles of a discrete distribution ν obtained by $\mathbf{Q}_\#\mu = \nu$.

Figure 1.3 illustrates the MK quantile contours and sign curves, as defined in the following.

Definition 1.3 (Quantile contours, ranks and signs)**.** *For ν supported on $\mathcal{X} \subset \mathbb{R}^d$ with center-outward quantile function $\mathbf{Q}$ and distribution function $\mathbf{F}$,*

(i) the quantile region $\mathbb{C}_\alpha$ of order $\alpha \in [0,1]$ is the image by $\mathbf{Q}$ of the ball $\mathbb{B}(0,\alpha)$,

(ii) the quantile contour $\mathcal{C}_\alpha$ of order $\alpha \in [0,1]$ is the boundary of $\mathbb{C}_\alpha$,

(iii) the rank function $\mathcal{R}_\nu : \mathcal{X} \to [0,1]$ is defined by $\mathcal{R}_\nu(x) = \|\mathbf{F}(x)\|$,

(iv) the sign function $\mathcal{D}_\nu : \mathcal{X} \to \mathbb{B}(0,1)$ is defined by $\mathcal{D}_\nu(x) = \mathbf{F}(x)/\|\mathbf{F}(x)\|$.

The ν-probability of $\mathbb{C}_\alpha$ is α, by the change of variables formula for push-forward maps, which is a first requirement for quantile regions. We emphasize that the choice of the reference measure $\mu = U_d$ plays a specific role for the interpretation of quantile regions indexed by a probability content level.

A major feature of our concepts is the ***distribution-freeness*** of MK ranks. In the real line, this mainly correspond to the fact that $F(X) \sim \text{Unif}([0,1])$ when the distribution function F is continuous, no matter the distribution ν. This is the main virtue of univariate ranks, or equivalently the empirical distribution function up to scaling by the number of observations. Indeed, on a sample $X_1, \cdots, X_n$, the usual ranks correspond to $F_n(X_1), \cdots, F_n(X_n)$ where F_n stands for the empirical distribution function. In other words, ranks are uniformly distributed on $\{1/n, 2/n, \cdots, 1\}$ with probability $1/n$ each, no matter the underlying distribution, and thus they are *distribution-free*. Univariate distribution-freeness can be rewritten as $F_{\#}\nu = \text{Unif}([0,1])$, and this is precisely the property that is put forward with the OT-based concepts from [33, 87]. As a matter of fact, this has inspired a wide range of literature interested in distribution-free statistical testing, see for instance the general framework of [50]. Among the existing statistical tests built upon these ideas, some have focused on goodness-of-fit, [50], others on independence [50, 132, 159-161], symmetry [92], various regression models, [82, 83, 86], or Pitman efficiency lower bounds [49]. We refer the reader to the introduction of [87] for a detailed discussion on rank-based inference and distribution-freeness, altogether with the notion of maximal ancillarity.

As opposed to these, we mainly focus on another major feature of MK quantiles : their ***descriptive power*** inherited from the fact that $\mathbf{Q}_{\#}\mu = \nu$. Notably, the data-adaptive ordering related to MK quantiles is not restricted to convex shapes, [33], in contrast with the majority of other existing definitions. In terms of random vectors, $\mathbf{Q}_{\#}\mu = \nu$ implies that $Q(U) \sim \nu$ as soon as $U \sim \mu$. A direct consequence is that $\mathbf{Q}$ must contain all the available information, which is appealing with the purpose of summing up unknown features of multivariate data. Indeed, as highlighted in [156], the utility of a multivariate proposal of quantile function for descriptive analysis is a key feature that is expected. Due to their intuitive definition, MK quantiles are easy to use, as for descriptive statistics that have already been derived either in terms of trimmed means, [60, 85], or in terms of dispersion measurements, [15].

1.2.2 The issue of regularized estimation

Finding regularized estimators that are also monotone in an appropriate fashion is a building block of MK quantiles [15, 87], and a pre-requisite before further investigation. Quite naturally, this is the first axis of the present thesis. A main idea towards this goal is to take advantage of the knowledge of the reference distribution μ to derive practical regularized estimators. To better highlight our contributions, we begin with an overview of the existing alternatives.

In practice, ν is often replaced by its empirical counterpart

$$\widehat{\nu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{Y_i}$$

built from a sample $\{Y_1, \cdots, Y_n\}$. Thus, it has a finite second order moment, and the estimation of the quantile map amounts to solve the plug-in version of the OT problem (P_0), that is

$$\min_{T:T_{\#}\mu=\widehat{\nu}_n} \mathbb{E}_{X\sim\mu}\left[\|X - T(X)\|^2\right]. \tag{1.8}$$

Such estimator, by nature, suffer from the impossibility to interpolate between the observations. Consequently, regularization naturally enters the picture, as within [15, 87] that provide interpolated maps built from empirical OT.

Optimal matchings and distribution-freeness

The estimation of $\mathbf{Q}$ using the plug-in estimator solving (1.8) can begin with various computational strategies to solve OT between μ and $\widehat{\nu}_n$. One can replace μ by a discrete measure $\widehat{\mu}_n$ on a regular grid and then solve a *discrete* OT problem between $\widehat{\mu}_n$ and $\widehat{\nu}_n$ as in [33, 87]. These solvers naturally render an optimal matching between two samples of same size. This is ideal to enforce the distribution-freeness of empirical MK ranks [74, 87], that may be crucial in practice, *e.g.* for statistical testing [50]. We refer the reader to Proposition 2.5 and Corollary 2.2 from [87]. However, one can think of other applications that require to compute out-of-sample estimates, where a mapping object $\mathbf{Q} : \mathbb{R}^d \to \mathbb{R}^d$ is highly desirable.

Map estimation by plug-in

A mapping object $\mathbf{Q} : \mathbb{R}^d \to \mathbb{R}^d$ can always be retrieved from the solution of (S_0). This was stressed in [33], with an estimator given by

$$\widehat{Q}_n(x) = \operatorname*{argsup}_{y\in\mathcal{Y}}\{\langle x, y\rangle - \frac{1}{2}\|y\|^2 + \widehat{\phi}_n(y)\}, \tag{1.9}$$

where $\widehat{\phi}_n$ is the solution of the semi-dual problem (S_0) between μ and $\widehat{\nu}_n$. However, this is not entirely satisfactory as $\widehat{Q}_n$ is restricted to the sample $\mathcal{Y} = \{Y_1, \cdots, Y_n\}$.

On the computational side, the cost of discrete OT solvers, [74, 87], is potentially very high because it scales cubically in the number of observations, [46]. Rather, it is proposed in [74] to compute the semi-dual problem (S_0) using the Newton-type algorithms proposed in [99]. Still, the estimator in [74] is of the form of (1.9), and suffers from the same restriction to $\mathcal{Y} = \{Y_1, \cdots, Y_n\}$. As a matter of fact, regularization is required to interpolate between these observations, [15, 87].

Regularized estimators

A satisfactory computation of quantile contours and sign curves requires a smooth empirical quantile function. One may cite the purpose of quantile regression where a prediction restricted to the observations is prohibitive, [11, 31, 136]. Smoothness was also the main consideration within [120], to use OT-based rank statistics as a differentiable loss for generative modeling. With Goodness-of-Fit purposes, computing the volumes of quantile regions in [15] also relied on differentiability of their estimator. Regularized empirical quantiles were also advocated in [85] for multivariate extensions of Lorenz curves, where it is required to estimate objects of the form of truncated expectations or integrals. As a last example, our multivariate superquantiles and expected shortfalls, [19], that will be discussed in Chapter 4, amount to integrals of $\mathbf{Q}$ along sign curves. There again, it is more satisfactory to use a smooth estimator for $\mathbf{Q}$ in practice.

The use of Moreau envelopes in [87] preserves the cyclical monotonicity as well as the optimal matchings (X_n, Y_n). These are ideal theoretical properties, because the result preserves the distribution-freeness of associated ranks. However, a supplementary gradient descent is required when computing a single $\mathbf{Q}(x)$ for $x \in \mathcal{X}$. This is alleviated in [15] with an approximation of $\mathbf{Q}$ rather than an interpolation. This yields a smooth estimator for $\mathbf{Q}$ that is consistent while ensuring the gradient-of-convex property. Nevertheless, the most famous regularization of OT is certainly the entropic one, that is a cornerstone of modern computational OT, [138]. In comparison with the approaches above, the use of entropic optimal transport (EOT) represents a step towards more regularization, with a monotone estimator related to recent advances in computational OT. Notably, insightful results were obtained in the context of MK quantiles in [31, 120].

Specifically, our context imposes the knowledge of the reference measure μ *a priori*, as it is fixed once for all. As a byproduct, it is appealing to take advantage of this additional information. This motivated the first main direction of the present thesis, to construct a solver for EOT that makes use of the knowledge of the reference measure μ, [19, 20]. In view of convergence guarantees, entropic regularization of OT provides known computational advantages, [138]. This comes with additional benefits regarding the smoothness and monotonicity of the regularized estimator. To better emphasize our contribution, we shall now present the entropic regularization of OT and the resulting *entropic map*.

Entropic maps under general assumptions

The EOT problem can be introduced without imposing finiteness of second-order moments, thanks to [24, 73]. In the quantiles context, this is an important concern of the recent papers [33, 87], as a quantile function must be defined for every measure ν. Define the Kullback-Leibler divergence between joint probabilities, for $\pi, \xi \in \Pi(\mu, \nu)$, by

$$\mathrm{KL}(\pi|\xi) = \int_{\mathcal{X}\times\mathcal{Y}} \left(\log\left(\frac{d\pi}{d\xi}(x,y) \right) - 1 \right) d\pi(x,y),$$

for $\frac{d\pi}{d\xi}$ the relative density of π with respect to ξ. Then, the primal version of EOT [47, 70] is given by

$$\min_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\pi(x,y) + \varepsilon \mathrm{KL}(\pi|\mu\otimes\nu), \tag{P_ε}$$

where $\varepsilon \geq 0$ stands for a regularization parameter. We say that this problem has a finite value when there exists $\pi_0 \in \Pi(\mu,\nu)$ such that

$$\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\pi_0(x,y) + \varepsilon \mathrm{KL}(\pi_0|\mu\otimes\nu) < +\infty.$$

Under such condition, the property of cyclical invariance introduced in [24, 73] has been showed to characterize optimality in (P_ε). In addition, for continuous costs, Theorem 1.3 from [73] ensures the existence of a unique cyclically invariant coupling π_ε in $\Pi(\mu,\nu)$ even if all couplings have infinite cost.

Thus, for the particular case of c being the quadratic cost, one can consider such a coupling π_ε even if μ and ν have infinite variance, for any $\varepsilon > 0$. Now, denote by $u^{c,\varepsilon}$ the smooth c-transform of $u \in L^1(\mu)$ given by

$$u^{c,\varepsilon}(y) = -\varepsilon \log\left(\int_{\mathcal{X}} \exp\left(\frac{u(x) - c(x,y)}{\varepsilon} \right) d\mu(x) \right). \tag{1.10}$$

For φ given in [73][Corollary 2.6], by taking $\mathbf{u}_\varepsilon(x) = \varepsilon \log \varphi(x)$, it exists a unique, up to additive constants, potential $\mathbf{u}_\varepsilon : \mathbb{R}^d \to \mathbb{R}$ satisfying

$$\mathbf{u}_\varepsilon = (\mathbf{u}_\varepsilon^{c,\varepsilon})^{c,\varepsilon}. \tag{1.11}$$

Moreover, from the proof of [73][Corollary 2.6], π_ε writes

$$d\pi_\varepsilon(x,y) = \exp\left(\frac{\mathbf{u}_\varepsilon(x) + \mathbf{u}_\varepsilon^{c,\varepsilon}(y) - c(x,y)}{\varepsilon} \right) d\mu(x)d\nu(y). \tag{1.12}$$

We now introduce the dual and semi-dual counterparts of (P_ε) as stated in [71][Pro-

position 2.1]. For

$$\mathcal{I}_\varepsilon(u,v) = \varepsilon \int \exp\left(\frac{u(x)+v(y)-c(x,y)}{\varepsilon}\right) d\mu(x)d\nu(y),$$

the *dual* EOT problem writes

$$\max_{(u,v)\in L^1(\mu)\times L^1(d\nu)} \int_{\mathcal{X}} u(x)d\mu(x) + \int_{\mathcal{Y}} v(y)d\nu(y) - \mathcal{I}_\varepsilon(u,v), \tag{D_ε}$$

and the *semi-dual* EOT problem is given by

$$\max_{u\in L^1(\mu)} \int_{\mathcal{X}} u(x)d\mu(x) + \int_{\mathcal{Y}} u^{c,\varepsilon}(y)d\nu(y) - \varepsilon. \tag{S_ε}$$

The regularized problems are equivalent to (P_0'), (D_0) and (S_0) when $\varepsilon = 0$, and $u^{c,0}(y) = u^c(y)$. As argued in [70, 71], first-order optimality conditions for (D_ε) write $u = v^{c,\varepsilon}$ and $v = u^{c,\varepsilon}$, see also [133]. The semi-dual (S_ε) is obtained by plugging $v = u^{c,\varepsilon}$ in (D_ε). In view of these first-order conditions, the potential $\mathbf{u}_\varepsilon$ solution of (S_ε) must verify (1.11), even when μ and ν have infinite variance.

From a solution of the EOT problem, one can retrieve the *entropic map*, [140, 153]. It leverages the efficiency of entropic regularization to obtain a smooth function, defined by barycentric projection of the optimal plan π_ε in (1.12), that is $\mathbf{Q}_\varepsilon(x) = \mathbb{E}_{(X,Y)\sim\pi_\varepsilon}[Y|X=x]$. Using (1.12) and (1.11), a closed-form expression of $\mathbf{Q}_\varepsilon$ is given by

$$\mathbf{Q}_\varepsilon(x) = \int_{\mathcal{Y}} y \frac{\exp\left(\frac{\mathbf{u}_\varepsilon^{c,\varepsilon}(y)-c(x,y)}{\varepsilon}\right)}{\int_{\mathcal{Y}} \exp\left(\frac{\mathbf{u}_\varepsilon^{c,\varepsilon}(z)-c(x,z)}{\varepsilon}\right) d\nu(z)} d\nu(y). \tag{1.13}$$

It follows once again from (1.11) together with a direct calculation that the above coincides with $\mathbf{Q}_\varepsilon(x) = x - \nabla \mathbf{u}_\varepsilon(x)$, the entropic counterpart of (1.6), see [140][Proposition 2]. Similarly, one can define an entropic map from ν towards μ, through

$$\mathbf{F}_\varepsilon(y) = \int_{\mathcal{X}} x \frac{\exp\left(\frac{\mathbf{u}_\varepsilon(x)-c(x,y)}{\varepsilon}\right)}{\int_{\mathcal{X}} \exp\left(\frac{\mathbf{u}_\varepsilon(z)-c(z,y)}{\varepsilon}\right) d\mu(z)} d\mu(x). \tag{1.14}$$

Here, it is important to note that $\mathbf{Q}_\varepsilon$ does not push μ forward to ν anymore, although it is expected to be close to it for ε close to 0. Indeed, the theoretical behavior for vanishing ε has been the subject of many works, showing convergence towards OT couplings, [24, 30, 107], potentials, [4, 52, 133] or maps [139, 140]. In addition, for a fixed regularization parameter $\varepsilon > 0$, plug-in estimators for the costs ([13, 72, 123]), potentials and maps ([13, 77, 78, 143]) can converge to their regularized population counterparts with a statistical rate independent from the dimension, which is a particularly appealing feature.

In practice, considering a fixed $\varepsilon > 0$ can be gainful because of the regularity it induces. Notably, the definition and continuity of $\mathbf{Q}_\varepsilon$ does not depend on the regularity of ν, which is in sharp contrast with the unregularized setting. In fact, $\mathbf{Q}_\varepsilon$ benefits from greater regularity. Let $\psi_\varepsilon(x) = \frac{1}{2}\|x\|^2 - \mathbf{u}_\varepsilon(x)$, so that $\mathbf{Q}_\varepsilon(x) = x - \nabla \mathbf{u}_\varepsilon(x) = \nabla \psi_\varepsilon(x)$. Direct computation of the Hessian of ψ_ε shows that it is a positive definite matrix, so that ψ_ε is strictly convex, see [34][Lemma 1]. Thus, Q_ε is the gradient of a convex function, that is the desired multivariate monotonicity in our context. As a byproduct, the interpolation that $\mathbf{Q}_\varepsilon$ provides between observations in practice preserves the structure of optimality. Of major importance, the continuity of $\mathbf{Q}_\varepsilon$ yields continuous quantile contours and strictly nested and closed quantile regions of any order $\alpha \in [0, 1]$.

Besides, this has direct implications regarding the invertibility of $\mathbf{Q}_\varepsilon$, that can be particularly relevant in view of MK ranks. Indeed, $\mathbf{Q}_\varepsilon$ being the gradient of the strictly convex function ψ_ε, [146][Theorem 26.5] ensures that $\mathbf{Q}_\varepsilon$ is one-to-one from $\mathcal{X}$ to $\mathcal{Y}$, with $\mathbf{Q}_\varepsilon^{-1}(y) = \nabla \psi_\varepsilon^*(y)$ for

$$\psi_\varepsilon^*(y) = \sup_{x \in \mathcal{X}} \big\{ \langle x, y \rangle - \psi_\varepsilon(x) \big\}$$

the Legendre transform of ψ_ε. Consequently, the envelope theorem allows to write

$$\mathbf{Q}_\varepsilon^{-1}(y) = \operatorname*{argsup}_{x \in \mathcal{X}} \big\{ \langle x, y \rangle - \psi_\varepsilon(x) \big\}. \tag{1.15}$$

For the above reasons, $\mathbf{Q}_\varepsilon$ is appealing for the estimation of regularized MK quantiles. On the computational side, from an estimate of $\mathbf{u}_\varepsilon^{c,\varepsilon}$, a plug-in estimator of $\mathbf{Q}_\varepsilon$ in (1.13) can be computed in linear time. Thus, the estimation of dual potentials $\mathbf{u}_\varepsilon$ and $\mathbf{u}_\varepsilon^{c,\varepsilon}$ is the main computational bottleneck. For this task, as previously argued, it is tempting to take advantage of the information of the reference measure μ to design a computational scheme specifically for MK quantiles.

1.3 Contributions and outline of the thesis

1.3.1 A new approach to solve continuous EOT

Making use of the knowledge of the reference measure μ allows to develop a fast solver that targets the *continuous* EOT problem between μ and ν at the limit of the iterations, [19]. This is the first main contribution of the present thesis, presented in Chapter 2. It relies on a parameterization of $\mathbf{u}_\varepsilon$ by its Fourier series, in the semi-dual problem (S_ε). This allows to compute partial derivatives, hence to construct a stochastic gradient descent in the space of Fourier coefficients, that are constrained to belong to the Banach space ℓ_1 by construction.

This algorithm is proven to be almost surely consistent with the help of tools from stochastic optimization. Importantly, a main difficulty arising here is the requirement that our stochastic sequence belongs to the infinite-dimensional Banach

space ℓ_1, that is customary to ensure well-defined gradients along the iterations.

Moreover, a main feature of our approach is that our convergence results are directly formulated as a function of the number n of observations $Y_1, \cdots, Y_n$ from ν, in contrast to *discrete* or *semi-discrete* solvers that are estimators of the plug-in problem between μ and $\widehat{\nu}_n$. In practice, the computational complexity at each iteration is inherited from our use of Fast Fourier transforms. Consequently, it depends on the size of a chosen grid of size p independent from n, which can prove beneficial in practice.

1.3.2 Directional counterparts

Going further, this computational approach can be extended beyond the setting of euclidean data, following recent advances, [84]. In Chapter 3, it will be stressed that the definition of the *entropic map* can be extended to a non-euclidean setting, with data on the hypersphere $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$. This is of major importance because the need for smooth and efficient estimators is the same in $\mathbb{S}^{d-1}$ than in $\mathbb{R}^d$. In this context, the regularized Kantorovich potentials can be parameterized by their series of spherical harmonics, that are spherical analogs of Fourier series. There again, a stochastic algorithm can be formulated that targets the *continuous* EOT problem between the reference measure μ and the target measure ν.

Such regularized estimation allows applications in descriptive analysis, to compare distributions depending on their dispersion, via volumes of quantile regions. Besides, a notion of statistical depth allows to classify an observation depending on its centrality with respect to several point clouds, [113]. Applications for classification are thus considered, with the help of a new directional version of the MK statistical depth, [33]. Our numerical approach based on EOT appears especially useful and convenient for these purposes. As an aside, this directional MK depth shows appealing properties with regards to the traditional Liu-Zuo-Serfling axioms. In comparison with existing notions of quantiles for spherical data, the adaptivity to the geometry of a distribution reveals particularly satisfying for descriptive analysis, as illustrated by numerical experiments. Moreover, another byproduct is that classification based on the spherical MK depth improves performance with respect to other notions of quantiles. These findings provide motivations for further use of directional MK quantiles for inference, in a regularized fashion.

1.3.3 New concepts of superquantiles and expected shortfalls

Inspired by the descriptive power of MK quantiles, another main direction of this thesis has been to develop multivariate notions of superquantiles and expected shortfalls derived from MK quantiles, that is an open problem of the founding paper [87]. Recall that the combination between MK quantile contours and the sign curves

provide a curvilinear coordinate system within the support of the distribution of interest ν, as illustrated in Figure 1.3. This is a convenient and promising way to render the available information, that shows how the directional information is contained in the sign curves. In dimension $d = 1$, superquantiles and expected shortfalls are defined from the quantile function Q by

$$S(\alpha) = \frac{1}{1-\alpha}\int_{\alpha}^{1} Q(t)dt \qquad \text{and} \qquad E(\alpha) = \frac{1}{\alpha}\int_{0}^{\alpha} Q(t)dt.$$

The definitions of Chapter 4 provide similar integral versions of the MK quantile function $\mathbf{Q}$, but along sign curves, that is

$$S(u) = \frac{1}{1-\|u\|}\int_{\|u\|}^{1} \mathbf{Q}\Big(t\frac{u}{\|u\|}\Big)dt \qquad \text{and} \qquad E(u) = \frac{1}{\|u\|}\int_{0}^{\|u\|} \mathbf{Q}\Big(t\frac{u}{\|u\|}\Big)dt.$$

Doing so, the MK superquantile and expected shortfall functions preserve the univariate interpretation of a *typical* observation *beyond* or *ahead* a quantile $\mathbf{Q}(u)$. These functions will be shown to have suitable properties, as they characterize random vectors and their convergence in distribution, under some assumptions. These theoretical findings extend univariate properties, [145], and highlight the importance of the proposed notions. In Figure 1.3, the circles and radius were transported by the MK quantile function $\mathbf{Q}$. Using instead the integrated counterparts S and E, one can define MK superquantile and expected shortfall contours and averaged sign curves, that are illustrated in the representative plots of Figure 1.4, showing how E and S characterize respectively central or peripheral areas of a point cloud.

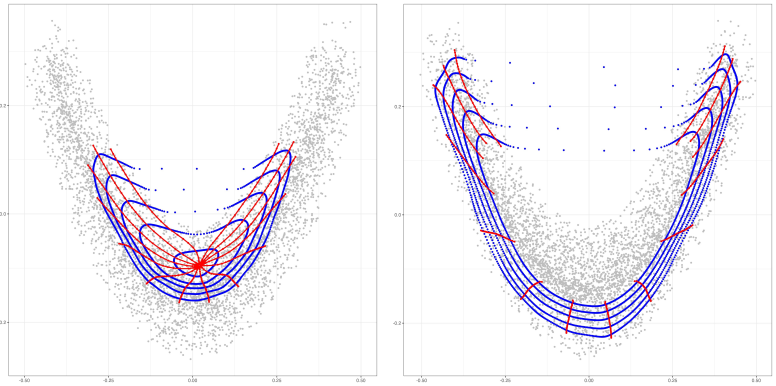


FIGURE 1.4 – Center-outward expected shortfall (left) and superquantile (right) contours in blue and averaged sign curves in red.

On the real line, a major application of these functions is risk analysis, [79], although several authors argue for a wider use in data science, [105, 145]. Indeed, in dimension $d = 1$, the most well-known risk measures certainly are the Value-at-Risk and Conditional-Value-at-Risk, that are simply defined by a quantile and by a superquantile. It is then natural to explore the multivariate extension of these notions with the help of our new integrated quantile functions, which is done with

an emphasis on the univariate interpretation. These are to be understood respectively as the worst risk encountered with ν-probability α, and as the averaged risk beyond this worst-case observation, but in a meaningful multivariate way.

We end this introduction with an outline of the thesis.

Chapter 2 : Stochastic algorithms to solve optimal transport

A first step for the estimation of the entropic map is to solve the dual version of entropic optimal transport (EOT). This Chapter is devoted to a new stochastic algorithm designed to directly solve the *continuous* problem between arbitrary probability measures μ and ν. Using the knowledge of the source measure, it is suggested to parametrize a Kantorovich dual potential by its Fourier coefficients. In this way, each iteration of the proposed stochastic algorithm reduces to two Fourier transforms that enables to make use of the Fast Fourier Transform (FFT) in order to implement a fast numerical method to solve EOT. Under suitable assumptions, we prove the almost sure convergence of our stochastic algorithm, that takes its values in an infinite-dimensional Banach space. Numerical experiments illustrate the performances on the computation of regularized Monge-Kantorovich quantiles. In particular, the potential benefits of entropic regularization are investigated for the smooth estimation of multivariate quantiles using data sampled from the target measure ν.

Chapter 3 : Directional statistics

The goal of this Chapter is to develop a computational approach for the estimation of regularized quantiles on the unit sphere $\mathbb{S}^2$. The definition of Monge-Kantorovich quantile and distribution functions has recently been extended for directional data in [84]. In practice, regularization is mandatory for applications that require out-of-sample estimates. To this end, the definition of the entropic map is extended to the spherical setting. A stochastic algorithm to directly solve a continuous EOT problem between the uniform distribution and a target distribution is proposed, by expanding Kantorovich potentials in the basis of spherical harmonics. In addition, a directional Monge-Kantorovich depth is defined, and it is shown to benefit from desirable properties related to the traditional Liu-Zuo-Serfling axioms for the statistical analysis of directional data.

Chapter 4 : Superquantiles and expected shortfalls

This chapter deals with new notions of center-outward superquantile and expected shortfall functions, that provide a natural way to characterize multivariate tail probabilities and central areas of point clouds. Crucially, our new concepts preserve the univariate interpretation of a typical observation that lies beyond or

ahead a quantile, but in a meaningful multivariate way. Moreover, these functions characterize random vectors, and their convergence is shown to be equivalent to convergence in distribution under some assumptions, which underlines their importance. Building on these notions appears fruitful to extend the standard notion of value at risk and conditional value at risk from the real line to $\mathbb{R}^d$. The behavior of our risk measurements is illustrated on both real and simulated datasets.

This thesis gave rise to the writing of the following papers.

Bernard, Bercu, Jérémie Bigot, and Gauthier Thurin.
Stochastic optimal transport in Banach Spaces for regularized estimation of multivariate quantiles. *arXiv :2302.00982* (2023)

Bernard, Bercu, Jérémie Bigot, and Gauthier Thurin.
Monge-Kantorovich superquantiles and expected shortfalls with applications to multivariate risk measurements. *arXiv :2307.01584* To appear in Electronic Journal of Statistics, vol 18, (2024)

Bernard, Bercu, Jérémie Bigot, and Gauthier Thurin.
Regularized estimation of Monge-Kantorovich quantiles for spherical data. *arXiv :2407.02085* (2024).

2

Stochastic algorithms to solve optimal transport

We introduce a new solver for entropic optimal transport (EOT) between two absolutely continuous probability measures μ and ν. Using the knowledge of the source measure, we propose to parametrize a Kantorovich dual potential by its Fourier coefficients. In this way, each iteration of our stochastic algorithm reduces to two Fourier transforms that enables us to make use of the Fast Fourier Transform (FFT) in order to implement a fast numerical method to solve EOT. Under suitable assumptions, we prove the almost sure convergence of our procedure that takes its values in an infinite-dimensional Banach space. This chapter is based on [19].

Contents

2.1 Solving entropic optimal transport

2.1.1 Different types of solvers

A main reason behind the wide use of OT-based methods in statistics and machine learning, [46], is certainly the efficiency of Sinkhorn algorithm for EOT, [47], whose complexity scales quadratically in the number of data, while being adapted to parallel implementation. A possible concern of it is the storage of the cost matrix between large sets of points, which motivated stochastic algorithms, [17, 22, 71], for either the *discrete*, the *semi-discrete* or the *continuous* setting. These refer to the assumptions on the measures μ and ν involved, either if both are discrete, if one is continuous, or if both are continuous, respectively. For the latter, the leading idea is to parametrize dual potentials by classes of functions. For this purpose, [153] used neural networks, making the problem lose its convexity. Instead, consistency holds for the RKHS parametrization of [71], or for kernel expansions based on regularized c-transforms, [124]. For both of them, to ensure convergence, the number of parameters grows with the iterations, increasing numerical complexity.

As opposed to these, we propose Fourier series when the reference measure is the uniform over the hypercube $[0,1]^d$. Indeed, the quantiles' context fixes the reference measure, and it is tempting to make use of this information *a priori*. We stress that our theoretical results do not depend on the orthonormal basis for the potentials, neither on the cost, so that our algorithm may be adapted, *e.g.* for wavelets on $\mathbb{R}^d$. As an aside, we mention that non equispaced Fast Fourier Transforms were used recently for better time and memory allocations in *discrete* Sinkhorn, [106], albeit not very much related to our proposal of Fourier parametrization of dual potentials.

2.1.2 Stochastic algorithms

Stochastic algorithms to solve EOT were first introduced in [71], by noticing that the dual (D_ε) and semi-dual (S_ε) objectives can be rewritten as expectations of a convex loss. This approach is particularly useful when the number of observations n is large and prevents the storage of the cost matrix, of size n^2. In the *semi-discrete* case, one is interested in the EOT problem (D_ε) between the absolutely continuous measure μ and the empirical measure $\widehat{\nu}_n$ associated with a sample $Y_1, \cdots, Y_n$. There, the potential $v \in L^1(\widehat{\nu}_n)$ can be identified with its values $(v(Y_i)) \in \mathbb{R}^n$. As studied in [17, 22, 69, 71], this yields the following optimisation problem, for $\varepsilon > 0$,

$$\min_{v \in \mathbb{R}^n} \widetilde{H}_\varepsilon(v) \qquad \text{with} \qquad \widetilde{H}_\varepsilon(v) = \mathbb{E}[\widetilde{h}_\varepsilon(v, X)] \tag{2.1}$$

where X is a random variable with distribution μ, and

$$\widetilde{h}_\varepsilon(v,x) = \varepsilon \log\left(\frac{1}{n}\sum_{j=1}^n \exp\left(\frac{v_j - c(x,Y_j)}{\varepsilon}\right)\right) - \frac{1}{n}\sum_{j=1}^n v_j + \varepsilon.$$

Based on independent samples $X_1,\dots,X_m$ from μ, these works approach the unique solution $\widetilde{v}_n \in \mathbb{R}^n$ of the problem (2.1) when $m \to +\infty$ and n is held fixed. From $\widetilde{v}_n$, as from any estimate of the regularized dual potential, one can derive an empirical entropic map. It is sufficient, for all $x \in \mathcal{X}$, to consider the plug-in of (1.13) by

$$\widetilde{Q}_\varepsilon^n(x) = \sum_{j=1}^n \widetilde{F}_j(x) Y_j \qquad \text{where} \qquad \widetilde{F}_j(x) = \frac{\exp\left(\frac{\widetilde{v}_{n,j} - c(x,Y_j)}{\varepsilon}\right)}{\sum_{\ell=1}^n \exp\left(\frac{\widetilde{v}_{n,\ell} - c(x,Y_\ell)}{\varepsilon}\right)}. \quad (2.2)$$

Hereafter, we shall propose another rewriting of (S_ε), to directly target the *continuous* EOT problem between *two arbitrary measures* μ *and* ν at the limit of the n iterations. The numerical performances of the associated empirical entropic map will be compared to $\widetilde{Q}_\varepsilon^n$ in Section 2.4.

2.2 From functions to Fourier coefficients

From now on, μ is assumed to be the uniform distribution on $\mathcal{X} = [0,1]^d$, except in some of the numerical experiments carried out in Section 2.4 where a change of variable enables to consider the spherical uniform distribution U_d for which $\mathcal{X} = \mathbb{B}(0,1)$. Moreover, ν is assumed to have a finite moment of order two. Then, we consider the normalization condition for the dual potential in (S_ε)

$$\int_{\mathcal{X}} u(x) d\mu(x) = 0. \quad (2.3)$$

Taking the support of μ to be equal to $[0,1]^d$ is motivated by the choice to parametrize a dual function $u \in L^1(\mu)$, satisfying the identifiability condition (2.3), by its decomposition in the standard Fourier basis $\phi_\lambda(x) = e^{2\pi i \langle \lambda, x\rangle}$, for $\lambda \in \mathbb{Z}^d$,

$$u(x) = \sum_{\lambda \in \Lambda} \theta_\lambda \phi_\lambda(x),$$

where $\Lambda = \mathbb{Z}^d \backslash \{0\}$ and $\theta = (\theta_\lambda)_{\lambda\in\Lambda}$ are the Fourier coefficients of u,

$$\theta_\lambda = \int_{\mathcal{X}} \overline{\phi_\lambda(x)} u(x) d\mu(x). \quad (2.4)$$

We refer to [164] for an introduction to multiple Fourier series on the *flat torus* $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$. Hereafter, $\mathbb{T}^d$ stands for the set of equivalence classes $[x] = \{x + k\,;\, k \in \mathbb{Z}^d\}$ for all $x \in [0,1[^d$. With a slight abuse of notation, we identify $\mathbb{T}^d$ to its fundamental

domain $[0,1[^d$, so that integration on $\mathbb{T}^d$ is Lebesgue-integration on $[0,1[^d$, see [164] or [38, 116] in the OT literature. Then, for a given regularization parameter $\varepsilon > 0$, we rewrite the dual problem (D_ε) with this parametrization, to consider, for $\ell_1(\Lambda)$ defined hereafter, the following *stochastic convex minimisation* problem

$$\theta^\varepsilon = \underset{\theta \in \ell_1(\Lambda)}{\operatorname{argmax}} \; H_\varepsilon(\theta) \qquad \text{with} \qquad H_\varepsilon(\theta) = \mathbb{E}\left[h_\varepsilon(\theta, Y)\right] \tag{2.5}$$

where Y is a random vector with distribution ν and

$$h_\varepsilon(\theta, y) = \varepsilon \log\left(\int_{\mathcal{X}} \exp\left(\frac{\sum_{\lambda\in\Lambda} \theta_\lambda \phi_\lambda(x) - c(x,y)}{\varepsilon}\right) d\mu(x)\right) + \varepsilon.$$

Remark 2.1. *We emphasize that the formulation (2.5) can be considered with any other orthonormal basis of functions ϕ_λ, or ground cost c, with (2.3) and (2.4). Our theoretical results would then continue to hold, under the assumption that $c \in L^1(\mu \times \nu)$. In particular, Chapter 3 deals with spherical harmonics on the 2-sphere, with a quadratic cost inherited from the Riemannian distance.*

In view of a gradient descent, one needs to compute first-order derivatives. We refer to [95, Chapter 8] for a basic course on Fréchet differentiability and Taylor formulas for functions between Banach spaces. In Section 2.3.2, it is shown that, for every $y \in \mathcal{Y}$, the function $\theta \mapsto h_\varepsilon(\theta, y)$ is Fréchet differentiable only if θ belongs to the convex set

$$\ell_1(\Lambda) = \left\{\theta = (\theta_\lambda)_{\lambda\in\Lambda} \in \mathbb{C}^\Lambda \;:\; \theta_{-\lambda} = \overline{\theta_\lambda} \text{ and } \|\theta\|_{\ell_1} = \sum_{\lambda\in\Lambda} |\theta_\lambda| < +\infty\right\}.$$

Moreover, its differential $D_\theta h_\varepsilon(\theta, y)$ is identified as an element of the dual Banach space

$$\ell_\infty(\Lambda) = \left\{v = (v_\lambda)_{\lambda\in\Lambda} \in \mathbb{C}^\Lambda : v_{-\lambda} = \overline{v_\lambda} \text{ and } \|v\|_{\ell_\infty} = \sup_{\lambda\in\Lambda} |v_\lambda| < +\infty\right\}.$$

The components of the first order Fréchet derivative $D_\theta h_\varepsilon(\theta, y)$ are the partial derivatives

$$\frac{\partial h_\varepsilon(\theta, y)}{\partial \theta_\lambda} = \int_{\mathcal{X}} \overline{\phi_\lambda(x)} F_{\theta,y}(x) d\mu(x) \tag{2.6}$$

that are the Fourier coefficients of the function

$$F_{\theta,y}(x) = \frac{\exp\left(\frac{\sum_{\lambda\in\Lambda} \theta_\lambda \phi_\lambda(x) - c(x,y)}{\varepsilon}\right)}{\int_{\mathcal{X}} \exp\left(\frac{\sum_{\lambda\in\Lambda} \theta_\lambda \phi_\lambda(x) - c(x,y)}{\varepsilon}\right) d\mu(x)}. \tag{2.7}$$

One can observe that $F_{\theta,y}$ is a probability density function, which is a key property that we shall repeatedly use.

Let (Y_n) be a sequence of independent random vectors sharing the same dis-

tribution ν. In the spirit of [144], we propose to estimate the solution of (2.5) by considering the stochastic algorithm in the Banach space $(\ell_1(\Lambda), \|\cdot\|_{\ell_1})$ defined, for all $n \geq 0$, by

$$\widehat{\theta}_{n+1} = \widehat{\theta}_n - \gamma_n W D_\theta h_\varepsilon(\widehat{\theta}_n, Y_{n+1}) \tag{2.8}$$

where $\gamma_n = \gamma n^{-\beta}$ with $\gamma > 0$ and $1/2 < \beta \leq 1$, which clearly implies the standard conditions

$$\sum_{n=0}^{\infty} \gamma_n = +\infty \qquad \text{and} \qquad \sum_{n=0}^{\infty} \gamma_n^2 < +\infty. \tag{2.9}$$

Moreover, W is the following linear operator

$$\left\{ \begin{array}{rcl} W : (\ell_\infty(\Lambda), \|\cdot\|_{\ell_\infty}) & \to & (\ell_1(\Lambda), \|\cdot\|_{\ell_1}) \\ v = (v_\lambda)_{\lambda\in\Lambda} & \mapsto & w \odot v = (w_\lambda v_\lambda)_{\lambda\in\Lambda} \end{array} \right.$$

where $w = (w_\lambda)_{\lambda\in\Lambda}$ is a *deterministic sequence of positive weights* satisfying the normalizing condition

$$\|w\|_{\ell_1} = \sum_{\lambda\in\Lambda} w_\lambda < +\infty. \tag{2.10}$$

A main difficulty arising here is that the space $\ell_1(\Lambda)$ of parameters differs from its dual space $\ell_\infty(\Lambda)$ to which the Fréchet derivative $D_\theta h_\varepsilon(\theta, y)$ belongs. This is a classical issue when considering convex optimization in Banach spaces, see e.g. [28], and this is the reason why we introduce the linear operator W in (2.8) that maps $\ell_\infty(\Lambda)$ to $\ell_1(\Lambda)$. The use of the linear operator W also induces two weighted norms on the space

$$\ell_2(\Lambda) = \left\{ \theta = (\theta_\lambda)_{\lambda\in\Lambda} \in \mathbb{C}^\Lambda : \theta_{-\lambda} = \overline{\theta_\lambda} \text{ and } \|\theta\|_{\ell_2}^2 = \sum_{\lambda\in\Lambda} |\theta_\lambda|^2 < +\infty \right\}.$$

Trivially, $\ell_1(\Lambda) \subset \ell_2(\Lambda)$.

Definition 2.2. *For every $\theta \in \ell_2(\Lambda)$ and for a sequence $w = (w_\lambda)_{\lambda\in\Lambda}$ of positive weights satisying (2.10), we define the two weighted norms*

$$\|\theta\|_W^2 = \sum_{\lambda\in\Lambda} w_\lambda |\theta_\lambda|^2 \qquad \text{and} \qquad \|\theta\|_{W^{-1}}^2 = \sum_{\lambda\in\Lambda} w_\lambda^{-1} |\theta_\lambda|^2. \tag{2.11}$$

Remark 2.3. *Imposing that the Fourier coefficients $\theta = (\theta_\lambda)_{\lambda\in\Lambda}$ form an absolutely convergent series implicitly requires that the optimal dual potential minimizing (D_ε) satisfy periodic conditions at the boundary of $[0,1]^d$. For the clarity of the exposure, a detailed discussion on sufficient conditions for the un-regularized optimal dual potential u_0 to be periodic is postponed to Appendix A.2.*

Hereafter, a *regularized estimator* of the potential defined from θ^ε in (2.5), by

$$u_\varepsilon(x) = \sum_{\lambda\in\Lambda} \theta_\lambda^\varepsilon \phi_\lambda(x) \tag{2.12}$$

is naturally given by

$$\widehat{u}_\varepsilon^n(x) = \sum_{\lambda \in \Lambda} \widehat{\theta}_{n,\lambda} \phi_\lambda(x). \tag{2.13}$$

In practice, our numerical procedure starts by considering a discretization of the dual potential u over a regular grid $\mathcal{X}_p = \{x_1, \ldots, x_p\}$ of points in $\mathcal{X}$. This allows us to compute the corresponding set of Fourier coefficients at frequencies Λ_p of size p by the Fast Fourier Transform (FFT). Then, the sequence $(\widehat{\theta}_{n,\lambda})_{\lambda \in \Lambda_p}$ satisfying (2.8) is easily implemented using, at each iteration, the FFT and its inverse, see Algorithm 1 below. Hence, the computational cost, at each iteration, of our algorithm is of order $\mathcal{O}(p \log(p))$, while the cost of the celebrated Sinkhorn algorithm [47] is $\mathcal{O}(pn)$, using a discrete source measure supported on $\mathcal{X}_p$, and the one of the semi-discrete stochastic algorithms proposed in [17, 22, 71] is $\mathcal{O}(n)$ at each iteration.

In our approach, the computational cost depends on the size p of the grid on $\mathcal{X}_p$ that is fixed by the user. This size p does not require to be particularly large, as showed by numerical experiments. However, we stress that this appealing computational cost of $\mathcal{O}(p \log(p))$ comes with a drawback regarding the dimension. Indeed, the number p of points in a uniform grid on $[0,1]^d$ grows exponentially with d. Thus, a standard implementation of the FFT on a uniform grid becomes difficult for medium dimensions such as $d = 10$. Extending our work to the high-dimensional setting would require the study of more sophisticated FFTs as proposed in [141], but this issue is beyond the scope of this chapter.

Algorithm 1 Stochastic algorithm

Initialize $N \in \mathbb{N}$, $\mathcal{X}_p = \{x_1, \cdots, x_p\}$, $u \in \mathbb{R}^p$ and $W \in \mathbb{R}^{p \times p}$
$\theta \leftarrow \mathrm{FFT}(u)$
while $n \leq N$ **do**
  $y \leftarrow Y_n$
  $u \leftarrow \mathrm{IFFT}(\theta)$
  for $i \in \{1, \cdots, p\}$ **do**
    $F[i] \leftarrow \exp\Big((u[i] - c(x_i, y))/\varepsilon\Big)$
  end for
  $F \leftarrow F/\mathrm{mean}(F)$ ▷ estimate of (2.7)
  $\mathrm{grad} \leftarrow \mathrm{FFT}(F)$
  $\theta \leftarrow \theta - \gamma_n W \cdot \mathrm{grad}$
end while

Barycentric projection

Since the entropic map can be estimated from any solution of (D_ε), we propose the following, derived from (1.13) and (2.13),

$$\widehat{Q}_\varepsilon^n(x) = \sum_{j=1}^n \widehat{F}_j(x) Y_j \qquad \text{where} \qquad \widehat{F}_j(x) = \frac{\exp\Big(\frac{(\widehat{u}_\varepsilon^n)^{c,\varepsilon}(Y_j) - c(x,Y_j)}{\varepsilon}\Big)}{\sum_{\ell=1}^n \exp\Big(\frac{(\widehat{u}_\varepsilon^n)^{c,\varepsilon}(Y_\ell) - c(x,Y_\ell)}{\varepsilon}\Big)}, \tag{2.14}$$

that is obtained by computing the smooth conjugate $(\widehat{u}_\varepsilon^n)^{c,\varepsilon} \in \mathbb{R}^n$ of $\widehat{u}_\varepsilon^n$ in (2.13). Note that if one denotes by $((\widehat{u}_\varepsilon^n)^{c,\varepsilon})^{c,\varepsilon}(x)$ the smooth conjugate of $(\widehat{u}_\varepsilon^n)^{c,\varepsilon}$ at x, then our estimator can also be expressed as

$$\widehat{Q}_\varepsilon^n(x) = x - \nabla((\widehat{u}_\varepsilon^n)^{c,\varepsilon})^{c,\varepsilon}(x).$$

Immediately, our regularized estimator $\widehat{Q}_\varepsilon^n$ is a smooth and monotone (see *e.g.* [34][Lemma 1]) estimator of the MK quantile function

2.3 Main results

2.3.1 Convergence of our algorithm

Hereafter, it is assumed that $\varepsilon > 0$. Moreover, the convergence of our algorithm, as stated below, is valid for any cost function c that is lower semi-continuous. Consequently, the restriction to the quadratic cost is not needed.

In order to state our main results, it is necessary to introduce two suitable assumptions related to the optimal sequence of Fourier coefficients $\theta^\varepsilon = (\theta_\lambda^\varepsilon)_{\lambda\in\Lambda}$ and the second order Fréchet derivative of the function H_ε given by (2.5).

Assumption 2.1. *The sequence $(\theta_\lambda^\varepsilon)_{\lambda\in\Lambda}$ satisfies $\|\theta^\varepsilon\|_{W^{-1}} < +\infty$.*

Assumption 2.2. *For any regularization parameter $\varepsilon > 0$, there exists a positive constant c_ε such that the second order Fréchet derivative of the function H_ε evaluated at the optimal value θ^ε satisfies, for any $\tau \in \ell_1(\Lambda)$,*

$$D^2 H_\varepsilon(\theta^\varepsilon)[\tau,\tau] \geq c_\varepsilon \|\tau\|_{\ell_2}^2. \tag{2.15}$$

Our main theoretical result is devoted to the almost sure convergence of the random sequence $(\widehat{\theta}_n)_n$ defined by (2.8). For the sake of readability, its proof is deferred to Section 2.5.

Theorem 2.4. *Suppose that the initial value $\widehat{\theta}_0$ is any random element in $\ell_2(\Lambda)$ such that $\|\widehat{\theta}_0\|_{W^{-1}} < +\infty$. Then, under Assumptions 2.1 and 2.2, the sequence $(\widehat{\theta}_n)$ converges almost surely in ℓ_2 towards the solution θ^ε of the stochastic convex*

minimisation problem (2.5)*, i.e.*

$$\lim_{n\to\infty} \|\widehat{\theta}_n - \theta^\varepsilon\|_{\ell_2} = 0 \qquad a.s. \tag{2.16}$$

Equivalently, we also have that

$$\lim_{n\to\infty} \int_{\mathcal{X}} |\widehat{u}_\varepsilon^n(x) - u_\varepsilon(x)|^2 d\mu(x) = 0 \qquad a.s. \tag{2.17}$$

Assumption 2.1 can be made more explicit by the choice of a specific sequence of weights $w = (w_\lambda)_{\lambda\in\Lambda}$ and by imposing regularity assumptions on the function $u_\varepsilon \in L^1(\mu)$ given by (2.12). For example, one may assume in dimension $d = 2$ that u_ε is differentiable (with periodic conditions on the boundary on $\mathcal{X}$) and that its gradient is square integrable,

$$\int_{\mathcal{X}} \|\nabla u(x)\|^2 d\mu(x) < +\infty.$$

Then, under such assumptions, one may use the fact that $\nabla u(x) = \sum_{\lambda\in\Lambda} 2\pi i\lambda\theta_\lambda^\varepsilon \phi_\lambda(x)$ and Parseval's identity, [164][Theorem 1.7], to obtain that

$$\sum_{\lambda\in\Lambda} \|\lambda\|^2 |\theta_\lambda^\varepsilon|^2 < +\infty.$$

Consequently, for the specific choice $w_\lambda = \|\lambda\|^{-2}$, we find that Assumption 2.1 holds properly. In higher dimension d, it is necessary to make additional assumptions on the differentiability of u_ε. Note that we shall also prove in Lemma 2.10 that for any $\theta, \tau \in \overline{\ell_1}(\Lambda)$,

$$D^2 H_\varepsilon(\theta^\varepsilon)[\tau,\tau] \geq \frac{1}{\varepsilon}\left(2 - \int_{\mathcal{Y}}\int_{\mathcal{X}} F^2_{\theta^\varepsilon,y}(x) d\mu(x) d\nu(y)\right) \|\tau\|^2_{\ell_2}.$$

Therefore a sufficient condition for Assumption 2.2 to hold is to assume that

$$\int_{\mathcal{Y}}\int_{\mathcal{X}} F^2_{\theta^\varepsilon,y}(x) d\mu(x) d\nu(y) < 2 \quad \text{with} \quad c_\varepsilon = \frac{1}{\varepsilon}\left(2 - \int_{\mathcal{Y}}\int_{\mathcal{X}} F^2_{\theta^\varepsilon,y}(x) d\mu(x) d\nu(y)\right).$$

2.3.2 Properties of the objective function H_ε

The purpose of this section is to discuss various keystone properties of the functions h_ε and H_ε that are needed to establish our main result on the convergence of our stochastic algorithm $\widehat{\theta}_n$.

Let us first discuss the first and second order Fréchet differentiability of the functions H_ε and h_ε that are functions from the Banach space $(\bar{\ell}_1(\Lambda), \|\cdot\|_{\ell_1})$ to $\mathbb{R}$. The following proposition gives the expression of the first order Fréchet derivative, that we shall sometimes refer to as the gradient, of h_ε and H_ε, as well as upper bounds on their operator norm.

Proposition 2.5. *For any $y \in \mathcal{Y}$, the first order Fréchet derivative of the function $h_\varepsilon(\cdot, y)$ at $\theta \in \bar{\ell}_1(\Lambda)$ is the linear operator $D_\theta h_\varepsilon(\theta, y) : \bar{\ell}_1(\Lambda) \to \mathbb{R}$ defined for any $\tau \in \bar{\ell}_1(\Lambda)$ as*

$$D_\theta h_\varepsilon(\theta, y)[\tau] = \sum_{\lambda \in \Lambda} \overline{\frac{\partial h_\varepsilon(\theta, y)}{\partial \theta_\lambda}} \tau_\lambda \tag{2.18}$$

where

$$\frac{\partial h_\varepsilon(\theta, y)}{\partial \theta_\lambda} = \int_{\mathcal{X}} \overline{\phi_\lambda(x)} F_{\theta,y}(x) d\mu(x). \tag{2.19}$$

Moreover, the linear operator $D_\theta h_\varepsilon(\theta, y)$ can be identified as an element of $\bar{\ell}_\infty(\Lambda)$ and its operator norm satisfies, for any $\theta \in \bar{\ell}_1(\Lambda)$ and $y \in \mathcal{Y}$,

$$\|D_\theta h_\varepsilon(\theta, y)\|_{op} = \sup_{\|\tau\|_{\ell_1} \leq 1} |D_\theta h_\varepsilon(\theta, y)[\tau]| \leq \sup_{\lambda \in \Lambda} \left| \frac{\partial h_\varepsilon(\theta, y)}{\partial \theta_\lambda} \right| \leq 1. \tag{2.20}$$

The first order Fréchet derivative of the function H_ε at $\theta \in \bar{\ell}_1(\Lambda)$ is the linear operator $DH_\varepsilon(\theta) : \bar{\ell}_1(\Lambda) \to \mathbb{R}$ defined for any $\tau \in \bar{\ell}_1(\Lambda)$ as

$$DH_\varepsilon(\theta)[\tau] = \sum_{\lambda \in \Lambda} \overline{\frac{\partial H_\varepsilon(\theta)}{\partial \theta_\lambda}} \tau_\lambda \tag{2.21}$$

where

$$\frac{\partial H_\varepsilon(\theta)}{\partial \theta_\lambda} = \int_{\mathcal{Y}} \frac{\partial h_\varepsilon(\theta, y)}{\partial \theta_\lambda} d\nu(y).$$

Moreover, the operator norm of the linear operator $DH_\varepsilon(\theta)$ satisfies, for any $\theta \in \bar{\ell}_1(\Lambda)$,

$$\|DH_\varepsilon(\theta)\|_{op} = \sup_{\|\tau\|_{\ell_1} \leq 1} |DH_\varepsilon(\theta)[\tau]| \leq \sup_{\lambda \in \Lambda} \left| \frac{\partial H_\varepsilon(\theta)}{\partial \theta_\lambda} \right| \leq 1. \tag{2.22}$$

The proposition below gives the expression of the second order Fréchet derivative, that we shall sometimes refer to as the Hessian, of h_ε and H_ε and upper bounds on their operator norm.

Proposition 2.6. *For any $y \in \mathcal{Y}$, the second order Fréchet derivative of the function $h_\varepsilon(\cdot, y)$ at $\theta \in \bar{\ell}_1(\Lambda)$ is the following symmetric bilinear mapping from $\bar{\ell}_1(\Lambda) \times \bar{\ell}_1(\Lambda)$ to $\mathbb{R}$*

$$\begin{aligned} D^2_\theta h_\varepsilon(\theta, y)[\tau, \tau'] = {} & \frac{1}{\varepsilon} \sum_{\lambda' \in \Lambda} \sum_{\lambda \in \Lambda} \tau'_{\lambda'} \overline{\tau_\lambda} \int_{\mathcal{X}} \phi_{\lambda'}(x) \overline{\phi_\lambda(x)} F_{\theta,y}(x) d\mu(x) \\ & - \frac{1}{\varepsilon} \left(\sum_{\lambda \in \Lambda} \tau'_\lambda \int_{\mathcal{X}} \phi_\lambda(x) F_{\theta,y}(x) d\mu(x) \right) \overline{\left(\sum_{\lambda \in \Lambda} \tau_\lambda \int_{\mathcal{X}} \phi_{\lambda(x)} F_{\theta,y}(x) d\mu(x) \right)}. \end{aligned} \tag{2.23}$$

and its operator norm satisfies, for any $\theta \in \bar{\ell}_1(\Lambda)$ and $y \in \mathcal{Y}$,

$$\|D^2_\theta h_\varepsilon(\theta, y)\|_{op} = \sup_{\|\tau\|_{\ell_1} \leq 1, \|\tau'\|_{\ell_1} \leq 1} |D^2_\theta h_\varepsilon(\theta, y)[\tau, \tau']| \leq \frac{1}{\varepsilon}. \tag{2.24}$$

Moreover, the second order Fréchet derivative of $H_\varepsilon : \bar{\ell}_1(\Lambda) \to \mathbb{R}$ *is the symmetric bilinear mapping from* $\bar{\ell}_1(\Lambda) \times \bar{\ell}_1(\Lambda)$ *to* $\mathbb{R}$ *defined by*

$$D^2 H_\varepsilon(\theta)[\tau, \tau'] = \int D^2_\theta h_\varepsilon(\theta, y)[\tau, \tau'] d\nu(y). \tag{2.25}$$

and its operator norm satisfies, for any $\theta \in \bar{\ell}_1(\Lambda)$,

$$\|D^2 H_\varepsilon(\theta)\|_{op} = \sup_{\|\tau\|_{\ell_1} \leq 1, \|\tau'\|_{\ell_1} \leq 1} |D^2 H_\varepsilon(\theta)[\tau, \tau']| \leq \frac{1}{\varepsilon}. \tag{2.26}$$

We now provide useful results on the regularity of H_ε.

Proposition 2.7. *For any* $y \in \mathcal{Y}$, *the functions* $h_\varepsilon(\cdot, y)$ *and* H_ε *are strictly convex on* $\bar{\ell}_1(\Lambda)$.

As already noticed in previous works [17, 71] dealing with related objective functions, the function H_ε is not strongly convex. Nevertheless, one can obtain a local strong convexity property of the function H_ε in the neighborhood of its minimizer θ^ε. This result is a consequence of the notion of generalized self-concordance introduced in [8], that has been shown to hold for semi-discrete EOT in [17], and which we extend to the setting of the functional H_ε on the Banach space $\bar{\ell}_1(\Lambda)$.

Proposition 2.8. *For all* $\theta \in \bar{\ell}_1(\Lambda)$, *we have*

$$H_\varepsilon(\theta) - H_\varepsilon(\theta^\varepsilon) \leq \frac{1}{\varepsilon} \|\theta - \theta^\varepsilon\|^2_{\ell_1}. \tag{2.27}$$

Moreover, for any $\theta \in \overline{\ell_1}(\Lambda)$, *the following local strong convexity property holds*

$$DH_\varepsilon(\theta)[\theta - \theta^\varepsilon] \geq g\Big(\frac{2}{\varepsilon}\|\theta - \theta^\varepsilon\|_{\ell_1}\Big) D^2 H_\varepsilon(\theta^\varepsilon)[\theta - \theta^\varepsilon, \theta - \theta^\varepsilon], \tag{2.28}$$

where, for all $x > 0$,

$$g(x) = \frac{1 - \exp(-x)}{x}. \tag{2.29}$$

2.4 Numerical experiments

Influence of the dimension d

We first investigate the convergence of our numerical scheme for the estimation of the entropic map using various values of the dimension d.

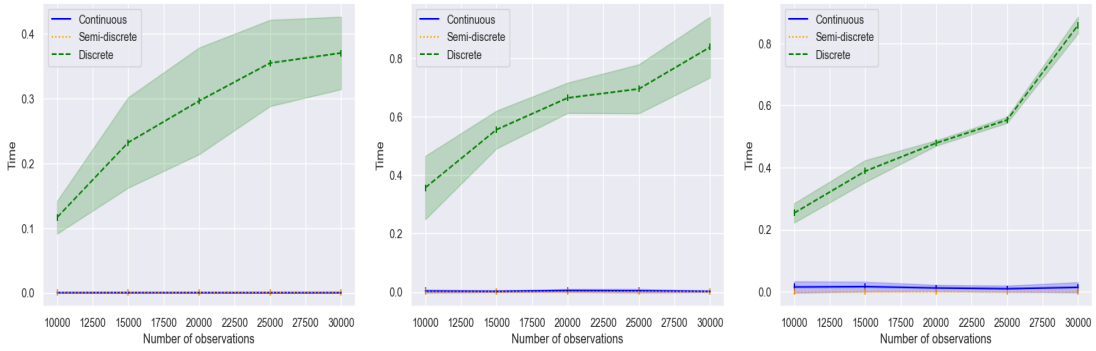
To do so, our estimator $\widehat{Q}^n_\varepsilon(x)$ in (2.14) is compared to $\widetilde{Q}^n_\varepsilon(x)$ in (2.2) where the dual potential $\widetilde{v}_n \in \mathbb{R}^n$ needed to compute $\widetilde{Q}^n_\varepsilon(x)$ is obtained with either the Sinkhorn algorithm [47] or a stochastic algorithm as proposed in [17, 71]. Starting from the uniform distribution on $[0, 1]^d$, we consider the map $Q : x \mapsto L^T L x + b$ where L is a lower triangular matrix and $b \in \mathbb{R}^d$, both filled with ones. Trivially,

Q is the gradient of a convex function, so that it is the MK quantile function of $\nu = Q_{\#}\mu$. Thus, by Monte-Carlo sampling, we are able to approximate the mean squared error of any estimator $\widehat{Q}$ defined as

$$\mathrm{MSE}(\widehat{Q}) = \mathbb{E}\left[\|\widehat{Q}(X) - Q(X)\|^2\right]. \tag{2.30}$$

The three ways of estimating Q are based on iterative schemes that we let running until convergence of the MSE below the value 10^{-2} for $d = 2, 3, 4$, and by taking $\varepsilon = 0.005$.

Figure 2.1 illustrates the time before convergence, in seconds, as a function of n (the number of observations). In what follows, the continuous, semi-discrete, and discrete approaches refer to Algorithm 1 with W the identity matrix, the stochastic algorithm from [17, 71], and the Sinkhorn algorithm [47] respectively. For $\mathcal{X}_p$ given in Algorithm 1, the uniform distribution on $\mathcal{X}_p$ is taken as a discrete reference measure for the Sinkhorn algorithm to ensure a fair comparison with our algorithm. The MSE is estimated through $m = 500$ other random samples from μ. The size p of the grid $\mathcal{X}_p$ is maintained comparable in every considered dimensions. Results are averaged over 10 experiments for several samples $(Y_1, \cdots, Y_n)$, and standard deviation is indicated around each MSE curve. Overall, these numerical experiments reveal a potentially faster convergence for approaches based on stochastic algorithms when the number of observations grows.



(a) $d = 2$, $p = 20^2$. (b) $d = 3$, $p = 10^3$. (c) $d = 4$, $p = 6^4$.

FIGURE 2.1 – Overall time, in seconds, until convergence of the MSE below 10^{-2} for different solvers for EOT.

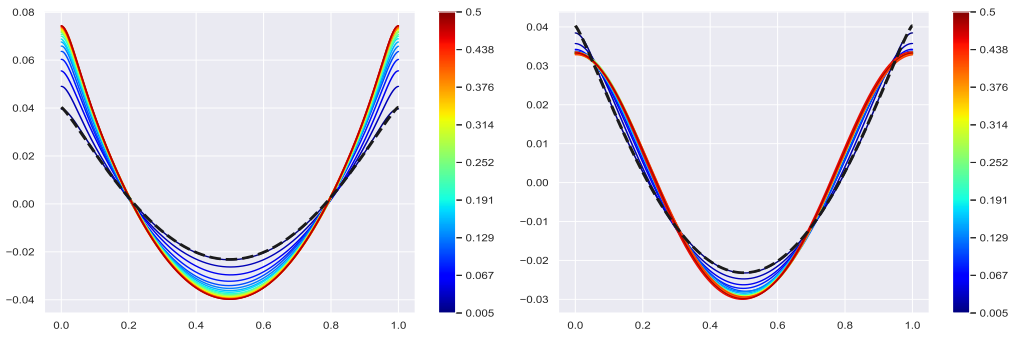
Numerical experiments in dimension $d = 1$

The univariate setting allows us an explicit knowledge of the ground truth Q. There, we study our algorithm with either the standard quadratic cost in $\mathbb{R}^d$ given by $c(x, y) = \frac{1}{2}\|x - y\|^2$ or the quadratic cost on the flat torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ that is

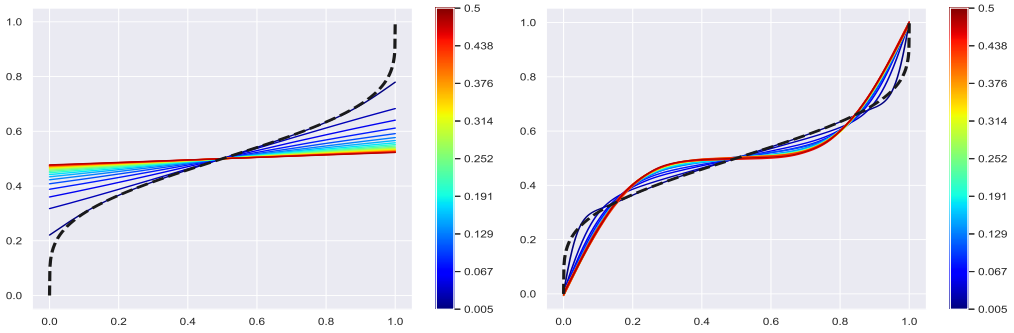
$$c(x, y) = \frac{1}{2} d_{\mathbb{T}^d}(x, y), \quad \text{with} \quad d_{\mathbb{T}^d}(x, y) = \min_{\lambda \in \mathbb{Z}^d} \|x - y + \lambda\|. \tag{2.31}$$

The choice of the quadratic cost on the torus is motivated by the discussion in the Appendix A.2 on sufficient conditions related to the summability of the Fourier coefficients of an optimal dual potential.

For the learning rate $\gamma_n = \gamma n^{-c}$, we took $\gamma = \varepsilon$ and $c = 3/4$. The sequence of weights $w = (w_\lambda)_{\lambda \in \Lambda}$ is chosen as $w_\lambda = |\lambda|^{-2}$ for $\lambda \in \mathbb{Z}\backslash\{0\}$. Taking a larger exposant than 2 results in smoother estimators of the optimal dual potential u_ε. For various values of $\varepsilon \in [0.005, 0.5]$, we consider a beta$(a, b)$ distribution ν on $\mathcal{Y} = [0, 1]$ with parameters $a = 5$ and $b = 5$. The optimal dual potential u_0 and quantile function Q_0 are straightforward to compute when $d = 1$ for the standard quadratic cost. For a sample of size $n = 10^5$, $\widehat{u}_\varepsilon^n$ and $\widehat{Q}_\varepsilon^n$ are displayed in Figure 2.2 using either the standard quadratic cost or the quadratic cost of the torus. One can observe that the choice of the cost yields a different regularization effect. Choosing $\varepsilon = 0.005$ yields values of $\widehat{u}_\varepsilon^n$ and $\widehat{Q}_\varepsilon^n$ that are very close to u_0 and Q_0 respectively.



(a) Standard quadratic cost (b) Quadratic cost on the torus



(c) Standard quadratic cost (d) Quadratic cost on the torus

FIGURE 2.2 – Estimators $\widehat{u}_\varepsilon^n$ and $\widehat{Q}_\varepsilon^n$ on the first and second lines respectively. The black and dashed curves are either the un-regularized optimal dual potential u_0 or the un-regularized quantile function Q_0 of the beta distribution for the standard quadratic cost.

From now on, let us consider a sample $(Y_1^*, \ldots, Y_J^*)$ of small size $J = 100$ of the same beta(a, b) distribution. We illustrate the potential benefits of using regularized OT to obtain a smoother estimator than the usual empirical quantile function $\widehat{Q}_0^J$

defined as the generalized inverse of the empirical cumulative distribution function

$$\widehat{F}_0^J(x) = \frac{1}{J}\sum_{j=1}^{J} \mathbb{1}_{\{Y_j^* \leq x\}}.$$

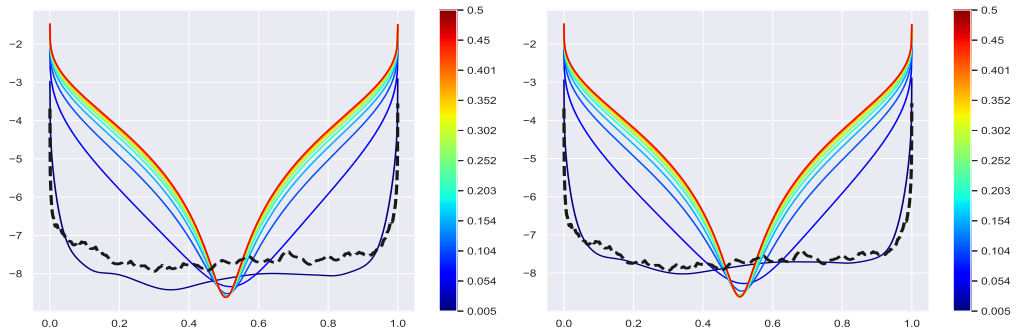
To this end, for various values of $\varepsilon \in [0.005, 0.5]$, we compute the two estimators $\widehat{Q}_\varepsilon^{n,J}$ from (2.14) and $\widetilde{Q}_\varepsilon^{m,J}$ from (2.2) with sequences of $n = m = 10^5$ random variables sampled from the discrete measure $\widehat{\nu}_J^*$ or the uniform measure on $[0,1]$ respectively. In Figure 2.3, we display in logarithmic scale the point-wise mean-squared errors

$$\mathrm{MSE}(\widehat{Q}_\varepsilon^{n,J}(x)) = \mathbb{E}\left[|\widehat{Q}_\varepsilon^{n,J}(x) - Q_0(x)|^2\right],$$

and

$$\mathrm{MSE}(\widetilde{Q}_\varepsilon^{m,J}(x)) = \mathbb{E}\left[|\widetilde{Q}_\varepsilon^{m,J}(x) - Q_0(x)|^2\right],$$

where the above expectations are approximated using Monte-Carlo experiments from 100 repetitions of the above described procedure. The MSE of these regularized estimators is then compared to the MSE of the usual empirical quantile function $\widehat{Q}_0^J$ defined accordingly. For all values of ε, it can be seen, from Figure 2.3, that regularization always improves the estimation of $Q_0(x)$ by $\widehat{Q}_0^J(x)$ around the median location $x = 0.5$. For the smallest values of ε, regularization also improves the estimation of $Q_0(x)$ for $x \in [0.1, 0.9]$, and the best results are obtained with the stochastic algorithm based on the FFT.



(a) $\log \mathrm{MSE}(\widehat{Q}_\varepsilon^{n,J}(x))$ (b) $\log \mathrm{MSE}(\widetilde{Q}_\varepsilon^{m,J}(x))$

FIGURE 2.3 – Point-wise error of the regularized estimators $\widehat{Q}_\varepsilon^{n,J}$ and $\widetilde{Q}_\varepsilon^{m,J}$ for various values of $\varepsilon \in [0.005, 0.5]$. The black and dashed curve is the point-wise error of the unregularized empirical quantile function $\widehat{Q}_0^J$.

Numerical experiments in dimension $d = 2$

As argued in [87], taking as reference the spherical uniform distribution μ_S on the unit ball $\mathbb{B}(0,1)$ induces different properties for MK quantiles. Thanks to a change in polar coordinates, one can parametrize on $\mathbb{B}(0,1)$ instead of $[0,1]^d$.

By definition, a random vector X with spherical uniform distribution is given by $X = R\Phi$ where R and Φ are independent and drawn uniformly from [0,1] and the unit hypersphere $\mathbb{S}^{d-1}$, respectively. In dimension $d = 2$, X writes in polar coordinates as

$$X = \begin{pmatrix} R\cos(2\pi\Psi) \\ R\sin(2\pi\Psi) \end{pmatrix} \in \mathbb{B}(0,1),$$

where (R, Ψ) is uniform on $[0,1]^2$. Then, for a function $u \in L^1(\mathbb{B}(0,1), \mu_S)$, its parametrization in polar coordinates is given, for all $(r, \psi) \in [0,1] \times [0,1]$, by

$$\overline{u}(r,\psi) = u\begin{pmatrix} r\cos(2\pi\psi) \\ r\sin(2\pi\psi) \end{pmatrix}.$$

Hence, by definition of μ_S, the function $\overline{u}$ is an element of $L^1([0,1]^2, \mu)$ where μ is the uniform distribution on $\mathcal{X} = [0,1]^2$. Consequently, thanks to this re-parametrization, we propose to solve in the Fourier domain, for $\Lambda = \mathbb{Z}^2\backslash\{0\}$, the following regularized OT problem

$$\theta^\varepsilon = \underset{\theta\in\ell_1(\Lambda)}{\operatorname{argmax}}\, \overline{H}_\varepsilon(\theta) \qquad \text{with} \qquad \overline{H}_\varepsilon(\theta) = \mathbb{E}\left[\,\overline{h}_\varepsilon(\theta, Y)\right], \tag{2.32}$$

where $Y = (Y_1, Y_2) \in \mathbb{R}^d 2$ is a random vector with distribution ν, and $\overline{h}_\varepsilon$ is given by

$$\overline{h}_\varepsilon(\theta, y) = \varepsilon \log\Big(\int_{\mathcal{X}} \exp\left(\frac{\sum_{\lambda\in\Lambda}\theta_\lambda\phi_\lambda(r,\psi) - c_y(r,\psi)}{\varepsilon}\right) d\mu(r,\psi)\Big) + \varepsilon,$$

with $\phi_\lambda(r,\psi) = e^{2\pi i(\lambda_1 r + \lambda_2\psi)}$ for $\lambda = (\lambda_1, \lambda_2) \in \mathbb{Z}^2$, and c_y refers to the quadratic cost,

$$c_y(r,\psi) = \frac{1}{2}\left((r\cos(2\pi\psi) - y_1)^2 + (r\sin(2\pi\psi) - y_2)^2\right).$$

In order to solve (2.32), we adapt the stochastic algorithm (2.8) which yields, after n iterations, the sequence $\overline{\theta}_n$ and the estimator, in polar coordinates,

$$\overline{u}_\varepsilon^n(r,\psi) = \sum_{\lambda\in\Lambda}\overline{\theta}_{n,\lambda}\phi_\lambda(r,\psi). \tag{2.33}$$

In practice, we discretize $[0,1]^2$ by choosing equi-spaced radius points $0 \le r_1 < \ldots < r_{p_1} \le 1$ and angles $0 \le \psi_1 < \ldots \psi_{p_2} < 1$ which results in taking a grid of $p = p_1 p_2$ points

$$\mathcal{X}_p = \left\{(r_{\ell_1}, \psi_{\ell_2})_{(\ell_1,\ell_2)\in\{1,p_1\}\times\{1,p_2\}}\right\} \subset [0,1]^2.$$

Finally, the stochastic algorithm (2.8) is implemented on this polar grid using the weight sequence $w_\lambda = 1$ for all $\lambda \in \Lambda_p$ that is with $\alpha = 0$. Of course, Assumption 2.1 is always verified if $(w_\lambda) \equiv (1, 1, \cdots, 1, 0, 0, \cdots)$. This is motivated by the fact that

choosing $w_\lambda = \|\lambda\|^{-\alpha}$ with $\alpha \geq 1$ would impose periodic constraints on the dual potentials $\bar{u}(r,\psi)$ along the radius coordinate. However, as shown by the following numerical experiments, an optimal dual potential typically does not satisfy the polar periodic conditions $\bar{u}(0,\psi) = \bar{u}(1,\psi)$ for all $\psi \in [0,1]$. The counterpart of $\widehat{Q}_\varepsilon^n$ in (2.14) directly follows from (2.33), that is

$$\overline{Q}_\varepsilon^n(x) = \sum_{j=1}^n \overline{F}_j(x) Y_j \qquad \text{where} \qquad \overline{F}_j(x) = \frac{\exp\Big(\frac{(\overline{u}_\varepsilon^n)^{c,\varepsilon}(Y_j) - c(x,Y_j)}{\varepsilon}\Big)}{\sum_{\ell=1}^n \exp\Big(\frac{(\overline{u}_\varepsilon^n)^{c,\varepsilon}(Y_\ell) - c(x,Y_\ell)}{\varepsilon}\Big)}, \tag{2.34}$$

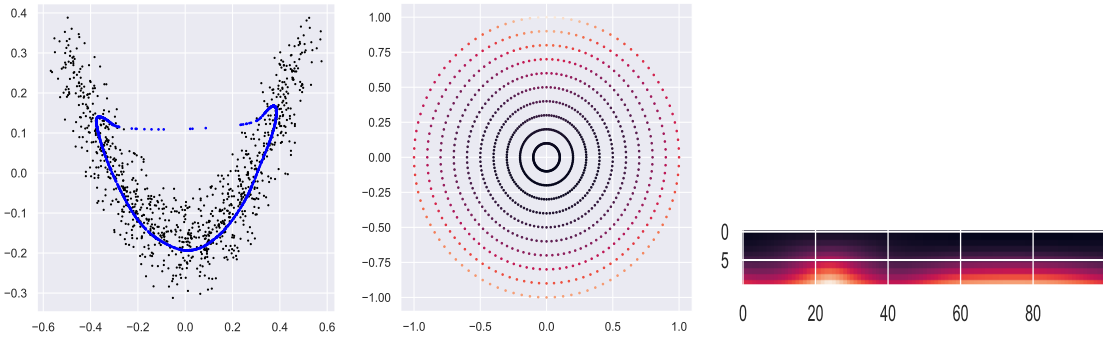
where the integral in the computation $(\overline{u}_\varepsilon^n)^{c,\varepsilon}(\cdot)$ is approximated with the polar grid $\mathcal{X}_p$. In what follows, we report numerical experiments for the banana-shaped distribution ν considered in [33]. It corresponds to sampling Y as the random vector

$$Y = \begin{pmatrix} U + R\cos(2\pi\Phi) \\ U^2 + R\sin(2\pi\Phi) \end{pmatrix},$$

where U is uniform on $[-1,1]$, Φ is uniform on $[0,1]$, $R = 0.2Z(1-(1-|U|)/2$ with Z uniform on $[0,1]$, and U, Φ and Z independent. In these simulations, the random variable Y is also centered and scaled so that it takes its values within the subset $[-0.6, 0.6] \times [-0.4, 0.5] \subset [0,1]^2$.

We first consider a sample $Y_1^*, \ldots, Y_J^*$ of size $J = 10^3$ that is held fixed and displayed in Figure 2.4. Then, we draw $n = 10^5$ random variables $Y_1, \ldots, Y_n$ from the associated discrete distribution $\widehat{\nu}_J^*$, and we run the stochastic algorithm (2.8) with different sizes $(p_1, p_2) = (10, 100)$ and $(p_1, p_2) = (100, 1000)$ for the discretization $\mathcal{X}_p$. Note that the cost of each iteration of the stochastic algorithm is of order $\mathcal{O}(p\log(p))$ for $p = p_1 p_2$. Therefore, the choice of discretization of the polar coordinates greatly influences the computational cost of the algorithm. In Figure 2.4, we display the resulting regularized dual potentials $\widehat{u}_\varepsilon^n$ in cartesian and polar coordinates for $\varepsilon = 0.005$. We also draw the resulting MK contour quantiles of level $r = 0.5$ for each choice of discretization, to illustrate that the results can be very similar with a much lowest computational cost for the discretization of size $(p_1, p_2) = (10, 100)$.

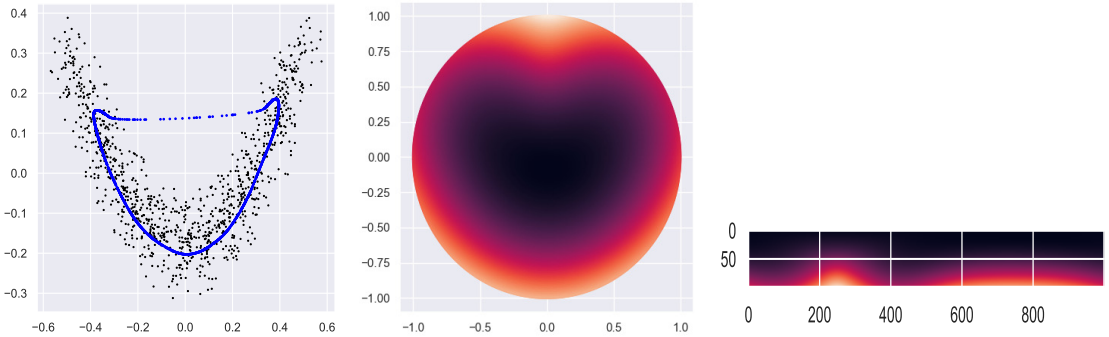
Figure 2.5 contains a comparison of the convergence between our FFT-based scheme (2.14) and (2.2), based on the stochastic gradient descent from [17], that we refer to as the regularized SGD. The reference distribution is taken to be the spherical uniform. Also, we compare these regularized approaches (using $\varepsilon = 0.005$) with classical un-regularized ones. To this end, we implement a subgradient descent for semi-discrete unregularized OT, namely the same Robbins-Monro scheme as (2.1) with $\varepsilon = 0$, and rely on (**??**). Finally, we use the OT network simplex solver from the Python library [64] to compute the solution of un-regularized OT between two empirical discrete distributions with supports $\mathcal{X}_p = \{x_1, \ldots, x_p\}$ and $(Y_1^*, \ldots, Y_J^*)$. We first consider a sample $Y_1^*, \ldots, Y_J^*$ of size $J = 10^4$ that is held fixed. For our



(a) Quantile contour of order $r = 0.5$; $(p_1, p_2) = (10, 100)$

(b) $\widehat{u}_\varepsilon^n(r\cos(2\pi\psi), r\sin(2\pi\psi))$; $(p_1, p_2) = (10, 100)$

(c) $\overline{u}_\varepsilon^n(r, \psi)$; $(p_1, p_2) = (10, 100)$



(d) Quantile contour of order $r = 0.5$; $(p_1, p_2) = (100, 1000)$

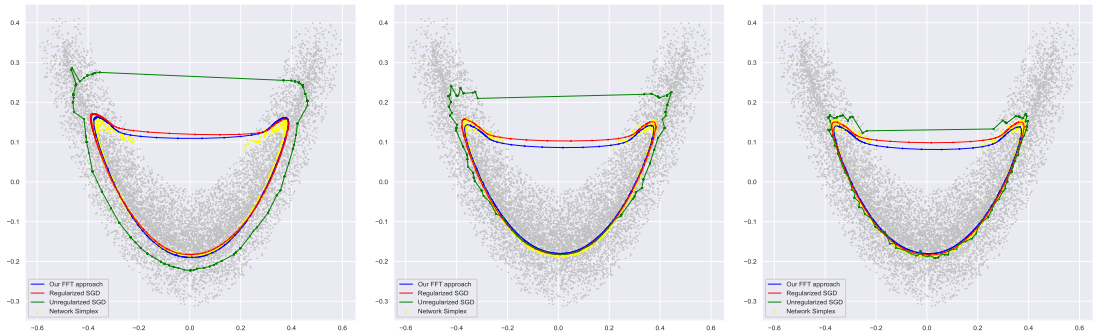
(e) $\widehat{u}_\varepsilon^n(r\cos(2\pi\psi), r\sin(2\pi\psi))$; $(p_1, p_2) = (100, 1000)$

(f) $\overline{u}_\varepsilon^n(r, \psi)$; $(p_1, p_2) = (100, 1000)$

FIGURE 2.4 – The blue curves are regularized MK quantile contours at level $r = 0.5$ for $\varepsilon = 0.005$ from the discrete measure $\hat{\nu}_J$ (displayed with black points) using two different discretizations $(p_1, p_2) = (10, 100)$ (first row) and $(p_1, p_2) = (100, 1000)$ (second row). The second (resp. thrid) columns represent the values of the regularized dual potentials in cartesian (resp. polar) coordinates.

FFT approach, we let $p_1 = 20$, $p_2 = 500$, so that $p = p_1 p_2 = 10^4$. For the three iterative schemes, the number of iterations varies between 10^4, 10^5 and 10^6. This corresponds, for our FFT approach, to a stochastic algorithm with $1, 10$ and 100 epochs, whereas the other approaches sample from the reference distribution μ_S.

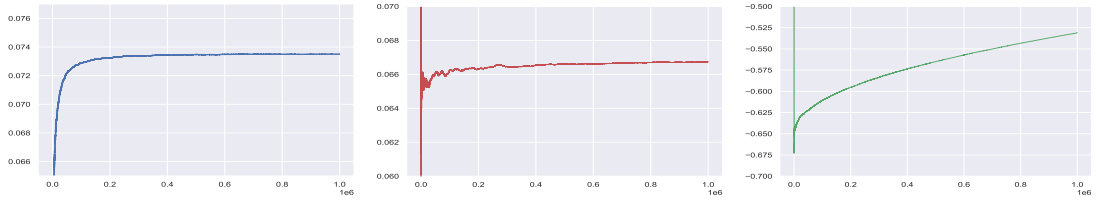
The first line of Figure 2.5 contains the corresponding quantile contours of order $r = 0.5$ for each method, for several number of iterations. Unlike regularized estimators, the values of un-regularized quantiles are restricted to $(Y_1^*, \ldots, Y_J^*)$. The second line of Figure 2.5 deals with convergence depending on the number of iterations. As customary, we consider a recursive estimation of the values of our objectives, respectively H_ε for (2.5), $\widetilde{H}_\varepsilon$ for (2.1) and $\widetilde{H}_0$ for (2.1) with $\varepsilon = 0$. These objectives are recursively estimated along the iterations by gradual averaging in order to account for convergence, as in [17]. For $J = p = 10^4$, the computational



(a) Quantile contour at level $r = 0.5$ with $n = 10^4$ iterations

(b) Quantile contour at level $r = 0.5$ with $n = 10^5$ iterations

(c) Quantile contour at level $r = 0.5$ with $n = 10^6$ iterations



(d) Convergence of our FFT-based scheme

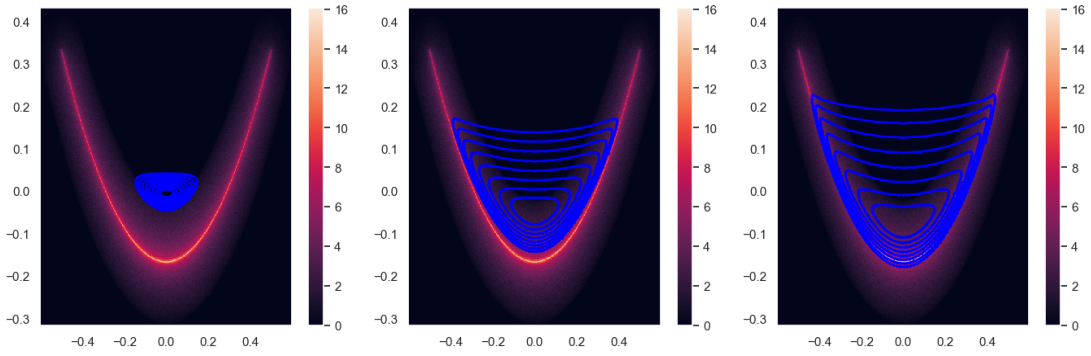
(e) Convergence of the regularized SGD

(f) Convergence of the unregularized SGD

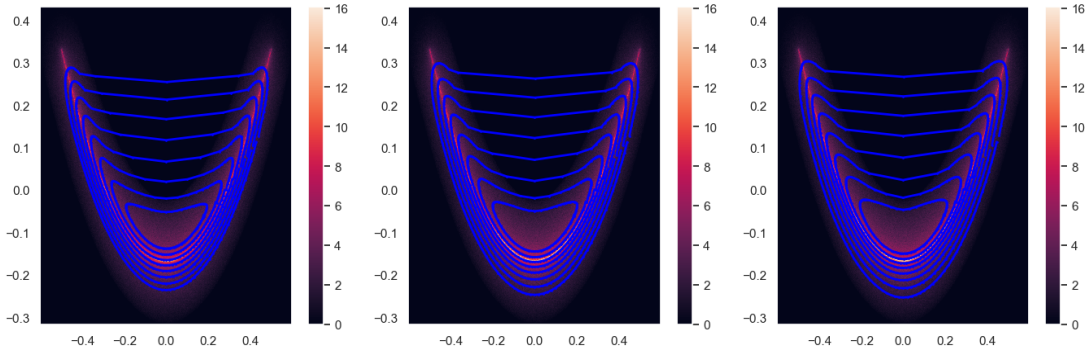
FIGURE 2.5 – Comparison between regularized (with $\varepsilon = 0.005$) and unregularized approaches.

cost at each iteration of the two regularized procedures is of the same order. It can be seen that the un-regularized SGD has not converged with 10^6 iterations, whereas the regularized approaches (2.14) and (2.2) have similar convergence behavior. Together with the first line of Figure 2.5, these results illustrate that entropically regularized methods converge faster towards a more suitable solution.

We finally propose a last numerical experiment to highlight the behavior of EOT when varying the regularization parameter ϵ. We chose to draw $n = 10^7$ random variables $Y_1, \dots, Y_n$ from the banana-shaped distribution, and we ran the stochastic algorithm (2.8) for the discretization $(p_1, p_2) = (10, 1000)$. Doing so, the obtained sample is very close to the true density, and the various resulting contours only depend on ε. In Figure 2.6, we display the resulting regularized MK quantile contours of levels $r \in \{0.2, 0.3, \dots, 1\}$ for different values of $\varepsilon \in [0.002, 0.5]$. This visualization warns on the choice of the regularization parameter that must be chosen small enough. Note that, for $n = 10^7$ observations, we have not been able to implement the Sinkhorn algorithm. Moreover, the cost at each iteration of either regularized or un-regularized SGD being $\mathcal{O}(n)$, these algorithms are much slower to converge than our approach.



(a) $\varepsilon = 0.5$ (b) $\varepsilon = 0.1$ (c) $\varepsilon = 0.05$



(d) $\varepsilon = 0.01$ (e) $\varepsilon = 0.005$ (f) $\varepsilon = 0.002$

FIGURE 2.6 – In the background, a density histogram from the empirical measure $\hat{\nu}_n = \frac{1}{n}\sum_{j=1}^{n} \delta_{Y_j}$ sampled from the banana-shaped distribution with $n = 10^7$. The blue curves correspond to regularized MK quantile contours of levels $r \in \{0.2, 0.3, \ldots, 1\}$ for $\varepsilon \in [0.002, 0.5]$.

2.5 Proofs of the main results

The proofs of Proposition 2.5, Proposition 2.6 and Proposition 2.7 are given in the Appendix, see Sections A.1.1, A.1.2 and A.1.3. We shall now proceed to the proofs of the main results from Section 2.3.

Proof of Proposition 2.8

For $\theta \in \overline{\ell_1}(\Lambda)$ and $t \in [0,1]$, we denote $\theta_t = \theta^\varepsilon + t(\theta - \theta^\varepsilon)$ and we define the function $\varphi(t) = H_\varepsilon(\theta_t)$. Then, we deduce from a second order Taylor expansion of φ with integral remainder that

$$\varphi(1) = \varphi(0) + \varphi'(0) + \int_0^1 (1-t)\varphi''(t)dt. \tag{2.35}$$

However, we clearly have

$$\varphi'(t) = DH_\varepsilon(\theta_t)[\theta - \theta^\varepsilon] \qquad \text{and} \qquad \varphi''(t) = D^2 H_\varepsilon(\theta_t)[\theta - \theta^\varepsilon, \theta - \theta^\varepsilon]. \tag{2.36}$$

Consequently, as $\varphi'(0) = DH_\varepsilon(\theta^\varepsilon)[\theta - \theta^\varepsilon] = 0$, (2.35) can be rewritten as

$$H_\varepsilon(\theta) - H_\varepsilon(\theta^\varepsilon) = \int_0^1 (1-t) D^2 H_\varepsilon(\theta_t)[\theta - \theta^\varepsilon, \theta - \theta^\varepsilon] dt. \tag{2.37}$$

Therefore, (2.27) immediately follows from (2.26) and (2.37). It only remains to prove (2.28). Our strategy is to adapt to the setting of this chapter the notion of self-concordance as introduced in [7, 8] and used in [17, 22] to study the statistical properties of stochastic optimal transport.

Lemma 2.9. *For $\theta \in \overline{\ell_1}(\Lambda)$ and for all $0 < t < 1$, denote $\theta_t = \theta^\varepsilon + t(\theta - \theta^\varepsilon)$. Then, the function $\varphi(t) = H_\varepsilon(\theta_t)$ verifies the self-concordance property*

$$|\varphi'''(t)| \leq \frac{2}{\varepsilon} \|\theta - \theta^\varepsilon\|_{\ell_1} \varphi''(t). \tag{2.38}$$

Proof. For a fixed $y \in \mathcal{Y}$, let $\phi(t) = h_\varepsilon(\theta_t, y)$. Firstly, we show that $\phi(t)$ verifies the self-concordance property. From the chain rule, we obtain that

$$\begin{aligned}
\phi'(t) &= Dh_\varepsilon(\theta_t, y)[\theta - \theta^\varepsilon], \\
\phi''(t) &= D^2 h_\varepsilon(\theta_t, y)[\theta - \theta^\varepsilon, \theta - \theta^\varepsilon], \\
\phi'''(t) &= D^3 h_\varepsilon(\theta_t, y)[\theta - \theta^\varepsilon, \theta - \theta^\varepsilon, \theta - \theta^\varepsilon],
\end{aligned}$$

where $D^3 h_\varepsilon$ denotes the third order Fréchet derivative of $h_\varepsilon(\cdot, y)$. It follows from (2.18) that

$$\phi'(t) = \int_{\mathcal{X}} S(x) F_{\theta_t, y}(x) d\mu(x) \qquad \text{where} \qquad S(x) = \sum_{\lambda \in \Lambda} (\theta_\lambda - \theta^\varepsilon_\lambda) \phi_\lambda(x). \tag{2.39}$$

Similarly, (2.23) yields

$$\varepsilon \phi''(t) = \int_{\mathcal{X}} S(x)^2 F_{\theta_t, y}(x) d\mu(x) - \left(\int_{\mathcal{X}} S(x) F_{\theta_t, y}(x) d\mu(x) \right)^2. \tag{2.40}$$

Hereafter, denoting by Z_t the random variable with density $F_{\theta_t, y}$ with respect to μ, it appears that $\varepsilon \phi''(t) = \mathbb{E}[S(Z_t)^2] - \mathbb{E}[S(Z_t)]^2 = \mathbb{E}[(S(Z_t) - \mathbb{E}[S(Z_t)])^2]$, that is

$$\varepsilon \phi''(t) = \int_{\mathcal{X}} \left(S(x) - \int_{\mathcal{X}} S(z) F_{\theta_t, y}(z) d\mu(z) \right)^2 F_{\theta_t, y}(x) d\mu(x). \tag{2.41}$$

Furthermore, using (A.6) in the derivation of (2.40), we have that

$$\begin{aligned}
\varepsilon \phi'''(t) = &\int_{\mathcal{X}} S(x)^2 \frac{d}{dt} F_{\theta_t, y}(x) d\mu(x) \\
&- 2 \left(\int_{\mathcal{X}} S(x) \frac{d}{dt} F_{\theta_t, y}(x) d\mu(x) \right) \int_{\mathcal{X}} S(x) F_{\theta_t, y}(x) d\mu(x),
\end{aligned}$$

which yields

$$\varepsilon^2\phi'''(t) = \int S^3(x)F_{\theta_t,y}(x)d\mu(x) - \int S^2(x)F_{\theta_t,y}(x)d\mu(x)\int S(z)F_{\theta_t,y}(z)d\mu(z)$$
$$- 2\int S(x)F_{\theta_t,y}(x)d\mu(x)\left[\int S^2(x)F_{\theta_t,y}(x)d\mu(x) - \left(\int S(x)F_{\theta_t,y}(x)d\mu(x)\right)^2\right].$$

Consequently,

$$\varepsilon^2\phi'''(t) = m_3 - m_2m_1 - 2m_1(m_2 - m_1^2) = m_3 - 3m_2m_1 + 2m_1^3,$$

where m_i stands for the i-th moment of the distribution of $S(Z_t)$. Then, one recognizes the formula for the cumulant of order 3 of a random variable, and so the above equality can be factorized as

$$\varepsilon^2\phi'''(t) = \mathbb{E}[(S(Z_t) - m_1)^3] = \int (S(x) - m_1)^3 F_{\theta_t,y}(x)d\mu(x). \tag{2.42}$$

Thanks to the connection between $\varepsilon\phi''(t)$ and the variance term in (2.41), the above (2.42) leads to

$$\varepsilon|\phi'''(t)| \leq \sup_{x\in\mathcal{X}} |S(x) - m_1|\phi''(t).$$

It is easy to see that $|S(x) - m_1| \leq |S(x)| + |m_1| \leq 2\|\theta - \theta^\varepsilon\|_{\ell_1}$. Hence

$$|\phi'''(t)| \leq \frac{2}{\varepsilon}\|\theta - \theta^\varepsilon\|_{\ell_1}\phi''(t). \tag{2.43}$$

Finally, given that $\varphi(t) = H_\varepsilon(\theta_t) = \int_{\mathcal{Y}} h_\varepsilon(\theta_t, y)d\nu(y) = \int_{\mathcal{Y}} \phi(t)d\nu(y)$, (2.43) induces the self-concordance property of φ. □

We are now in a position to prove inequality (2.28). Denote $\delta = 2\|\theta-\theta^\varepsilon\|_{\ell_1}/\varepsilon$. It follows from inequality (2.38) that, for all $0 < t < 1$, $|\varphi'''(t)| \leq \delta\varphi''(t)$, which leads to $\frac{\varphi'''(t)}{\varphi''(t)} \geq -\delta$. By integrating the above inequality between 0 and t, we obtain that $\log\varphi''(t) - \log\varphi''(0) \geq -\delta t$, which means that $\frac{\varphi''(t)}{\varphi''(0)} \geq e^{-\delta t}$. Integrating once again the previous inequality between 0 and 1, we obtain that

$$\varphi'(1) - \varphi'(0) \geq \left(\frac{1 - e^{-\delta}}{\delta}\right)\varphi''(0). \tag{2.44}$$

Finally, as $\varphi'(1) = DH_\varepsilon(\theta)(\theta - \theta^\varepsilon)$, $\varphi'(0) = 0$ and $\varphi''(0) = D^2H_\varepsilon(\theta^\varepsilon)[\theta - \theta^\varepsilon, \theta - \theta^\varepsilon]$, inequality (2.28) holds, which completes the proof of Proposition 2.8. □

A sufficient condition for Assumption 2.2

Lemma 2.10. *For any* $\tau \in \overline{\ell_1}(\Lambda)$,

$$D^2 H_\varepsilon(\theta^\varepsilon)[\tau,\tau] \geq \frac{1}{\varepsilon}\left(2 - \int_{\mathcal{Y}}\int_{\mathcal{X}} F^2_{\theta^\varepsilon,y}(x)d\mu(x)d\nu(y)\right)\|\tau\|^2_{\ell_2}. \tag{2.45}$$

Proof. We already saw from (2.25) that for any $\tau \in \overline{\ell_1}(\Lambda)$,

$$\begin{aligned} D^2 H_\varepsilon(\theta^\varepsilon)[\tau,\tau] &= \frac{1}{\varepsilon}\sum_{\lambda'\in\Lambda}\sum_{\lambda\in\Lambda}\tau_{\lambda'}\overline{\tau_\lambda}\int_{\mathcal{Y}}\int_{\mathcal{X}}\phi_{\lambda'}(x)\overline{\phi_\lambda(x)}F_{\theta^\varepsilon,y}(x)d\mu(x)d\nu(y) \\ &- \frac{1}{\varepsilon}\int_{\mathcal{Y}}\left|\sum_{\lambda\in\Lambda}\tau_\lambda\int_{\mathcal{X}}\phi_\lambda(x)F_{\theta^\varepsilon,y}(x)d\mu(x)\right|^2 d\nu(y). \end{aligned} \tag{2.46}$$

Our proof consists in a study of the two terms in the right-hand side of (2.46). Since (2.21) only defines $DH_\varepsilon(\theta^\varepsilon)_\lambda$ for all $\lambda \neq 0$, we deduce from (2.19) and (2.21) that, for all $\lambda \neq \lambda'$,

$$\int_{\mathcal{X}}\phi_{\lambda'}(x)\overline{\phi_\lambda(x)}F_{\theta^\varepsilon,y}(x)d\mu(x)d\nu(y) = \int_{\mathcal{Y}}\int_{\mathcal{X}}\phi_{\lambda'-\lambda}(x)F_{\theta^\varepsilon,y}(x)d\mu(x)d\nu(y) = DH_\varepsilon(\theta^\varepsilon)_{\lambda'-\lambda}.$$

Moreover, as soon as $\lambda = \lambda'$,

$$\int_{\mathcal{Y}}\int_{\mathcal{X}}\phi_\lambda(x)\overline{\phi_\lambda(x)}F_{\theta^\varepsilon,y}(x)d\mu(x)d\nu(y) = \int_{\mathcal{Y}}\int_{\mathcal{X}}F_{\theta^\varepsilon,y}(x)d\mu(x)d\nu(y) = 1.$$

Hence, from the optimality condition $DH_\varepsilon(\theta^\varepsilon) = 0$, we obtain that

$$\int_{\mathcal{Y}}\int_{\mathcal{X}}\phi_{\lambda'}(x)\overline{\phi_\lambda(x)}F_{\theta^\varepsilon,y}(x)d\mu(x)d\nu(y) = \delta_0(\lambda'-\lambda),$$

where δ_0 stands for the dirac function at 0. Therefore, it follows that

$$\frac{1}{\varepsilon}\sum_{\lambda'\in\Lambda}\sum_{\lambda\in\Lambda}\tau_{\lambda'}\overline{\tau_\lambda}\int_{\mathcal{Y}}\int_{\mathcal{X}}\phi_{\lambda'}(x)\overline{\phi_\lambda(x)}F_{\theta^\varepsilon,y}(x)d\mu(x)d\nu(y) = \frac{1}{\varepsilon}\|\tau\|^2_{\ell_2}. \tag{2.47}$$

From now on, our goal is to find an upper bound for the second term in the right-hand side of (2.46). By Cauchy-Schwarz's inequality, we have that

$$\left|\sum_{\lambda\in\Lambda}\tau_\lambda\int_{\mathcal{X}}\phi_\lambda(x)F_{\theta^\varepsilon,y}(x)d\mu(x)\right|^2 \leq \|\tau\|^2_{\ell_2(\Lambda)}\|Dh_\varepsilon(\theta^\varepsilon,y)\|^2_{\ell_2(\Lambda)}. \tag{2.48}$$

Moreover, it follows from Parseval's identity, [164][Theorem 1.7] together with the fact that $\int_{\mathcal{X}} F_{\theta^\varepsilon,y}(x)d\mu(x) = 1$, that

$$\|Dh_\varepsilon(\theta^\varepsilon,y)\|^2_{\ell_2(\Lambda)} = \int_{\mathcal{X}} F^2_{\theta^\varepsilon,y}(x)d\mu(x) - 1. \tag{2.49}$$

Hence, combining (2.48) and (2.49), we obtain that

$$\int_{\mathcal{Y}} \left| \sum_{\lambda \in \Lambda} \tau_\lambda \int_{\mathcal{X}} \phi_\lambda(x) F_{\theta^\varepsilon, y}(x) d\mu(x) \right|^2 d\nu(y) \leq \|\tau\|_{\ell_2}^2 \left(\int_{\mathcal{Y}} \int_{\mathcal{X}} F_{\theta^\varepsilon, y}^2(x) d\mu(x) d\nu(y) - 1 \right). \tag{2.50}$$

Finally, we deduce (2.45) from (2.46), (2.47) and (2.50). □

Proof of Theorem 2.4

We shall proceed to the almost sure convergence of the random sequence $(\widehat{\theta}_n)_n$. Let (V_n) be the Lyapunov sequence defined, for all $n \geq 1$, by

$$V_n = \|\widehat{\theta}_n - \theta^\varepsilon\|_{W^{-1}}^2.$$

Assumption 2.1 ensures that $\|\theta^\varepsilon\|_{W^{-1}} < +\infty$. Moreover, we clearly have from (2.8) that

$$W^{-1/2}\widehat{\theta}_{n+1} = W^{-1/2}\widehat{\theta}_n - \gamma_n W^{1/2} D_\theta h_\varepsilon(\widehat{\theta}_n, Y_{n+1}),$$

where W^α stands for the linear operator, for $\alpha \in \{-1/2, 1/2\}$, that maps $v = (v_\lambda)_{\lambda \in \Lambda} \in \ell_\infty(\Lambda)$ to $(w_\lambda^\alpha v_\lambda)_{\lambda \in \Lambda}$. It follows from (2.11) that $\|\widehat{\theta}_n\|_{W^{-1}} = \|W^{-1/2}\widehat{\theta}_n\|_{\ell_2}$. Consequently,

$$\|\widehat{\theta}_{n+1}\|_{W^{-1}} \leq \|\widehat{\theta}_n\|_{W^{-1}} + \gamma_n \|W^{1/2} D_\theta h_\varepsilon(\widehat{\theta}_n, Y_{n+1})\|_{\ell_2}.$$

Furthermore, we obtain from (2.20) that

$$\|W^{1/2} D_\theta h_\varepsilon(\widehat{\theta}_n, Y_{n+1})\|_{\ell_2}^2 \leq \|w\|_{\ell_1} \sup_{\lambda \in \Lambda} \left| \frac{\partial h_\varepsilon(\theta, y)}{\partial \theta_\lambda} \right|^2 \leq \|w\|_{\ell_1} < \infty.$$

Therefore, thanks to the assumption that $\|\widehat{\theta}_0\|_{W^{-1}} < +\infty$, we deduce by induction that $\|\widehat{\theta}_n\|_{W^{-1}} < +\infty$, which means that the Lyapunov sequence (V_n) is well defined. From now on, it follows from (2.8) and (2.11) that for all $n \geq 0$,

$$\begin{aligned} V_{n+1} &= \|\widehat{\theta}_n - \theta^\varepsilon - \gamma_n W D_\theta h_\varepsilon(\widehat{\theta}_n, Y_{n+1})\|_{W^{-1}}^2, \\ &= V_n - 2\gamma_n \langle \widehat{\theta}_n - \theta^\varepsilon, D_\theta h_\varepsilon(\widehat{\theta}_n, Y_{n+1}) \rangle + \gamma_n^2 \|D_\theta h_\varepsilon(\widehat{\theta}_n, Y_{n+1})\|_W^2. \end{aligned}$$

Moreover, (2.20) implies that $\|D_\theta h_\varepsilon(\widehat{\theta}_n, Y^{n+1})\|_W^2 \leq \|w\|_{\ell_1}$ which ensures that for all $n \geq 0$,

$$V_{n+1} \leq V_n - 2\gamma_n \langle \widehat{\theta}_n - \theta^\varepsilon, D_\theta h_\varepsilon(\widehat{\theta}_n, Y_{n+1}) \rangle + \gamma_n^2 \|w\|_{\ell_1}. \tag{2.51}$$

Denote by $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$ the σ-algebra generated by $Y_1, \ldots, Y_n$ drawn from ν. From Proposition 2.5, $\mathbb{E}[D_\theta h_\varepsilon(\widehat{\theta}_n, Y_{n+1}) | \mathcal{F}_n] = DH_\varepsilon(\widehat{\theta}_n)$, which implies via (2.51) that for all $n \geq 0$,

$$\mathbb{E}[V_{n+1} | \mathcal{F}_n] \leq V_n + A_n - B_n \qquad \text{a.s.} \tag{2.52}$$

where (A_n) and (B_n) are the two positive sequences given, for all $n \geq 0$, by

$$A_n = \gamma_n^2 \|w\|_{\ell_1} \qquad \text{and} \qquad B_n = 2\gamma_n DH_\varepsilon(\widehat{\theta}_n)[\widehat{\theta}_n - \theta^\varepsilon].$$

Therefore, as $\sum_{n=0}^{\infty} A_n < \infty$, we deduce from the Robbins-Siegmund theorem [144] that the sequence (V_n) converges almost surely to a finite random variable V and that the series

$$\sum_{n=0}^{\infty} B_n = 2 \sum_{n=0}^{\infty} \gamma_n DH_\varepsilon(\widehat{\theta}_n)[\widehat{\theta}_n - \theta^\varepsilon] < \infty \qquad \text{a.s.} \tag{2.53}$$

Hence, by combining (2.53) with the first condition in (2.9), it necessarily follows that

$$\lim_{n\to\infty} DH_\varepsilon(\widehat{\theta}_n)[\widehat{\theta}_n - \theta^\varepsilon] = 0 \qquad \text{a.s.} \tag{2.54}$$

Hereafter, our goal is to prove that $\|\widehat{\theta}_n - \theta^\varepsilon\|_{\ell_2}$ goes to zero almost surely as n tends to infinity. From now on, let g be the function defined in (2.29). One can easily see that g is a continuous and strictly decreasing function. Moreover, using the Cauchy-Schwarz inequality, one has that $\|\widehat{\theta}_n - \theta^\varepsilon\|_{\ell_1}^2 \leq \|w\|_{\ell_1} V_n$. Hence, it follows from inequality (2.28) that for all $n \geq 0$,

$$DH_\varepsilon(\widehat{\theta}_n)[\widehat{\theta}_n - \theta^\varepsilon] \geq g\Big(\frac{2}{\varepsilon}\|w\|_{\ell_1}^{1/2} V_n^{1/2}\Big) D^2 H_\varepsilon(\theta^\varepsilon)[\widehat{\theta}_n - \theta^\varepsilon, \widehat{\theta}_n - \theta^\varepsilon]. \tag{2.55}$$

Therefore, we obtain from Assumption 2.2 and inequality (2.55) that for all $n \geq 0$,

$$DH_\varepsilon(\widehat{\theta}_n)[\widehat{\theta}_n - \theta^\varepsilon] \geq c_\varepsilon g\Big(\frac{2}{\varepsilon}\|w\|_{\ell_1}^{1/2} V_n^{1/2}\Big) \|\widehat{\theta}_n - \theta^\varepsilon\|_{\ell_2}^2. \tag{2.56}$$

Since (V_n) converges a.s. to a finite random variable V, it follows by continuity of g that

$$\lim_{n\to\infty} g\Big(\frac{2}{\varepsilon}\|w\|_{\ell_1}^{1/2} V_n^{1/2}\Big) = g\Big(\frac{2}{\varepsilon}\|w\|_{\ell_1}^{1/2} V^{1/2}\Big) \qquad \text{a.s.} \tag{2.57}$$

and the limit in the right-hand side of (2.57) is positive almost surely. Therefore, we conclude from (2.54), (2.56) and (2.57) that

$$\lim_{n\to\infty} \|\widehat{\theta}_n - \theta^\varepsilon\|_{\ell_2} = 0 \qquad \text{a.s.}$$

Finally, we deduce from Parseval's identity, [164][Theorem 1.7] that

$$\int_{\mathcal{X}} |\widehat{u}_\varepsilon^n(x) - u_\varepsilon(x)|^2 d\mu(x) = \|\widehat{\theta}_n - \theta^\varepsilon\|_{\ell_2}^2$$

which achieves the proof of Theorem 2.4. □

3

Directional statistics

The definition of Monge-Kantorovich quantile and distribution functions has recently been extended to non-euclidean data in [84]. Again, providing regularized estimators is highly desirable. Thus, we generalize the entropic map on the unit hypersphere $\mathbb{S}^{d-1}$ and motivate its use. In addition, we extend our algorithm from Chapter 2 to solve entropic optimal transport, by expanding dual potentials in the basis of spherical harmonics. Besides, we define a directional Monge-Kantorovich depth, and show that it benefits from desirable properties related to Liu-Zuo-Serfling axioms for the statistical analysis of directional data. This chapter is based on [20].

Contents

3.1 Introduction

3.1.1 Context and bibliography

Quantiles for directional data

In various situations, data naturally correspond to directions that are modeled as observations belonging to the circle or the unit d-sphere $\mathbb{S}^{d-1}$ for $d \geq 2$. Such observations, referred to as directional data, can be found in various applications including wildfires [5], gene expressions [58], or cosmology [119] to name but a few. Directional statistics [109, 110, 118, 137] is the field that brings together the corresponding models, methods and applications for statistical inference. In this chapter, we focus on the concept of quantiles for directional data. Beyond the setting of distributions with rotational symmetry [108], the absence of a canonical ordering on $\mathbb{S}^{d-1}$ precludes a consensual definition of quantiles. In this context, Monge-Kantorovich quantiles have recently been introduced in [84], in line with the euclidean definitions, [33, 87]. Starting from independent and identically distributed ($i.i.d.$) directional data $X_1, \cdots, X_n$ sampled from a target measure ν, the main idea in [84] is to define an empirical quantile function $\mathbf{Q}_n$ as a Monge map, from $\mu_{\mathbb{S}^{d-1}}$ the uniform probability distribution on $\mathbb{S}^{d-1}$ towards the empirical measure $\widehat{\nu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, with a transport cost equal to the squared Riemannian distance. In this manner, if U denotes a random variable with distribution $\mu_{\mathbb{S}^{d-1}}$, the random vector $\mathbf{Q}_n(U)$ follows the empirical distribution of the observations, which is consistent with the standard univariate quantile function. Many statistical properties of these directional quantiles and related notions of ranks, signs and MANOVA are then investigated in [84], with various numerical experiments.

Related works

The first notion of directional depth function was introduced in [162], followed by the work of [113] that developed three different approaches. The properties of the latter have been studied and applied for inference in [3, 149]. The required computational effort led the authors of [108] to build the angular Mahalanobis depth, and, doing so, they provided the first concept of directional quantiles. Despite appealing properties, the obtained contours are constrained to be rotationally symmetric, motivating the elliptic counterpart from [90]. Still, the elliptic assumption is a strong one as discussed in [84]. Facing either this lack of adaptiveness or the computational burdens of previous references, distance-based depths were proposed in [135], even though not explicitly related to the notion of quantiles. These directional depth functions can been applied, for instance, in data analysis and inference [3, 102, 113, 162], classification [54, 55, 102, 113, 130, 135] or clustering [134]. Among recent years, two concepts of multivariate quantiles have emerged in $\mathbb{R}^d$, namely the spatial quantiles [32] and the center-outward ones [33, 87], both gathering most of commonly sought-after properties. More importantly here, these

promising ideas have successfully been extended to directional data [84, 102], improving on the lack of adaptiveness of Mahalanobis quantiles [108], with desirable asymptotic results inherited from the formalism of quantiles.

To put it in a nutshell, existing concepts of directional quantiles include Mahalanobis quantiles, Spatial quantiles and Monge-Kantorovich ones. On the one hand, a statistical depth associated to a notion of quantiles is amenable to benefit from the best of both worlds, that is adaptivity to the underlying geometry and consistency of empirical versions, as argued for instance in [102]. On another hand, in comparison with other directional quantiles, Monge-Kantorovich ones present an additional descriptive power inherited from the fact that $\mathbf{Q}(U) \sim \nu$. A direct consequence is that $\mathbf{Q}$ must contain all the available information, which is appealing with the purpose of summing up unknown features of multivariate data. Even more, these concepts provide a curvilinear coordinate system within the support of the distribution of interest, [84], which is a promising way to render the information.

Main contributions

To compute the estimator $\mathbf{Q}_n$, it is proposed in [84] to first approximate the uniform measure $\mu_{\mathbb{S}^{d-1}}$ by a "regular" grid of n-points over the unit sphere $\mathbb{S}^{d-1}$. Then, the empirical quantile function is defined through the discrete OT problem between this n-points grid and $\widehat{\nu}_n$. However, being a matching between two discrete distributions, the resulting quantile function does not provide out-of-sample estimates which is desirable in many statistical applications. There, regularization naturally enters the picture, as argued in [15, 87]. Consequently, the entropic map is extended here to the spherical setting, similarly to what is done in [48] for general costs in $\mathbb{R}^d$. To the best of our knowledge, this appears to be new, although it benefits from explicit formulation tractable in linear time, given the dual potentials solving EOT. This belongs to the line of work estimating OT maps on manifolds, see *e.g.* [37, 66, 136].

In addition, the computational cost of finding a numerical solution to a discrete OT problem is known to scale cubically in the number of observations [46]. To circumvent this issue, we suggest to adapt the stochastic algorithm developed in [19] dealing with multivariate data in $\mathbb{R}^d$, that is not assumed to belong to the unit d-sphere. On the sphere $\mathbb{S}^2$, dual potentials can be parameterized via spherical harmonics coefficients instead, that is the analog of the Fourier basis for square-integrable functions on $\mathbb{S}^2$. In this manner, using a sequence of random variables $X_1, \dots, X_n$ sampled from a target distribution ν supported on $\mathbb{S}^2$, we construct a stochastic algorithm in the space of spherical harmonics coefficients. In practice, this algorithm depends on the choice of a grid of points in $\mathbb{S}^2$ of size $\mathcal{O}(p^2)$ to implement a FFT on $\mathbb{S}^2$ [175]. The computational cost at each iteration is thus of order $\mathcal{O}\left(p^2 \log^2(p)\right)$ [104].

Finally, we introduce the directional version of the MK statistical depth, in accordance with the euclidean MK depth [33]. We discuss its properties relative to

traditional Liu-Zuo-Serfling axioms for the statistical analysis of directional data. Moreover, we study statistical applications built from it, and provide a comparison with other estimators, to better highlight the potential of entropic regularization for empirical spherical quantiles.

3.1.2 Mathematical prerequisites

In this section, we introduce the main definitions of OT-based distribution and quantile functions for spherical data, beginning with notation related to spherical harmonics and differentiation on $\mathbb{S}^2$.

Preliminaries and notations

The unit 2-sphere is defined by $\mathbb{S}^2 = \{x \in \mathbb{R}^3 : \|x\| = 1\}$. The points $x \in \mathbb{S}^2$ can be written in spherical coordinates, with longitude $\phi \in [-\pi, \pi]$ and colatitude $\theta \in [0, \pi]$, as

$$x = \Phi(\theta, \phi) := (\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta). \tag{3.1}$$

On the 2-sphere, the geodesic distance is $d(x, y) = \arccos(\langle x, y\rangle)$, and the squared Riemannian distance is

$$c(x, y) = \frac{1}{2} d(x, y)^2. \tag{3.2}$$

Both d and c are continuous and bounded as $d(x, y) \in [0, \pi]$. Moreover, $(\mathbb{S}^2, d)$ is a separable complete metric space, with Borel algebra $\mathcal{B}^2$. The surface measure $\sigma_{\mathbb{S}^2}$ on $\mathbb{S}^2$ is given by

$$\int_{\mathbb{S}^2} f(x) d\sigma_{\mathbb{S}^2}(x) = \int_0^{\pi} \int_{-\pi}^{\pi} f(\Phi(\theta, \phi)) \sin\theta d\phi d\theta,$$

and the uniform measure on $\mathbb{S}^2$ writes $\mu_{\mathbb{S}^2} = \frac{1}{4\pi}\sigma_{\mathbb{S}^2}$. The space of all equivalence classes of square integrable functions on $\mathbb{S}^2$ is denoted by $L^2(\mathbb{S}^2)$. We define the *spherical harmonic* function of degree $l \in \mathbb{N}^*$ and order $m \in \{-l, \cdots, l\}$ by

$$Y_l^m(x) = Y_l^m(\Phi(\theta, \phi)) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) e^{im\phi},$$

where the associated Legendre functions $P_l^m : [-1, 1] \to \mathbb{R}$ verify, for $l \in \mathbb{N}^*$ and $m \geq 0$,

$$P_l^m(t) = \frac{(-1)^m}{2^l l!} (1 - t^2)^{m/2} \frac{d^{l+m}(t^2 - 1)^l}{dt^{l+m}} \quad \text{and} \quad P_l^{-m}(t) = (-1)^m \frac{(l-m)!}{(l+m)!} P_l^m(t).$$

Importantly, the spherical harmonics form an orthonormal basis of $L^2(\mathbb{S}^2)$, so that every function $f \in L^2(\mathbb{S}^2)$ is uniquely decomposed, for $x \in \mathbb{S}^2$, as

$$f(x) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \bar{f}_l^m Y_l^m(x), \tag{3.3}$$

where the sequence of spherical harmonic coefficients $\bar{f} = (\bar{f}_l^m)$ verifies

$$\bar{f}_l^m = \frac{1}{4\pi} \int_{\mathbb{S}^2} f(x) \overline{Y_l^m(x)} d\sigma_{\mathbb{S}^2}(x). \tag{3.4}$$

We refer to [35] for an introduction to Fourier analysis on the sphere.

Below, we introduce a few notation from differential geometry [163], that one can find for instance in [61]. At any point $x \in \mathbb{S}^2$, the tangent space is $\mathcal{T}_x\mathbb{S}^2 = \{y \in \mathbb{R}^3 : \langle x, y\rangle = 0\}$, and the associated orthogonal projection $\rho_x : \mathbb{R}^3 \to \mathcal{T}_x\mathbb{S}^2$ verifies

$$\rho_x \xi = (I - xx^T)\xi = \xi - \langle \xi, x\rangle x. \tag{3.5}$$

The exponential map at $x \in \mathbb{S}^2$, $\mathrm{Exp}_x : \mathcal{T}_x\mathbb{S}^2 \to \mathbb{S}^2$, has the explicit form

$$\mathrm{Exp}_x(v) = \cos(\|v\|)x + \sin(\|v\|)\frac{v}{\|v\|},$$

and its inverse Log_x writes

$$\mathrm{Log}_x(z) = \frac{d(x,z)}{\sqrt{1 - \langle x, z\rangle^2}} \rho_x z = d(x,z) \frac{\rho_x(z-x)}{\|\rho_x(z-x)\|}. \tag{3.6}$$

We take the extrinsic viewpoint for the manifold $\mathbb{S}^2$ embedded in $\mathbb{R}^3$, so that the Riemannian gradient is given by orthogonally projecting Euclidean derivatives onto the tangent space. For a smooth function $f : \mathbb{S}^2 \to \mathbb{R}$, its Riemannian gradient $\nabla f(x)$ at x is thus defined by

$$\nabla f(x) = \rho_x Df(x), \qquad \text{where} \qquad Df(x) = \Big(\frac{\partial f(x)}{\partial x_i}\Big)_{i,j\in\{1,2,3\}}. \tag{3.7}$$

The Riemannian Hessian $\nabla^2 f(x) : \mathcal{T}_x\mathbb{S}^2 \to \mathcal{T}_x\mathbb{S}^2$ at x is defined by the same token,

$$\nabla^2 f(x) = \rho_x \Big[D^2 f(x) - \langle Df(x), x\rangle I\Big] \qquad \text{with} \qquad D^2 f(x) = \Big(\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\Big)_{i,j\in\{1,2,3\}}. \tag{3.8}$$

Main definitions for directional MK quantiles

Here, we fix the definitions of directional distribution and quantile functions as introduced in [84]. Given μ and ν two probability measures supported on $\mathbb{S}^2$, optimal transportation follows the same definitions than in Chapter 1. Before considering the existence of Monge maps, we recall the key definition of c-transforms as stated in [122] on Riemannian manifolds, that is equivalent to the formulation of [84][Definition 2].

Definition 3.1. *Given a function $\psi : \mathbb{S}^2 \to \mathbb{R}$, its c-transform is defined by*

$$\psi^c(y) = \inf_{x\in\mathbb{S}^2} \{c(x,y) - \psi(x)\}.$$

Then, ψ is said to be c-concave when $\psi^{cc} := (\psi^c)^c = \psi$.

The proper summary of [84][Proposition 1] highlights that c-concavity is related to optimality in Monge's OT problem (P_0). For our continuous and bounded cost c, an appropriate assumption is that the reference measure belongs to $\mathbf{B}_2$, the family of $\sigma_{\mathbb{S}^2}$-absolutely continuous distributions with densities bounded away from 0 and ∞, see [122][Theorem 9], which is the case for the uniform measure $\mu_{\mathbb{S}^2}$. This enables the definition of directional distribution and quantile functions, first introduced in [84].

Definition 3.2. *The directional MK quantile function of the arbitrary probability measure ν is the $\mu_{\mathbb{S}^2}$-a.s. unique map $\mathbf{Q} : \mathbb{S}^2 \to \mathbb{S}^2$ such that $\mathbf{Q}_{\#}\mu_{\mathbb{S}^2} = \nu$ and there exists a c-concave differentiable mapping $\psi : \mathbb{S}^2 \to \mathbb{R}$ such that, $\sigma_{\mathbb{S}^2}$-a.e.,*

$$\mathbf{Q}(x) = Exp_x(-\nabla\psi(x)).$$

In addition, the directional MK distribution function of ν is given by

$$\mathbf{F}(x) = Exp_x(-\nabla\psi^c(x)).$$

As soon as ν belongs to $\mathbf{B}_2$, [122][Corollary 10] ensures that $\mathbf{F} = \mathbf{Q}^{-1}$ almost everywhere, whereas $\mathbf{Q}^{-1}$ might not exist if ν is not absolutely continuous. Compared to [84], our definition begins with $\mathbf{Q}$ instead of $\mathbf{F}$, and it does not require the absolute continuity for ν. This follows developments from [74] for measures supported in $\mathbb{R}^d$.

Remark 3.3 (Regularity)**.** *The regularity of OT maps on the sphere is a delicate subject that has inspired a number of works, including [53, 114, 115, 131]. Firstly, any c-concave potential ψ is twice differentiable almost everywhere [39][Proposition 3.14]. For further regularity, an appropriate requirement is that the underlying measures are smooth and belong to $\mathbf{B}_2$. In particular, if, a minima, ν has density $f \in C^{1,1}(\mathbb{S}^2)$ with respect to $\sigma_{\mathbb{S}^2}$, then the MK quantile function $\mathbf{Q}$ belongs to $C^{2,\beta}(\mathbb{S}^2)$ for all $\beta \in]0,1[$, see [115] for more details.*

Remark 3.4 (Gradient mappings)**.** *A gradient mapping is built from a c-concave potential ψ, through $x \mapsto Exp_x(-\nabla\psi(x))$. A statistical model based on convex combinations of such maps was introduced in [154], and further used for directional data in [155]. With the viewpoint of [33, 84, 87], this amounts to a barycenter model for MK quantile functions.*

In view of the proof of [122][Theorem 9], ψ solves the semi-dual version of Kantorovich's problem (S_0). Our proposal is to build upon this to tackle the issue of finding regularized estimators for $\mathbf{F}$ and $\mathbf{Q}$. Before that, we recall the definitions of directional quantile contours and regions from [84], that simplify in dimension $d = 3$. A central point in $\mathbb{S}^2$ must be chosen for the uniform distribution $\mu_{\mathbb{S}^2}$,

in view of defining nested regions with $\mu_{\mathbb{S}^2}$-content $\tau \in [0,1]$. In our Riemannian framework, a well-suited notion of central point is the Fréchet median

$$\theta_M = \underset{z \in \mathbb{S}^2}{\operatorname{argmax}}\, \mathbb{E}_{Z \sim \nu}[d(Z,z)], \tag{3.9}$$

that can be computed with the package *geomstats* [126], in Python. Then, the spherical cap with $\mu_{\mathbb{S}^2}$ probability $\tau \in [0,1]$ centered at $\mathbf{F}(\theta_M)$ is

$$\mathbb{C}_\tau^U = \left\{ x \in \mathbb{S}^2 : \langle x, \mathbf{F}(\theta_M) \rangle \geq 1 - 2\tau \right\},$$

with boundary $\mathcal{C}_\tau^U = \{x \in \mathbb{S}^2 : \langle x, \mathbf{F}(\theta_M)\rangle = 1 - 2\tau\}$ a *parallel* of order τ. This defines a rotated version of the usual latitude-longitude coordinate system (3.1), with respect to the pole $\mathbf{F}(\theta_M)$, as follows. Any $x \in \mathbb{S}^2$ decomposes into

$$x = \langle x, \mathbf{F}(\theta_M)\rangle \mathbf{F}(\theta_M) + \sqrt{1 - \langle x, \mathbf{F}(\theta_M)\rangle^2}\, \mathbf{S}_{\mathbf{F}(\theta_M)}(x), \tag{3.10}$$

where $\langle x, \mathbf{F}(\theta_M)\rangle$ is a latitude, constant over the parallel $\mathcal{C}_\tau^U$, while the *directional sign*

$$\mathbf{S}_{\mathbf{F}(\theta_M)}(x) = \frac{x - \langle x, \mathbf{F}(\theta_M)\rangle \mathbf{F}(\theta_M)}{\| x - \langle x, \mathbf{F}(\theta_M)\rangle \mathbf{F}(\theta_M) \|} \tag{3.11}$$

is a longitude, with the convention $\mathbf{0}/0 = \mathbf{0}$ for $x = \pm \mathbf{F}(\theta_M)$. The unit vector $\mathbf{S}_{\mathbf{F}(\theta_M)}(x)$ takes values on the rotated equator $\mathcal{C}_{1/2}^U$, and thus allows to characterize meridians crossing $s \in \mathcal{C}_{1/2}^U$ through $\mathcal{M}_s^U = \{x \in \mathbb{S}^2 : \mathbf{S}_{\mathbf{F}(\theta_M)}(x) = s\}$. For ease of understanding, we take the example of $\mathbf{F}(\theta_M) = (0,0,1)^T$: taking $x \in \mathcal{C}_\tau^U$ such that $\langle x, \mathbf{F}(\theta_M)\rangle = 1 - 2\tau$ is equivalent to $x_3 = 1 - 2\tau$. Thus, we retrieve that for a fixed τ, quantile contours $\mathcal{C}_\tau^U$ have indeed a fixed latitude in the classical longitude-latitude system (3.1). In fact, one can use contours of constant latitude in the system (3.1) to discretize a reference quantile contour oriented towards $\mathbf{F}(\theta_M)$, by choosing the appropriate rotation matrix $\mathbf{O}$ that sends $(0,0,1)^T$ towards $\mathbf{F}(\theta_M)$. Numerically, it can be computed using Rodrigues' rotation formula, for instance.

The image by $\mathbf{Q}$ of the parallel / meridian system (3.10) provides curvilinear parallels $\mathbf{Q}(\mathcal{C}_\tau^U)$ and curvilinear meridians $\mathbf{Q}(\mathcal{M}_s^U)$ adapted to the geometry of the support of ν, giving rise to suitable directional concepts of quantile contours and signs [84]. Intuitively, a change in coordinates in a data-adaptive fashion must retain all the available information, in a simpler form amenable to be summed up.

Definition 3.5 (Quantile contours, regions and signs)**.** *Let $\nu \in \mathbf{B}_2$, with directional quantile function $\mathbf{Q}$. Then,*

- *the quantile contour of order $\tau \in [0,1]$ is $\mathcal{C}_\tau = \mathbf{Q}(\mathcal{C}_\tau^U)$,*
- *the quantile region of order $\tau \in [0,1]$ is $\mathbb{C}_\tau = \mathbf{Q}(\mathbb{C}_\tau^U)$,*
- *the sign curve associated with $s \in \mathcal{C}_{1/2}^U$ is $\mathbf{Q}(\mathcal{M}_s^U)$.*

Since $\mathbf{Q}$ is a push-forward mapping, $\mathbf{Q}_{\#}\mu_{\mathbb{S}^2} = \nu$, the ν-probability content of $\mathbb{C}_\tau$ is τ. Moreover, the quantile contours of $\nu \in \mathbf{B}_2$ are continuous and the quantile

regions are closed, connected and nested, as stated in [84]. Invariance properties, that were shown in [84], are gathered in Appendix B.1, for the sake of completeness.

3.2 Regularized estimation of MK quantiles on the 2-sphere

3.2.1 Entropic OT on the 2-sphere

We now introduce an algorithm based on the spherical Fourier transform to solve the regularized Kantorovich problem on the 2-sphere. In the Euclidean case, Kantorovich's problem is known to be easier to solve than Monge's problem [46]. Even more so, since the founding work of [47], adding an entropic regularization term to (P_0') has been a cornerstone for the development of OT-based methods in statistics and machine learning. In [71], rewriting the dual objective function to be optimized allowed the introduction of stochastic algorithms to obtain provably convergent algorithms, see also [17]. For arbitrary measures, (not only discrete ones), the dual variables cannot be viewed as finite-dimensional vectors anymore. Therefore, one requires the use of nonparametric families of dual functions, as proposed in [71] with reproducing kernel Hilbert spaces or in [153] with deep neural networks. In Chapter 2, we suggested the use of Fourier series in the specific context of center-outward quantiles to take advantage of the knowledge of the reference measure. The resulting algorithm directly targets the continuous OT problem between the reference measure and the underlying ν, instead of the discrete or semi-discrete OT problem towards the empirical measure $\widehat{\nu}_n$. The idea that we are introducing here for spherical distributions is in the same spirit. We mention that several algorithms exist in order to solve the unregularized OT problem on the sphere, see [88] and the references therein. Also, the Network simplex and Sinkhorn algorithm only depend on the cost matrix, so that they can be adapted trivially on any space. While these discrete solvers require the storage of the cost matrix, of size n^2 for two samples of sizes n, stochastic algorithms are designed to avoid it.

In our spherical setting, the cost is bounded, thus the problem (S_ε) admits a solution in L^∞, unique up to additive constants, [70][Theorem 7]. To leverage unicity, we impose that $\int_{\mathbb{S}^2} u(x) d\mu_{\mathbb{S}^2}(x) = 0$, so that the problem (S_ε) becomes

$$\max_{u \in L^\infty(\mathbb{S}^2)} \int_{\mathbb{S}^2} u^{c,\varepsilon}(y) d\nu(y).$$

It is well-known, [71, 133], that $\mathbf{u}_\varepsilon$ is solution of (S_ε) if and only if $\mathbf{u}_\varepsilon = ((\mathbf{u}_\varepsilon)^{c,\varepsilon})^{c,\varepsilon}$, see (1.11). Note that a Lipschitz continuous function on $\mathbb{S}^2$ shall equal its spherical Fourier series (3.3) pointwise [125][Theorem 5.26]. One can find in [122][Lemma 2] that the cost c and the true unregularized potentials are Lipschitz, because $\mathbb{S}^2$ has a finite diameter $|\mathbb{S}^2| = \pi$. Furthermore, the same holds in the regularized case $\varepsilon > 0$, from the optimality condition (1.11), see Proposition 12 from [62][Appendix

B] or [133][Lemma 3.1]. Consequently, we suggest to parameterize the dual variable in (S_ε) by its spherical harmonic coefficients. For a given $\varepsilon > 0$, we consider the optimal sequence of coefficients $\bar{\mathbf{u}}_\varepsilon$ defined as the solution of the following *stochastic convex minimisation* problem

$$\bar{\mathbf{u}}_\varepsilon = \underset{\bar{\mathbf{u}} \in \ell_1}{\operatorname{argmax}} H_\varepsilon(\bar{\mathbf{u}}) \qquad \text{with} \qquad H_\varepsilon(\bar{\mathbf{u}}) = \mathbb{E}\left[h_\varepsilon(\bar{\mathbf{u}}, X)\right] \tag{3.12}$$

where X is a random vector with distribution ν , $\bar{\mathbf{u}} = (\bar{\mathbf{u}}_l^m)_{l \geq 1}$ and

$$h_\varepsilon(\bar{\mathbf{u}}, x) = -u^{c,\varepsilon}(x) \quad \text{with} \quad u(z) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \bar{\mathbf{u}}_l^m Y_l^m(z).$$

Note that the spherical harmonic coefficient $\bar{\mathbf{u}}_0^0$ equals 0 because of the identifiability condition $\int_{\mathbb{S}^2} u(x) d\mu_{\mathbb{S}^2}(x) = 0$.

We shall now discuss the equivalence between (3.12) and the original problem (S_ε), and in particular the restriction $\bar{\mathbf{u}} \in \ell_1$. On the 2-sphere, the series of spherical harmonics of a continuously differentiable function is uniformly convergent, see [98][p.259] or [96]. To obtain the stronger result that the sequence of spherical harmonics belongs to ℓ_1, the function needs to be twice continuously differentiable, [96][Theorem 2]. Regarding the unregularized Kantorovich potentials, such differentiability requires smoothness of the measures involved, as highlighted in Remark 3.3. A sufficient condition when the reference measure is $\mu_{\mathbb{S}^2}$ is that the density of ν is differentiable and bounded above and below by positive constants, see [115][Corollary 6.2]. It appears that such conditions on ν are not needed for the regularized potential solving (S_ε). Indeed, it shall be proven in Proposition 3.14 that $\mathbf{u}_\varepsilon$ is always twice continuously differentiable.

Consequently, the problem (3.12) is equivalent to the original one (S_ε). The main virtue of this parameterization is that partial derivatives of h_ε with respect to the parameters $\bar{\mathbf{u}}_l^m$ can be derived easily, which is appealing in view of a stochastic gradient scheme. Here, the objective $h_\varepsilon(\cdot, x)$ is of the same mathematical nature than in [19], so that it is differentiable and the following property holds.

Proposition 3.6. *For every $x \in \mathbb{S}^2$, the function $h_\varepsilon(\cdot, x) : \ell_1 \to \mathbb{R}$ is Fréchet differentiable and its differential $D_{\bar{\mathbf{u}}} h_\varepsilon(\bar{\mathbf{u}}, x)$ belongs to the dual Banach space $(\ell_\infty, \|\cdot\|_{\ell_\infty})$ where $\|\bar{\mathbf{u}}\|_{\ell_\infty} = \sup_{l,m} |\bar{\mathbf{u}}_l^m|$. The components of $D_{\bar{\mathbf{u}}} h_\varepsilon(\bar{\mathbf{u}}, x)$ are the partial derivatives*

$$\frac{\partial h_\varepsilon(\bar{\mathbf{u}}, x)}{\partial \bar{\mathbf{u}}_l^m} = \frac{1}{4\pi} \int_{\mathbb{S}^2} g_{\bar{\mathbf{u}}, x}(z) Y_l^m(z) d\sigma_{\mathbb{S}^2}(z), \tag{3.13}$$

that are the spherical harmonics coefficients of the function

$$g_{\bar{\mathbf{u}}, x}(z) = \frac{\exp\left(\frac{u(z) - c(z,x)}{\varepsilon}\right)}{\int \exp\left(\frac{u(y) - c(y,x)}{\varepsilon}\right) d\mu_{\mathbb{S}^2}(y)} \quad \text{with} \quad u(z) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \bar{\mathbf{u}}_l^m Y_l^m(z). \tag{3.14}$$

For (X_n) a sequence of independent random vectors with distribution ν, we consider the stochastic algorithm in the Banach space $(\ell_1, \|\cdot\|_{\ell_1})$ defined, for all $n \geq 0$, by

$$\widehat{u}_{n+1} = \widehat{u}_n - \gamma_n W D_{\widehat{u}} h_\varepsilon(\widehat{u}_n, X_{n+1}) \tag{3.15}$$

where $\gamma_n = \gamma n^{-\alpha}$ is a decreasing sequence of positive numbers with $1/2 < \alpha < 1$ and $\gamma > 0$. Because ℓ_1 differs from its dual space ℓ_∞, the linear operator W is defined by

$$\left\{ \begin{array}{rcl} W : (\ell_\infty, \|\cdot\|_{\ell_\infty}) & \to & (\ell_1, \|\cdot\|_{\ell_1}) \\ \bar{v} = (\bar{v}_l^m) & \mapsto & \bar{w} \odot \bar{v} = (\bar{w}_l^m \bar{v}_l^m) \end{array} \right.$$

where $\bar{w} = (\bar{w}_l^m)$ is a *deterministic sequence of positive weights* satisfying the condition

$$\|\bar{w}\|_{\ell_1} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \bar{w}_l^m < +\infty. \tag{3.16}$$

In all the experiments carried out hereafter, we use the sequence $\bar{w}_l^m = (l^2 + m^2)^{-1}$. For a given regularization parameter $\varepsilon > 0$, a *regularized estimator* of the optimal potential $\mathbf{u}_\varepsilon(x) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \bar{\mathbf{u}}_{\varepsilon,l}^m Y_l^m(x)$ is naturally given by

$$\widehat{\mathbf{u}}_{\varepsilon,n}(x) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \widehat{u}_{n,l}^m Y_l^m(x). \tag{3.17}$$

From a practical point of view, the stochastic sequence (3.15) must be discretized. To do so, one must consider a grid of p^2 points on the 2-sphere and the associated spherical harmonics coefficients. We emphasize that this discretization takes place in the space of frequencies, willing to take advantage from implicit interpolation in this space. The Python library *pyshtools* [175], implements spherical harmonics transforms and reconstructions. Our numerical procedure builds upon it as the stochastic algorithm (3.15) requires computing the spherical harmonics coefficients of the function $g_{\widehat{u},y}(x)$ in (3.14), that relies on $u(x)$ reconstructed from $\widehat{u}$ thanks to the inverse Fourier transform on the 2-sphere. With the help of the fast routine within *pyshtools*, the computational cost at each iteration of (3.15) is of order $\mathcal{O}\left(p^2 \log^2(p)\right)$ [175]. Finally, the estimator $\widehat{\mathbf{u}}_{\varepsilon,n}(x)$ in (3.17) can be quickly recovered for any x thanks to the Python library *sphericart* [25].

Remark 3.7. *To study the convergence of the stochastic algorithm (3.15), we could adapt the theoretical results in [19] to our setting, because they were irrespective of the orthonormal basis and the cost c.*

3.2.2 Regularized distribution and quantile functions

We now introduce the regularized counterpart of Definition 3.2. When dealing with measures supported on $\mathbb{R}^d$, one can use the *entropic map*, defined as the ba-

rycentric projection of the entropic optimal plan, see *e.g.* [140, 143]. At first sight, this requires a notion of average on the sphere, as done in [66] from the unregularized empirical OT plan. Nonetheless, this map is alternatively characterized by analogy with Brenier theorem [140][Proposition 2], whose building block is the gradient of Kantorovich potential. Based on such differentiation, entropic maps were introduced in [48] for OT problems involving general convex costs in $\mathbb{R}^d$. Because this enforces the structure of optimality, we pursue this idea for our non-Euclidean setting. Our numerical experiments in Section 3.4 flesh out the empirical benefits and shortcomings when varying ε.

Definition 3.8. *Let ν be an arbitrary probability measure supported on $\mathbb{S}^2$, and $\mathbf{u}_\varepsilon : \mathbb{S}^2 \to \mathbb{R}$ be a solution of (S_ε) between $\mu_{\mathbb{S}^2}$ and ν. Then, the regularized distribution function of ν is given by*

$$\mathbf{F}_\varepsilon(z) = Exp_z(-\nabla \mathbf{u}_\varepsilon^{c,\varepsilon}(z)), \tag{3.18}$$

and the regularized quantile function of ν is

$$\mathbf{Q}_\varepsilon(x) = Exp_x(-\nabla \mathbf{u}_\varepsilon(x)). \tag{3.19}$$

This requires the differentiation of entropic Kantorovich potentials. For a given regularization parameter $\varepsilon > 0$, partial derivatives can be retrieved by

$$\frac{\partial \mathbf{u}_\varepsilon(x)}{\partial x_i} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \bar{\mathbf{u}}_l^m \frac{\partial Y_l^m(x)}{\partial x_i},$$

where the package *sphericart* [25], allows the computation of $\frac{\partial Y_l^m(x)}{\partial x_i}$ easily. The Riemannian gradient $\nabla \mathbf{u}_\varepsilon$ follows using (3.7). But because this may lead to numerical instabilities, we suggest instead to make use of first-order conditions (1.11). With discrete counterparts in practice, changing a couple of potentials (u, v) to $(v^{c,\varepsilon}, u^{c,\varepsilon})$ improves the objective to solve. As (1.10) depends on the measure $\mu_{\mathbb{S}^2}$, its symmetric version (1.11) depends on ν instead. Notably, Sinkhorn's algorithm corresponds to perform such alternative smooth conjugates [70][Proposition 10]. The next proposition gives a generalized entropic map on the hypersphere $\mathbb{S}^2$.

Proposition 3.9. *Denote by*

$$g_\varepsilon(x, z) = \frac{-d(x, z)}{\sqrt{1 - \langle x, z \rangle^2}} \exp\Big(\frac{\mathbf{u}_\varepsilon(x) - c(x, z) + \mathbf{u}_\varepsilon^{c,\varepsilon}(z)}{\varepsilon}\Big). \tag{3.20}$$

Then, the Euclidean partial derivatives of $\mathbf{u}_\varepsilon$ admit the closed-form expression

$$\partial_{x_i} \mathbf{u}_\varepsilon(x) = \int z_i g_\varepsilon(x, z) d\nu(z). \tag{3.21}$$

Similarly, the Euclidean gradient of $\mathbf{u}_\varepsilon^{c,\varepsilon}$ *verifies*

$$\partial_{z_i} \mathbf{u}_\varepsilon^{c,\varepsilon}(z) = \int x_i g_\varepsilon(x,z) d\mu_{\mathbb{S}^2}(x). \tag{3.22}$$

Proof. Denote by $v = \mathbf{u}_\varepsilon^{c,\varepsilon}$. Then, using the chain rule in (1.11) gives

$$\partial_{x_i} \mathbf{u}_\varepsilon(x) = -\varepsilon \frac{\partial_{x_i} J(x)}{J(x)} \quad \text{for} \quad J(x) = \int \exp\Big(\frac{v(z) - c(x,z)}{\varepsilon}\Big) d\nu(z). \tag{3.23}$$

We now turn to the differentiation of J. As shown in [133][Lemma 2.1],

$$\inf_{x \in \mathcal{X}} \{c(x,z) - \mathbf{u}_\varepsilon(x)\} \leq v(z) \leq \int c(x,z) d\mu_{\mathbb{S}^2}(x).$$

By boundedness of $\mathbb{S}^2$, v is bounded and so is the integrand in (3.23). As we deal with probability measures, this justifies using the differentiation under the integral sign, that induces

$$\partial_{x_i} J(x) = \int -\frac{\partial_{x_i} c(x,z)}{\varepsilon} \exp\Big(\frac{v(z) - c(x,z)}{\varepsilon}\Big) d\nu(z). \tag{3.24}$$

Fix $z \in \mathbb{S}^2$, so that

$$\partial_{x_i} c(x,z) = d(x,z) \partial_{x_i} d(x,z) \quad \text{and} \quad \partial_{x_i} d(x,z) = \frac{-1}{\sqrt{1 - \langle x,z \rangle^2}} z_i. \tag{3.25}$$

Combining (3.24) with (3.25),

$$\partial_{x_i} J(x) = \int \frac{z_i}{\varepsilon} \frac{d(x,z)}{\sqrt{1 - \langle x,z \rangle^2}} \exp\Big(\frac{v(z) - c(x,z)}{\varepsilon}\Big) d\nu(z). \tag{3.26}$$

Plugging (3.26) in (3.23) gives (3.22), where g_ε defined in (3.20) shows up because, by properties of exp,

$$\exp\Big(\frac{v(z) - c(x,z) + \mathbf{u}_\varepsilon(x)}{\varepsilon}\Big) = \frac{\exp\Big(\frac{v(z) - c(x,z)}{\varepsilon}\Big)}{\int \exp\Big(\frac{v(y) - c(x,y)}{\varepsilon}\Big) d\nu(y)} \tag{3.27}$$

By symmetry, the same arguments on $\mathbf{u}_\varepsilon^{c,\varepsilon}(z) = -\varepsilon \log \int \exp\Big(\frac{\mathbf{u}_\varepsilon(x) - c(x,z)}{\varepsilon}\Big) d\mu_{\mathbb{S}^2}(x)$ yield (3.21). □

Combining Proposition 3.9 with the definition of the Riemannian gradient on $\mathbb{S}^2$, (3.7), yields the following corollary, recalling that

$$(x,z) \mapsto \exp\Big(\frac{\mathbf{u}_\varepsilon(x) - c(x,z) + \mathbf{u}_\varepsilon^{c,\varepsilon}(z)}{\varepsilon}\Big)$$

is the density of the optimal entropic plan with respect to $\mu \otimes \nu$, [133].

Corollary 3.10. *The regularized distribution and quantile functions of ν on $\mathbb{S}^2$ admit closed-form expressions through*

$$\mathbf{Q}_\varepsilon(x) = Exp_x \int Log_x(z) \exp\Big(\frac{\mathbf{u}_\varepsilon(x) - c(x,z) + \mathbf{u}_\varepsilon^{c,\varepsilon}(z)}{\varepsilon}\Big) d\nu(z),$$

and

$$\mathbf{F}_\varepsilon(z) = Exp_x \int Log_z(x) \exp\Big(\frac{\mathbf{u}_\varepsilon(x) - c(x,z) + \mathbf{u}_\varepsilon^{c,\varepsilon}(z)}{\varepsilon}\Big) d\mu_{\mathbb{S}^2}(x).$$

Proof. Using the Euclidean derivatives obtained in Proposition 3.9,

$$\nabla \mathbf{u}_\varepsilon(x) = \rho_x \int z g_\varepsilon(x,z) d\nu(z) \quad \text{and} \quad \nabla \mathbf{u}_\varepsilon^{c,\varepsilon}(z) = \rho_z \int x g_\varepsilon(x,z) d\mu_{\mathbb{S}^2}(x).$$

But this is equivalent to

$$\nabla \mathbf{u}_\varepsilon(x) = \int -\frac{d(x,z)}{\sqrt{1-\langle x,z\rangle^2}} \rho_x(z) \exp\Big(\frac{\mathbf{u}_\varepsilon(x) - c(x,z) + \mathbf{u}_\varepsilon^{c,\varepsilon}(z)}{\varepsilon}\Big) d\nu(z),$$

and

$$\nabla \mathbf{u}_\varepsilon^{c,\varepsilon}(z) = \int -\frac{d(x,z)}{\sqrt{1-\langle x,z\rangle^2}} \rho_z(x) \exp\Big(\frac{\mathbf{u}_\varepsilon(x) - c(x,z) + \mathbf{u}_\varepsilon^{c,\varepsilon}(z)}{\varepsilon}\Big) d\mu_{\mathbb{S}^2}(x).$$

There, one recovers $\mathrm{Log}_x = \mathrm{Exp}_x^{-1}$, (3.6), that gives

$$\nabla \mathbf{u}_\varepsilon(x) = -\int \mathrm{Log}_x(z) \exp\Big(\frac{\mathbf{u}_\varepsilon(x) - c(x,z) + \mathbf{u}_\varepsilon^{c,\varepsilon}(z)}{\varepsilon}\Big) d\nu(z),$$

and

$$\nabla \mathbf{u}_\varepsilon^{c,\varepsilon}(z) = -\int \mathrm{Log}_z(x) \exp\Big(\frac{\mathbf{u}_\varepsilon(x) - c(x,z) + \mathbf{u}_\varepsilon^{c,\varepsilon}(z)}{\varepsilon}\Big) d\mu_{\mathbb{S}^2}(x).$$

□

Remark 3.11. *It should be noted that $\mathbf{F}_\varepsilon$, resp. $\mathbf{Q}_\varepsilon$, does not push ν forward to $\mu_{\mathbb{S}^2}$ anymore, resp. $\mu_{\mathbb{S}^2}$ forward to ν. However, they are expected to be close to their unregularized counterparts, for small values of $\varepsilon > 0$, as studied, for the quadratic cost in $\mathbb{R}^d$, in [77, 140, 165]. The limit $\varepsilon \to 0$ has been considered outside the Euclidean setting [24, 133], although not directly about the generalized entropic map itself. In particular, up to some sequence (ε_k) such that $\lim_{k\to+\infty} \varepsilon_k = 0$, Proposition 3.2 from [133] gives us the uniform convergence of potentials $(\mathbf{u}_{\varepsilon_k})$ on compact subsets of $\mathbb{S}^2$, towards ψ solving (S_0).*

From Corollary 3.10, $\mathbf{F}_\varepsilon$ and $\mathbf{Q}_\varepsilon$ can be seen as weighted averages in the tangent space. We argue that this can be gainful in practice, because of the regularity it induces. In particular, one can infer from the next proposition that $\mathbf{F}_\varepsilon$ and $\mathbf{Q}_\varepsilon$ must be continuous.

Proposition 3.12. *The potential* $\mathbf{u}_\varepsilon$ *is twice-differentiable everywhere, and*

$$\frac{\partial^2 \mathbf{u}_\varepsilon}{\partial_{x_i}\partial_{x_j}}(x) = \int \tilde{c}_{ij}(x,z)\exp\Big(\frac{\mathbf{u}_\varepsilon(x) - c(x,z) + \mathbf{u}_\varepsilon^{c,\varepsilon}(z)}{\varepsilon}\Big)d\nu(z) + \frac{\partial_{x_i}\mathbf{u}_\varepsilon(x)\partial_{x_j}\mathbf{u}_\varepsilon(x)}{\varepsilon},$$

where

$$\tilde{c}_{ij}(x,z) = \frac{\partial^2 c(x,z)}{\partial_{x_i}\partial_{x_j}} - \frac{\partial_{x_i}c(x,z)\partial_{x_j}c(x,z)}{\varepsilon}.$$

Besides, the same holds for $\mathbf{u}_\varepsilon^{c,\varepsilon}$*, replacing* $\tilde{c}_{ij}(x,z)$ *by*

$$\bar{c}_{ij}(x,z) = \frac{\partial^2 c(x,z)}{\partial_{z_i}\partial_{z_j}} - \frac{\partial_{z_i}c(x,z)\partial_{z_j}c(x,z)}{\varepsilon}.$$

Proof. The same calculus can be found in [70][Lemma 3], up to the fact that one can recognize partial derivatives of $\mathbf{u}_\varepsilon$, that is (3.21), in the result, at the very end of our proof. First of all, g_ε is bounded by using [133][Lemma 2.1] and the compacity of $\mathbb{S}^2$. Thus, one can differentiate in (3.21) under the integral sign, and

$$\frac{\partial^2 \mathbf{u}_\varepsilon}{\partial_{x_i}\partial_{x_j}}(x) = \int \partial_{x_j} z_i g_\varepsilon(x,z)d\nu(z). \tag{3.28}$$

In view of using classical rules of differentiation, note that

$$z_i g_\varepsilon(x,z) = (\partial_{x_i}c(x,z))G(x,z) \quad \text{for} \quad G(x,z) = \exp\Big(\frac{\mathbf{u}_\varepsilon(x) - c(x,z) + \mathbf{u}_\varepsilon^{c,\varepsilon}(z)}{\varepsilon}\Big). \tag{3.29}$$

Besides,

$$\partial_{x_j}G(x,z) = \frac{1}{\varepsilon}G(x,z)\partial_{x_j}(\mathbf{u}_\varepsilon(x) - c(x,z)).$$

As a byproduct,

$$\begin{aligned}\partial_{x_j} z_i g_\varepsilon(x,z) &= \frac{\partial^2 c(x,z)}{\partial_{x_i}\partial_{x_j}}G(x,z) + \partial_{x_i}c(x,z)\frac{1}{\varepsilon}G(x,z)\Big(\partial_{x_j}\mathbf{u}_\varepsilon(x) - \partial_{x_j}c(x,z)\Big),\\ &= G(x,z)\Big(\frac{\partial^2 c(x,z)}{\partial_{x_i}\partial_{x_j}} + \partial_{x_i}c(x,z)\frac{\partial_{x_j}\mathbf{u}_\varepsilon(x) - \partial_{x_j}c(x,z)}{\varepsilon}\Big).\end{aligned} \tag{3.30}$$

Plugging (3.30) in (3.28) and using (3.29) when rearranging,

$$\begin{aligned}\frac{\partial^2 \mathbf{u}_\varepsilon}{\partial_{x_i}\partial_{x_j}}(x) &= \int G(x,z)\Big(\frac{\partial^2 c(x,z)}{\partial_{x_i}\partial_{x_j}} - \frac{\partial_{x_i}c(x,z)\partial_{x_j}c(x,z)}{\varepsilon}\Big) + \frac{\partial_{x_j}\mathbf{u}_\varepsilon(x)}{\varepsilon}z_i g_\varepsilon(x,z)d\nu(z),\\ &= \int G(x,z)\Big(\frac{\partial^2 c(x,z)}{\partial_{x_i}\partial_{x_j}} - \frac{\partial_{x_i}c(x,z)\partial_{x_j}c(x,z)}{\varepsilon}\Big)d\nu(z) + \frac{1}{\varepsilon}\partial_{x_i}\mathbf{u}_\varepsilon(x)\partial_{x_j}\mathbf{u}_\varepsilon(x).\end{aligned}$$

where we also used the explicit derivatives of $\mathbf{u}_\varepsilon$ from (3.21). The Hessian of $\mathbf{u}_\varepsilon^{c,\varepsilon}$ follows by symmetry. □

Remark 3.13. *One can find e.g. in [61] that first-order derivatives of the cost c are given by*

$$\partial_{x_i} c(x,z)\partial_{x_j} c(x,z) = \frac{d(x,z)^2}{1-\langle x,z\rangle^2} z_i z_j,$$

and that second-order derivatives write

$$\frac{\partial^2 c(x,z)}{\partial_{x_i}\partial_{x_j}} = d(x,z)\frac{\langle x,z\rangle}{\sqrt{1-\langle x,z\rangle^2}}\Big(\mathbb{1}_{i=j} - \frac{1}{1-\langle x,z\rangle^2} z_i z_j\Big) + \frac{1}{1-\langle x,z\rangle^2} z_i z_j.$$

Before turning to the description of the empirical counterparts in the next section, twice differentiability of $\mathbf{u}_\varepsilon$ allows to infer regularity for the associated sequence of spherical harmonics, as already discussed in Section 3.2.1.

Proposition 3.14. *Let $\mathbf{u}_\varepsilon$ be a solution of (S_ε). Then, $\mathbf{u}_\varepsilon$ is twice continuously differentiable, and, as a byproduct, its series of spherical harmonics belongs to ℓ_1.*

Proof. From Proposition 3.12, $\mathbf{u}_\varepsilon$ is twice-differentiable everywhere, that gives us the continuity of $\mathbf{Q}_\varepsilon$, and of $\partial_{x_i}\mathbf{u}_\varepsilon$. In the expression of the second-order partial derivatives given in (3.12), the term $\frac{1}{\varepsilon}\partial_{x_i}\mathbf{u}_\varepsilon(x)\partial_{x_j}\mathbf{u}_\varepsilon(x)$ is thus continuous. The remaining term takes the form of a parameter-dependant integral, whose integrand is continuous and bounded. Thus, the result follows by a direct application of the theorem for continuity under the integral sign, and by using the property that a sequence of spherical harmonics belongs to ℓ_1 for functions that are twice continuously differentiable, see [96][Theorem 2]. □

3.2.3 Regularized estimators

Suppose that the estimator $\widehat{\mathbf{u}}_{\varepsilon,n}$, defined in (3.17), has been computed using the stochastic algorithm (3.15) from *i.i.d.* observations $X_1, \cdots, X_n$ sampled from ν supported on $\mathbb{S}^2$. To obtain a regularized quantile function, the empirical counterpart of Corollary 3.10 involves integrals with respect to $\mu_{\mathbb{S}^2}$ to compute the smooth conjugate of $\widehat{\mathbf{u}}_{\varepsilon,n}$. Therefore, we consider a random sample $U_1, \cdots, U_N$ uniformly drawn on $\mathbb{S}^2$, and we define the following estimator (as an approximation of $\widehat{\mathbf{u}}^{c,\varepsilon}_{\varepsilon,n}$)

$$\widehat{\mathbf{u}}^{c,\varepsilon}_{N,n}(z) = -\varepsilon \log \frac{1}{N}\sum_{i=1}^{N} \exp\Big(\frac{\widehat{\mathbf{u}}_{\varepsilon,n}(U_i) - c(U_i,z)}{\varepsilon}\Big). \tag{3.31}$$

Hence, plugging (3.31) into (3.27), we propose the following estimator, for the regularized quantile function $\mathbf{Q}_\varepsilon$ defined in Corollary 3.10,

$$\hat{\mathbf{Q}}^{\varepsilon}_{N,n}(x) = \mathrm{Exp}_x\Big(\sum_{i=1}^{n} \hat{g}^{\varepsilon}_{N,n}(x,X_i)\mathrm{Log}_x(X_i)\Big), \tag{3.32}$$

where

$$\hat{g}^{\varepsilon}_{N,n}(x,z) = \frac{\exp\left(\frac{\widehat{\mathbf{u}}^{c,\varepsilon}_{N,n}(z)-c(x,z)}{\varepsilon}\right)}{\sum_{j=1}^{n}\exp\left(\frac{\widehat{\mathbf{u}}^{c,\varepsilon}_{N,n}(X_j)-c(x,X_j)}{\varepsilon}\right)}.$$

In the same token, an estimator of $\mathbf{F}_{\varepsilon}$ is given by

$$\hat{\mathbf{F}}^{\varepsilon}_{N,n}(z) = \mathrm{Exp}_z\Big(\sum_{i=1}^{N}\tilde{g}^{\varepsilon}_{N,n}(U_i,z)\mathrm{Log}_z(U_i)\Big), \tag{3.33}$$

with

$$\tilde{g}^{\varepsilon}_{N,n}(x,z) = \frac{\exp\left(\frac{\widehat{\mathbf{u}}_{\varepsilon,n}(x)-c(x,z)}{\varepsilon}\right)}{\sum_{j=1}^{N}\exp\left(\frac{\widehat{\mathbf{u}}_{\varepsilon,n}(U_j)-c(U_j,z)}{\varepsilon}\right)}.$$

Note that the empirical version of unregularized MK quantiles that is proposed in [84] relies on discrete OT, yielding a bijection between a reference grid of points in $\mathbb{S}^2$ and the samples. This is beneficial for statistical testing where distribution-freeness of the ranks is highly desirable. On the contrary, regularization yields, even empirically, smooth maps that are not constrained to belong to the set of observed data, which is crucial for the descriptive analysis of Section 3.3.

Besides, the estimation of contours in [84] requires to solve two different discrete OT problems. The first one estimates the central point $\mathbf{F}(\theta_M)$, whereas the second involves a grid oriented towards the estimate of $\mathbf{F}(\theta_M)$, to render MK contours. On the contrary, with our algorithm targeting continuous OT, there is no need to solve two different OT problems, as the estimate $\widehat{\mathbf{u}}_{\varepsilon,n}$ yields both $\hat{\mathbf{F}}^{\varepsilon}_{N,n}$ and $\hat{\mathbf{Q}}^{\varepsilon}_{N,n}$, and a fortiori $\hat{\mathbf{F}}^{\varepsilon}_{N,n}(\theta_M)$.

3.3 Depth-based data analysis

This section is dedicated to study a companion concept of directional MK quantiles, the MK statistical depth. We state a directional definition and discuss its properties. After that, we introduce descriptive tools in the spirit of the ones presented in [112] in the euclidean setting.

For the sake of completeness, we first study the Euclidean setting $\mathbb{R}^d$ before the directional one, that is of particular interest for us. Indeed, the results that we derive below do not appear as such in the literature, at least to the best of our knowledge.

3.3.1 Euclidean setting

From the euclidean definitions of MK quantile and distribution functions, see Definition 1.2, the MK depth is defined in $\mathbb{R}^d$ as follows, [33].

Definition 3.15. *Let ν be an arbitrary probability measure on $\mathbb{R}^d$. Its MK quantile function is the unique $\mathbf{Q} = \nabla\psi$ for some convex $\psi : \mathbb{R}^d \to \mathbb{R}$ such that $\mathbf{Q}_{\#}U_d = \nu$. Then, the MK depth of $x \in \mathbb{R}^d$ is the depth of $\nabla\psi^*$ under Tukey's depth, [169],*

$$D_\nu(x) = D_{U_d}^{Tukey}\Big(\nabla\psi^*(x)\Big).$$

The Liu-Zuo-Serfling axioms [111, 176], describe desirable properties for depth concepts. The MK-depth softens some of them, to reach more relevant contours [33]. Firstly, MK depth corresponds to Tukey depth for elliptical families [33]. Moreover, it benefits from invariance properties [74][Lemmas A.7,A.8], with respect to scaling (multiplication by a positive constant), translations, and orthogonal transformations (multiplication by an orthogonal matrix). Note that the affine-invariance does not hold. Another axiom is the *linear monotonicity relative to the deepest points*, that is $D_\nu(x) \leq D_\nu((1-t)x_0 + tx)$ for all $t \in [0,1]$ if x_0 is a deepest point. This is not fulfilled by the MK depth [33], although it verifies a similar property along sign curves, as we shall see now.

Proposition 3.16 (Curvilinear monotonicity relative to the deepest points)**.** *Assume that ν is continuous. The MK depth is monotonically decreasing along sign curves, that is, for each $u \in \mathbb{B}(0,1)$ and $t \in [0,1]$,*

$$D_\nu(\mathbf{Q}(u)) \leq D_\nu(\mathbf{Q}(tu)).$$

Proof. Recall that Tukey's depth verifies linear monotonicity relative to the deepest points [176]. As the origin is the deepest point for U_d, this writes, for any $t \in [0,1]$,

$$D_{U_d}^{Tukey}(u) \leq D_{U_d}^{Tukey}(tu). \tag{3.34}$$

From Definition 3.15, $D_\nu(\mathbf{Q}(u)) = D_{U_d}^{Tukey}(\nabla\psi^* \circ \mathbf{Q}(u))$. By continuity of ν, $\nabla\psi^* \circ \mathbf{Q}(u) = u$ *a.e.*, see [74] or [172][Theorem 2.12 and Corollary 2.3]. Thus the result follows, with (3.34). □

This corresponds to the classical linear monotonicity under distributions with straight sign curves, including spherical families due to the particular form of the MK quantile function in this setting, taken from [33].

Corollary 3.17. *For spherically symmetric distributions, sign curves are straight lines, and the MK depth verifies linear monotonicity relative to the deepest point. For any x in the support of ν, for $x_0 = \mathbb{E}(X)$ the deepest point of ν,*

$$\forall t \in [0,1], D_\nu(x) \leq D_\nu((1-t)x_0 + tx). \tag{3.35}$$

Proof. Let X be a random vector associated with a spherically symmetric distribution, for which $\mathbb{E}(X)$ and the deepest point shall coincide. From [33], the MK

distribution function of X is known. By inverting it, we get its quantile function

$$\mathbf{Q}(u) = \frac{u}{\|u\|} G^{-1}(\|u\|) + \mathbb{E}(X),$$

where G is the univariate distribution function of the radial part $\|X - \mathbb{E}(X)\|$. Because $\|X\| \geq 0$ *a.s.* and G^{-1} is increasing, $G^{-1}(t\|u\|)/G^{-1}(\|u\|) \in [0,1]$ and

$$\mathbf{Q}(tu) = \frac{u}{\|u\|} G^{-1}(t\|u\|) + \mathbb{E}(X) = \frac{G^{-1}(t\|u\|)}{G^{-1}(\|u\|)} \Big(\mathbf{Q}(u) - \mathbb{E}(X) \Big) + \mathbb{E}(X).$$

This rewrites, for $\delta_t = G^{-1}(t\|u\|)/G^{-1}(\|u\|)$, $\mathbf{Q}(tu) = \delta_t \mathbf{Q}(u) + (1-\delta_t)\mathbb{E}(X)$. Besides, δ_t takes all values between 0 and 1 for $t \in [0,1]$. This, combined with Proposition (3.16) induces

$$\forall u \in \mathbb{B}(0,1), \forall \delta \in [0,1], D_\nu(\mathbf{Q}(u)) \leq D_\nu(\delta \mathbf{Q}(u) + (1-\delta)\mathbb{E}(X)).$$

But any x in the support of ν writes $\mathbf{Q}(u)$ for $u = \mathbf{F}(x)$, which gives (3.35). □

We now turn to the properties of the directional MK depth.

3.3.2 Directional setting

Using the same ideas than in [33], one can define the MK depth on the sphere through any statistical depth with respect to the uniform $\mu_{\mathbb{S}^2}$ oriented towards $\mathbf{F}(\theta_M)$. The simplest is certainly to consider the proximity with $\mathbf{F}(\theta_M)$.

Definition 3.18. *Let ν be an arbitrary probability measure on $\mathbb{S}^2$, with directional distribution function $\mathbf{F}$. The directional MK depth of $x \in \mathbb{S}^2$ is defined by*

$$D_\nu(x) = 1 - d(\mathbf{F}(x), \mathbf{F}(\theta_M))/\pi.$$

Regarding the $\mathbb{S}^2$-adapted versions of Liu-Zuo-Serfling axioms, [111, 176], the directional MK depth behaves like its Euclidean counterpart. We begin with the four classical properties that are direct spherical counterparts of the Euclidean axioms, see *e.g.* [108]. The affine-invariance is replaced on $\mathbb{S}^2$ by rotational invariance, which holds true from [84], see also Proposition B.1. Moreover, it is straightforward that D_ν attains its maximum at the center $\mathbf{F}(\theta_M)$, and that it vanishes at $-\mathbf{F}(\theta_M)$, the spherical counterpart of infinity. Finally, monotonicity along great circles is not fulfilled, but it is replaced in the same data-adaptive fashion than in $\mathbb{R}^d$.

Proposition 3.19 (Curvilinear monotonicity relative to the deepest points)**.** *Assume that ν is continuous. The directional MK depth is monotonically decreasing along sign curves. For each $x \in \mathbb{S}^2$ and $t \in [\langle x, \mathbf{F}(\theta_m)\rangle, 1]$, let $x_t \in \mathcal{M}_s^U$, for $s = \mathbf{S}_{\mathbf{F}(\theta_M)}(x)$, such that*

$$x_t = t\mathbf{F}(\theta_M) + \sqrt{1-t^2}s. \tag{3.36}$$

Then,

$$D_\nu(\mathbf{Q}(x)) \leq D_\nu(\mathbf{Q}(x_t)). \tag{3.37}$$

Proof. Fix $x \in \mathbb{S}^2$. Let $s = \mathbf{S}_{\mathbf{F}(\theta_M)}(x)$ be the directional sign associated to x, from the decomposition (3.10). For $t \in [-1, 1]$, let $x_t \in \mathcal{M}_s^U$ be a parameterization of the reference sign curve associated to s, as in (3.36). Hence, one may note that $\langle x_t, \mathbf{F}(\theta_M)\rangle = t$, and $x_t = x$ for $t = \langle x, \mathbf{F}(\theta_M)\rangle$. Besides, $D_\nu(\mathbf{Q}(x_t)) = 1 - d(x_t, \mathbf{F}(\theta_M))/\pi = 1 - \arccos(t)/\pi$. Thus, as soon as $t \geq \langle x, \mathbf{F}(\theta_m)\rangle$, (3.37) holds. □

Explicit formulations for rotationally invariant distributions are given in [84], and recalled in Appendix B.2. As a direct byproduct, the multivariate optimal transport only acts on $\langle x, \theta_M\rangle$, thus it reduces to univariate optimal transport along the axis $\pm\theta_M$. In particular, for such distributions, MK quantile contours coincide with Mahalanobis ones, [108], so that the following is straightforward.

Corollary 3.20. *For rotationally invariant distributions, sign curves are great circles. Thus, the MK depth verifies linear monotonicity along great circles, relative to the deepest point.*

Other desirable axioms have been put forward recently in [129, 130], namely the *upper semi-continuity* and the *non-rigidity of central regions*. *Upper semi-continuity* is ensured to hold as soon as the MK distribution function $\mathbf{F}$ is continuous, thus at least for $\nu \in \mathbf{B}_2$. Even more, when ν is arbitrary, taking the regularized directional MK depth built from $\mathbf{F}_\varepsilon$, for $\varepsilon > 0$, imposes continuity, which may motivate such regularized estimator. Lastly, the *non-rigidity of central regions* states that quantile regions are not restricted to be spherical caps, which is readily true for the MK depth. In fact, its adaptivity to the underlying support is one of its main feature, and it can be seen as a stronger *non-rigidity* axiom, requiring that $\mathbf{Q}(U) \sim \nu$ as soon as $U \sim \mu$. Furthermore, Proposition 3.19 and Corollary 3.20 shed some light on the non-verified axiom of *monotony along great circles*. Our results suggest that the directional MK depth alleviates these axioms when necessary, *e.g.* for complex distributions such as mixtures, whereas the axioms are fulfilled for distributions for which it is useful, in particular for rotationally invariant ones.

3.3.3 Descriptive tools

The seminal paper [112] gathers descriptive tools based on data depths. Monge-Kantorovich analogs already exist for data in $\mathbb{R}^d$, and we shall now extend some of them to the directional setting. We stress that the ability of our regularized estimator to interpolate between data points is crucial for (i) smooth contours in practice and (ii) computing volumes of quantile regions.

Representative plots

Firstly, [112] study representative plots for bivariate data, and the MK analog is given by the descriptive plots from [84, 87], the latter with the added information

of sign curves. Figure 3.1 illustrates it on a Tangent von-Mises Fisher distribution, [68], and on a Mixture of two von-Mises Fisher distributions, with the help of our empirical regularized quantile function. One can observe that the shapes of the distributions are well recovered.

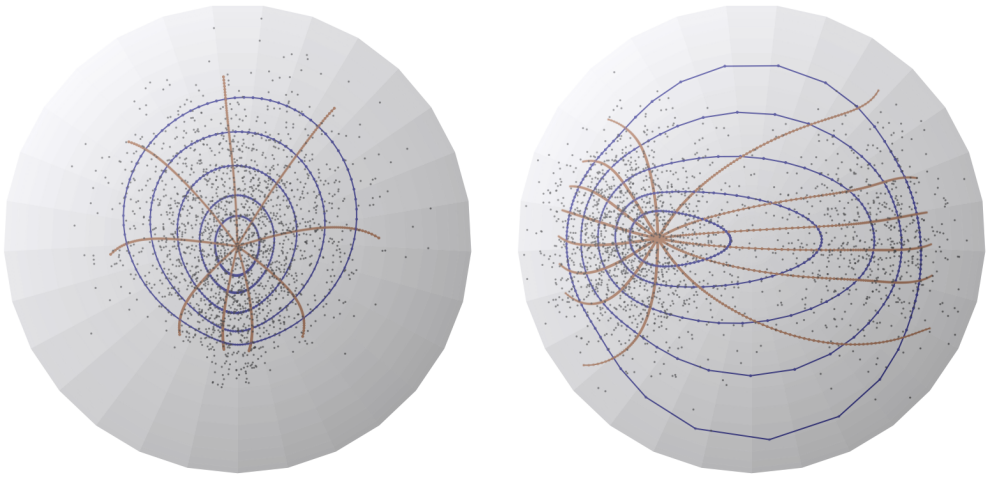


FIGURE 3.1 – Regularized quantile contours of levels $\{0.1, 0.25, 0.5, 0.75, 0.9\}$ and associated sign curves, with $\epsilon = 10^{-1}$.

Scale or dispersion

Hereafter, we present a graphical tool to describe the amount of dispersion, called the *scale curve* in [112] and whose MK analog has been introduced in [15] for Euclidean data. We mention that, in [112], this tool is also used to compare the variance of vector-valued estimators. Put simply, given any level $\alpha \in [0, 1]$, we consider the volumes $V(\alpha)$ of MK quantile regions. Plotting such volumes with respect to $\alpha \in [0, 1]$ yields a *scale curve* [112]. The faster it grows, the greater the dispersion. Thus, if the scale curve of ν_1 is consistently above the one of ν_2, then ν_1 is more spread out than ν_2. On $\mathbb{S}^2$, the volume is bounded, so we consider the normalized $\mu_{\mathbb{S}^2}$ instead of $\sigma_{\mathbb{S}^2}$. Define

$$V(\alpha) = \int_{\mathcal{C}_\alpha} d\mu_{\mathbb{S}^2}(x) = \int_{\mathbb{S}^2} \mathbb{1}_{\{x \in \mathcal{C}_\alpha\}} d\mu_{\mathbb{S}^2}(x) = \int_{\mathbb{S}^2} \mathbb{1}_{\{\langle \mathbf{F}(x), \mathbf{F}(\theta_M) \rangle \geq 1-2\alpha\}} d\mu_{\mathbb{S}^2}(x).$$

This can be estimated with a sample $U_1, \cdots, U_N$ from $\mu_{\mathbb{S}^2}$, by the proportion

$$V_{\varepsilon,n}(\alpha) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{\langle \hat{\mathbf{F}}^{\varepsilon}_{N,n}(U_i), \mathbf{F}(\theta_M) \rangle \geq 1-2\alpha\}}.$$

On the left-hand side of Figure 3.2, we draw the scale curves of von-Mises Fisher distributions with varying concentration parameter $\kappa \in \{1, 2, 5, 15\}$. It is well-captured that the lower the value of κ, the more spread out is the underlying

distribution. Besides, univariate order statistics (and equivalently, quantiles) are fundamental to analyse the presence of outliers. In our spherical setting, the scale curve is able to conveniently summarize this type of information, as illustrated in the right-hand side of Figure 3.2. We consider $n = 500$ observations coming from three identical von-Mises Fisher distributions with dispersion parameter $\kappa = 15$ and mean $(0, 1, 0)^T$, but each with a certain number $N \in \{5, 20, 50\}$ of outliers localized near from $(0, 0, 1)^T$. It appears that the volumes of quantile regions up to the order $\alpha \approx 0.6$ are identical, whereas the dispersion increases with the number of outliers for peripheric quantile regions, which is precisely the expected behavior.

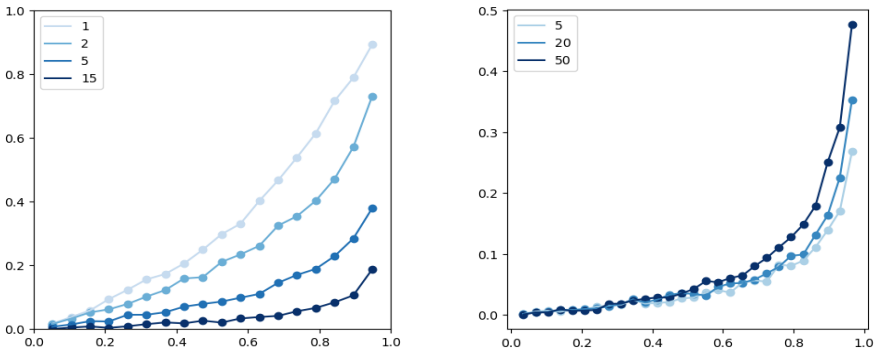
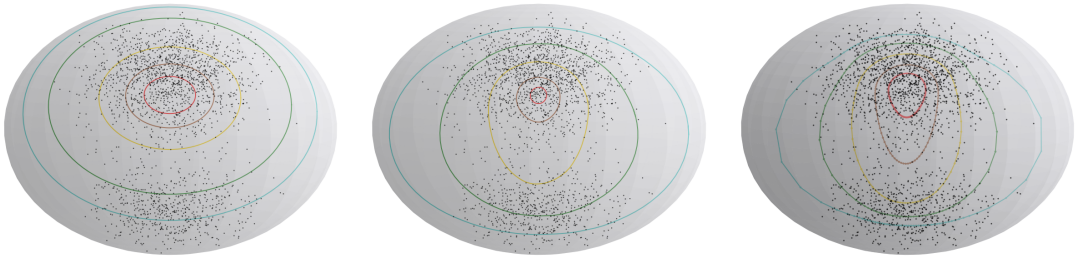


FIGURE 3.2 – Scale curves of von-Mises Fisher distributions with various (left) dispersion parameter κ and (right) number N of added outliers.

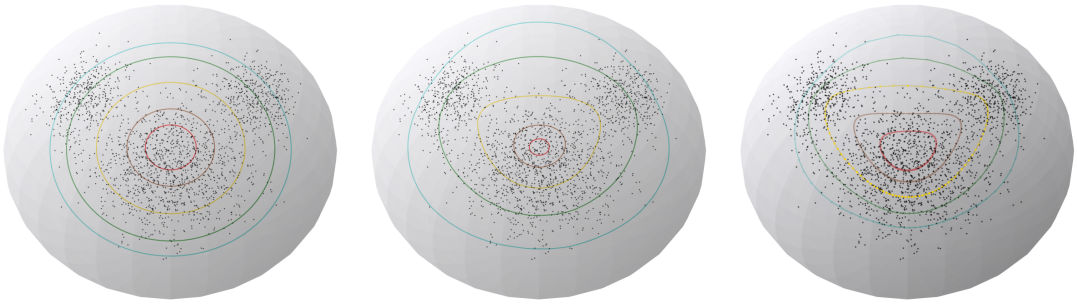
3.4 Numerical experiments

Other concepts of quantiles

In Figure 3.3, we display a visual comparison of existing concepts for quantiles on the sphere by focusing on mixtures of two (first row) and three (second row) von-Mises Fisher distributions. It can be observed that Mahalanobis quantile regions [108] are concentric spherical caps, whereas spatial quantiles [102] and our regularized MK quantiles can exhibit more complex shapes that better fit the geometry of the data. To that extent, spatial and regularized MK quantiles, obtained through our regularized estimator $\mathbf{Q}_\varepsilon$, are both more satisfactory. For the spatial quantiles, our naive implementation forces them to belong to data samples. For each notion of quantiles, 100 points are drawn within each contours, with straight lines to link them. We emphasize that spatial quantiles are not indexed by their probability content, as opposed to the MK ones [84]. Because entropic MK quantiles interpolate between data points, contours cross the void between mixture components. A careful inspection shows that the number of points per contour within this void is much lower than in the high density areas. This illustrates how the variation of mass, that is the underlying geometry, is captured by our regularized estimator.



(a) Mahalanobis quantiles (b) Spatial quantiles (c) Monge-Kantorovich quantiles



(d) Mahalanobis quantiles (e) Spatial quantiles (f) Monge-Kantorovich quantiles

FIGURE 3.3 – Quantile contours of orders $\{0.1, 0.25, 0.5, 0.75, 0.9\}$

Influence of the regularization strength

In Figure 3.4, we study the influence of the regularization parameter on the estimation of known quantile contours, that were described in [84] and that are recalled in Appendix B.2 for the sake of completeness. The mean-squared error from uniform samples $(x_i) \subset \mathbb{S}^2$ is

$$\mathcal{R}_n(\widehat{Q}) = \frac{1}{n}\sum_{i=1}^{n} c(Q(x_i), \widehat{Q}(x_i)),$$

for $\widehat{Q}$ denoting either our regularized MK quantile estimator $\mathbf{Q}_\varepsilon$ or the unregularized one $\widehat{\mathbf{Q}}_0$ proposed in [84]. Samples of size $n = 500$ are drawn from the uniform $\mu_{\mathbb{S}^2}$ and from a von-Mises fisher distribution of location $(0, 0, 1)^T$ and concentration $\kappa = 10$. For several values of ε, $\mathbf{Q}_\varepsilon$ and $\widehat{\mathbf{Q}}_0$ are computed, and compared to the ground truth by $\mathcal{R}_n(\widehat{Q})$.

By doing this experiment 50 times, we obtain a boxplot of MSE values for each ε, that are reported in Figure 3.4 where $\varepsilon = 0$ refers to $\widehat{\mathbf{Q}}_0$. The dashed horizontal line illustrates the median value for the MSE of $\widehat{\mathbf{Q}}_0$. It can be observed that the entropic regularization is able to significantly outperform the estimation of the

quantile map, in particular for values around $\varepsilon \approx 0.09$.

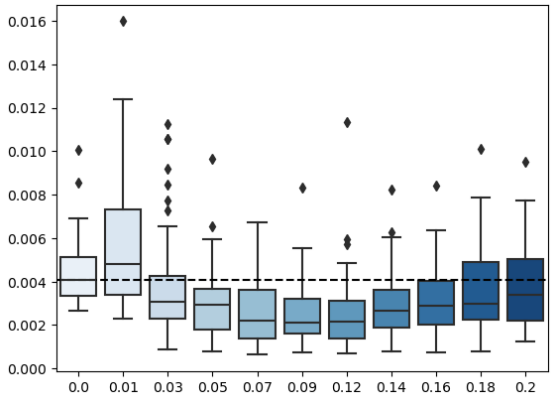
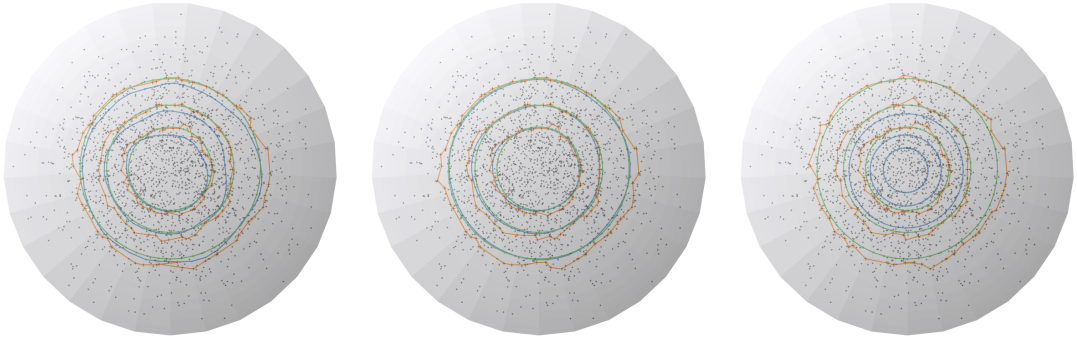


FIGURE 3.4 – Mean squared error $\mathcal{R}_n(\widehat{\mathbf{Q}}_\varepsilon)$ as a function of $\varepsilon \in [0, 0.2]$.

In Figure 3.5, we visually compare regularized and unregularized MK spherical quantile contours of orders $24.4\%, 48.8\%, 75.6\%$ on the same von-Mises fisher distribution than in Figure 3.4. Such uncommon probability contents are inherent to empirical unregularized contours because the number of contours as well as their size depends on the sample size, that is here fixed at $n = 2001$. Ground-truth contours deduced from (B.3) are presented together with unregularized and regularized ones, for $\varepsilon \in \{0.01, 0.05, 1\}$. Each contour contains 100 points, linked by straight lines. For $\varepsilon = 0.01$, contours adapt too much on the finite-sample data, causing errors as ground-truth contours are smoother. For $\varepsilon = 1$, contours are smoother, but there is too much bias in the approximation between $\widehat{\mathbf{Q}}_\varepsilon$ and the underlying ground truth. For the well-chosen $\varepsilon = 0.05$, the trade-off between regularity and low-bias allows the better estimation. This sheds some light on the behavior of regularization. The lower the ε, the more adapted $\widehat{\mathbf{Q}}_\varepsilon$ is to the finite-sample data and its irregularities. Larger values of ε induce smoother contours, as a byproduct of a greater regularity for $\widehat{\mathbf{Q}}_\varepsilon$. Thus, this emphasizes the need for calibration of the regularization strength.

Max-depth classification

Furthermore, following experiments in [102], we report a *quantitative comparison* between different notions of directional quantiles on the task of supervised classification, based on the *max-depth* approach from [75]. Because statistical depths are designed to measure *outlyingness* within a distribution, one can classify $x \in \mathbb{S}^2$ as coming from the distribution ν_1 instead of ν_2 if the depth of x with respect to (*w.r.t.*) ν_1 is greater than the one *w.r.t.* ν_2. As highlighted in [102], the vanishing property of many statistical depths [113, 169], raises issues when the sample x lies outside the convex hull of empirical data, where new data is classified randomly. On the contrary, the depth associated to spatial quantiles [102][Section 5], is strictly positive everywhere, and the same holds trivially for the MK depth. In each



(a) $\varepsilon = 0.01$ (b) $\varepsilon = 0.05$ (c) $\varepsilon = 1$

FIGURE 3.5 – Empirical regularized quantile contours in blue, unregularized ones in orange, and ground truth in green.

setting, we learn the quantile functions on 200 points coming from distributions (P_1, P_2), before computing misclassification rates on a test data of 200 different observations. Define O the rotation matrix fixing the first coordinate and mapping $m_1 = (0, 0, 1)^T$ to $m_2 = (0, \sin(\pi/3), \cos(\pi/3))^T$. We consider for (P_1, P_2) :

1. *von-Mises Fisher :* P_1 is a vMF with location m_1 and concentration $\kappa = 3$, P_2 is the distribution of OX for $X \sim P_1$.
2. *Tangent von-Mises Fisher :* P_1 is the distribution of

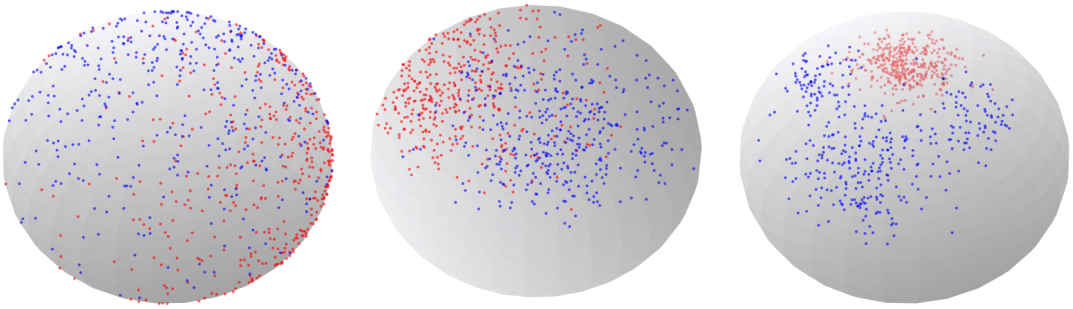
$$X = Zm_1 + \sqrt{1 - Z^2}S,$$

where $Z = 2V - 1$ for V a Beta$(2, 8)$ distribution, $S_3 = 0$ and $(S_1, S_2)^T$ follows a von-Mises Fisher distribution with location $(1, 0)^T$ and concentration $\kappa' = 5$. Again, P_2 is the distribution of OX for $X \sim P_1$.

3. *Non-convex data :* P_1 is the same mixture of three vMF distributions than in Figure 3.3. P_2 is a vMF distribution centered at $(0, 0, 1)^T$ with concentration parameter $\kappa = 50$.

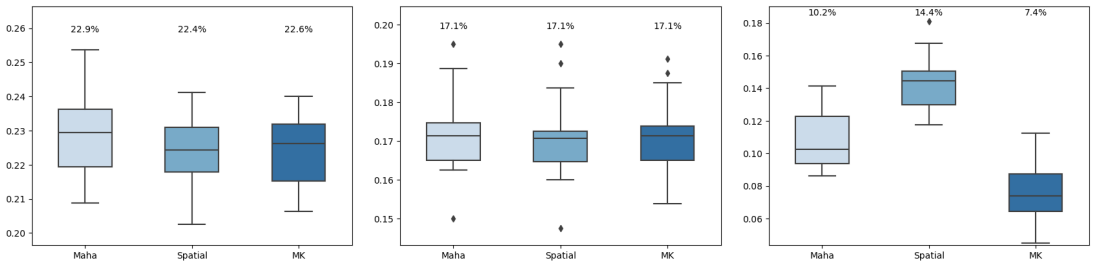
Examples of data sampled from these three distributions are displayed in Figure 3.6, with observations sampled from P_1 in blue and P_2 in red.

For each notion of depth, we run 20 times the classification, and we gather the misclassification rates. Figure 3.7 gives the results for the MK depth with $\varepsilon = 10^{-1}$ in comparison with existing notions of quantiles. All methods provide comparative results when both the supports of P_1 and P_2 are convex. However, in the third setting, classification based on the MK depth significatively outperforms the others. This illustrates the potential of entropically regularized quantiles for inference.



(a) von-Mises Fisher (b) Tangent von-Mises Fisher (c) Non-convex data

FIGURE 3.6 – Data for max-depth classification.



(a) Von-Mises Fisher (b) Tangent von-Mises Fisher (c) Non-convex data

FIGURE 3.7 – Misclassification rates with several statistical depths.

4

Superquantiles and expected shortfalls

"*The superquantile function [...] is as fundamental to a random variable as the distribution and quantile functions.*"

Rockafellar & Royset, [145]

This chapter defines center-outward superquantile and expected shortfall functions, to provide a natural way to characterize multivariate tail probabilities and central areas of point clouds. These notions characterize the underlying distributions and their weak convergence, which underlines their importance. Applications in risk analysis are provided, with multivariate extensions of the standard notions of value-at-risk and conditional-value-at-risk. This chapter is based on [18].

Contents

4.1 Introduction

Superquantile, expected shortfall

Modeling the dependency between the components of a random vector is at the core of multivariate statistics. To that end, one way to proceed is to characterize the multivariate probability tails. For distributions supported on the real line, this is often tackled with the use of superquantiles or expected shortfalls, that complement the information given by the quantiles. Let X be an integrable absolutely continuous random variable with cumulative distribution function F. For all $\alpha \in]0,1[$, recall that the quantile $Q(\alpha)$ of level α is given by

$$Q(\alpha) = \inf\{x : F(x) \geq \alpha\},$$

whereas the superquantile $S(\alpha)$ and expected shortfall $E(\alpha)$ are defined by

$$S(\alpha) = \mathbb{E}[X \big| X \geq Q(\alpha)] = \frac{\mathbb{E}[X\mathbb{1}_{X \geq Q(\alpha)}]}{\mathbb{P}(X \geq Q(\alpha))} = \frac{1}{1-\alpha}\mathbb{E}[X\mathbb{1}_{X \geq Q(\alpha)}], \tag{4.1}$$

and

$$E(\alpha) = \mathbb{E}[X \big| X \leq Q(\alpha)] = \frac{\mathbb{E}[X\mathbb{1}_{X \leq Q(\alpha)}]}{\mathbb{P}(X \leq Q(\alpha))} = \frac{1}{\alpha}\mathbb{E}[X\mathbb{1}_{X \leq Q(\alpha)}]. \tag{4.2}$$

As illustrated in Figure 4.1, $S(\alpha)$ focuses on the upper-tail while $E(\alpha)$ targets the lower-tail. We emphasize that using the terms of superquantile and expected shortfall is a subjective consideration taken from [1, 145]. Most of the time, one does not consider the upper and the lower tails together, so that a single name is required, up to considering the distribution of $-X$. In this vein, depending on the application, the expected shortfall may refer to the same as the Conditional-Value-at-Risk, Conditional-Tail-Expectation, see *e.g.* [2], or even the superquantile, that aims to be a neutral alternative name in statistics [145].

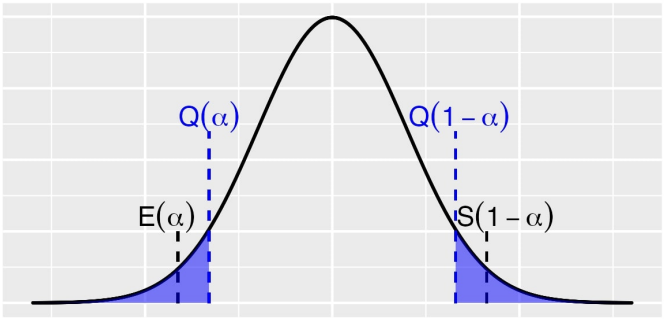


FIGURE 4.1 – Illustration of the notions of superquantile $S(\alpha)$ and expected shortfall $E(\alpha)$ for an univariate Gaussian distribution.

The main contribution of the present chapter is to extend (4.1) and (4.2) towards a notion of multivariate superquantile and expected shortfall. As part of the difficulty, both the mathematical meanings of "ahead", "beyond" and "typical"

do not adapt canonically in $\mathbb{R}^d$. We argue that sufficient notions are provided by the Monge-Kantorovich (MK) quantiles, ranks and signs, introduced in [33, 87]. In particular, the traditional left-to-right ordering is replaced in our approach by a center-outward one that is more intuitive for a point cloud [33]. Hence, the two subsets of observations that we are interested in are located at the outward or near the mean value, which requires to adapt the concepts of (4.1) and (4.2) in $\mathbb{R}^d$. It is well-known from a simple change of variables in $\mathbb{R}$ that S and E average observations beyond and ahead the quantile of level α, in the sense that

$$S(\alpha) = \frac{1}{1-\alpha}\int_\alpha^1 Q(t)dt \qquad \text{and} \qquad E(\alpha) = \frac{1}{\alpha}\int_0^\alpha Q(t)dt. \tag{4.3}$$

In Section 4.2, our definitions generalize the formulation (4.3). If Q stands for the multivariate MK quantile function instead of the classical univariate one, center-outward superquantile and expected shortfall functions are defined, for any u in the unit ball $\mathbb{B}(0,1)\backslash\{0\}$, by

$$S(u) = \frac{1}{1-\|u\|}\int_{\|u\|}^1 Q\Big(t\frac{u}{\|u\|}\Big)dt \qquad \text{and} \qquad E(u) = \frac{1}{\|u\|}\int_0^{\|u\|} Q\Big(t\frac{u}{\|u\|}\Big)dt.$$

Importantly, the univariate quantile and related functions are deeply rooted in risk analysis. On the one hand, risk measures which are both coherent and regular can be characterized by integrated quantile functions [79]. On the other hand, given a level α, fundamental risk measures are given by $Q(\alpha)$ and $S(\alpha)$, called Value-at-Risk (VaR) and Conditional-Value-at-Risk (CVaR), respectively. As a matter of fact, a natural main contribution of the present paper is to provide meaningful multivariate extensions of VaR and CVaR.

Background on multivariate risk measurement

MK quantiles and the related concepts have already been applied to multivariate risk measurement in [15], and, in some sense, in the previous works [59, 67]. The theory in [59] states ideal theoretical properties for coherent regular risk measures, while the maximal correlation risk measure of [15] furnishes a real-valued risk measure with these properties. This constitutes, to the best of our knowledge, the short literature on risk measurement based on the MK quantile function. In this work, we argue that an adequate procedure of multivariate risk measurement shall account for all the information on the tails, both in terms of direction and spreadness. To answer this issue, vector-valued risk measures are natural candidates. There also exist several extensions of VaR or CVaR to the multivariate setting, including [6, 29, 40, 41, 76, 91, 142, 167, 170], but none of them is based on the theory of optimal transportation and its associated potential benefits. In particular, our concepts do not require any assumption on the tail behavior of the data, nor any statistical model, because MK quantiles adapt naturally to the shape

of a point cloud. On the real line, the VaR and the CVaR have a clear interpretation : for a level $\alpha \in [0,1]$, the VaR is the worst observation encountered with probability $1-\alpha$ whereas CVaR is the average value beyond this worst observation. Such a meaningful definition is surely part of the reason for their wide use in practice. Under the name of Conditional-Tail-Expectation, the idea proposed in [41, 57] preserves this interpretation, but relies on level sets defined from the theory of copulas. Specifically, the obtained quantile levels do not adapt automatically to the shape of the data. Still, this notion averages over a certain quantile level, and it returns a tail observation of the same dimension as the data. Our work is inspired by this approach, as we aim to give the same information about multivariate tails, but we use the MK quantile function, which yields, to our opinion, concepts with even better interpretability. A result of independent interest is also provided, giving a new family of Monge maps between known probability distributions. This may motivate changing the reference distribution under generalized gamma models.

4.2 Center-outward superquantiles and expected-shortfalls

4.2.1 Main definitions

On the real line, the notion of superquantile and expected shortfall relies heavily on the one of quantile. It is then natural to make use of the Monge-Kantorovich (MK) quantile function and its appealing properties in order to define associated superquantile and expected shortfall functions. In the remaining, the reference measure is taken as U_d, which is highlighted by the denomination of *center-outward* quantiles instead of *Monge-Kantorovich*. By simplicity, we shall restrict ourselves to the set of integrable probability measures over $\mathbb{R}^d$, that is

$$\mathcal{P}_1(\mathbb{R}^d) = \big\{\nu : \mathbb{E}_{X\sim\nu}[\|X\|] < +\infty\big\}.$$

We also make some assumptions to ensure continuity of $\mathbf{Q}$, that are described hereafter. Following [87], we assume, without loss of generality, that ψ in Definition 1.2 satisfies $\psi(0) = 0$ and, for $u \in \mathbb{R}^d$ such that $\|u\| = 1$,

$$\psi(u) = \liminf_{\substack{v\to u \\ \|v\|<1}} \psi(v). \tag{4.4}$$

Moreover, we impose, for all $u \in \mathbb{R}^d$ such that $\|u\| > 1$,

$$\psi(u) = +\infty. \tag{4.5}$$

This being said, the convex potential ψ is uniquely defined over its domain $\mathrm{Dom}(\psi) = \{u|\psi(u) < +\infty\}$, that verifies $\mathbb{B}(0,1) \subset \mathrm{Dom}(\psi) \subset \overline{\mathbb{B}}(0,1)$. Although ψ can be

chosen to be continuous over $\mathbb{B}(0,1)$, [146][Theorem 10.3], the gradient $\nabla\psi$ is only defined almost everywhere. At every u where ψ is not differentiable, one can still define the subdifferential

$$\partial\psi(u) = \{z \in \mathbb{R}^d : \forall x \in \mathbb{R}^d, \psi(x) - \psi(u) \geq \langle z, x - u\rangle\}.$$

Then, for all $u \in \mathbb{B}(0,1)$, we can define $\mathbf{Q}(u)$ as the average of $\partial\psi(u)$ so that $\mathbf{Q}$ is defined everywhere, without ambiguity. Note that considering any other point in the subdifferential $\partial\psi(u)$ would be as satisfactory. In fact, from [63], as soon as ν is a continuous probability measure with non vanishing density on $\mathbb{R}^d$, ψ can be shown to be differentiable everywhere on $\mathbb{B}(0,1)\backslash\{0\}$. The same result is showed under milder assumptions in [51], that are presented hereafter.

Assumption 4.1. *Let ν be an absolutely continuous measure with probability density function p defined on its support $\mathcal{X}$. For every $R > 0$, there exist two constants $0 < \lambda_R < \Lambda_R$ such that, for all $x \in \mathcal{X} \cap \mathbb{B}(0,R)$,*

$$\lambda_R \leq p(x) \leq \Lambda_R.$$

Assumption 4.2. *The support $\mathcal{X} \subset \mathbb{R}^d$ of ν is convex.*

Under these assumptions, the next theorem is given in [51].

Theorem 4.1 (Regularity of the center-outward quantile function, [51])**.** *Under Assumptions 4.1 and 4.2, there exists a compact and convex set K with Lebesgue measure 0 such that the center-outward quantile function $\mathbf{Q}$ is a homeomorphism from $\mathbb{B}(0,1)\backslash\{0\}$ to $\mathcal{X}\backslash K$, with inverse*

$$\mathbf{Q}^{-1}(x) = \nabla\psi^*(x) = \operatorname*{argsup}_{u\in\mathbb{B}(0,1)}\{\langle x, u\rangle - \psi(u)\}.$$

We also define $\mathcal{P}_*(\mathbb{R}^d)$ as the set of integrable probability measures for which Q is a homeomorphism from $\mathbb{B}(0,1)\backslash\{0\}$ to its image, see Theorem 4.1. This includes any ν satisfying assumptions 4.1 and 4.2, that is also integrable, to ensure finiteness of the center-outward superquantiles. We stress that, for any $\nu \in \mathcal{P}_*(\mathbb{R}^d)$, Q is defined everywhere on $\mathbb{B}(0,1)\backslash\{0\}$, as opposed to almost everywhere on $\mathbb{B}(0,1)$.

Hereafter, this allows us to properly define quantile contours, ranks and signs, where continuity and invertibility of $\mathbf{Q}$ are required. The next definitions, taken from [33, 87], gather the main concepts that we use in the following.

Definition 4.2 (Quantile contours, ranks and signs)**.** *Let ν with center-outward quantile function $\mathbf{Q}$ and supported on $\mathcal{X} \subset \mathbb{R}^d$. Then, for the distribution ν,*

(i) the quantile region $\mathbb{C}_\alpha$ of order $\alpha \in [0,1]$ is the image by $\mathbf{Q}$ of the ball $\mathbb{B}(0,\alpha)$,

(ii) the quantile contour $\mathcal{C}_\alpha$ of order $\alpha \in [0,1]$ is the boundary of $\mathbb{C}_\alpha$,

(iii) the rank function $\mathcal{R}_\nu : \mathcal{X} \to [0,1]$ *is defined by* $\mathcal{R}_\nu(x) = \|\mathbf{Q}^{-1}(x)\|$,

(iv) the sign function $\mathcal{D}_\nu : \mathcal{X} \to \mathbb{B}(0,1)$ *is defined by* $\mathcal{D}_\nu(x) = \mathbf{Q}^{-1}(x)/\|\mathbf{Q}^{-1}(x)\|$.

A few remarks follow from Definition 4.2. The ν-probability of $\mathbb{C}_\alpha$ is α, by the change of variables formula for push-forward maps, which is a first requirement for quantile regions. As already stressed, one may note that the rank and sign functions require the invertibility of $\mathbf{Q}$. Also, the ranks and signs are independent for any distribution ν. Furthermore, one can consider the *sign curve* associated to $u \in \mathbb{B}(0,1)$, that is the image by $\mathbf{Q}$ of the radius

$$L_u = \Big\{ t\frac{u}{\|u\|} : t \in [0,1] \Big\}.$$

These sign curves gather a sort of curvilinear directional information in a convenient fashion, that will be critical hereafter. Together with the quantile contours, they provide a curvilinear polar coordinate system adapted to a point cloud, as illustrated in Figure 4.2.

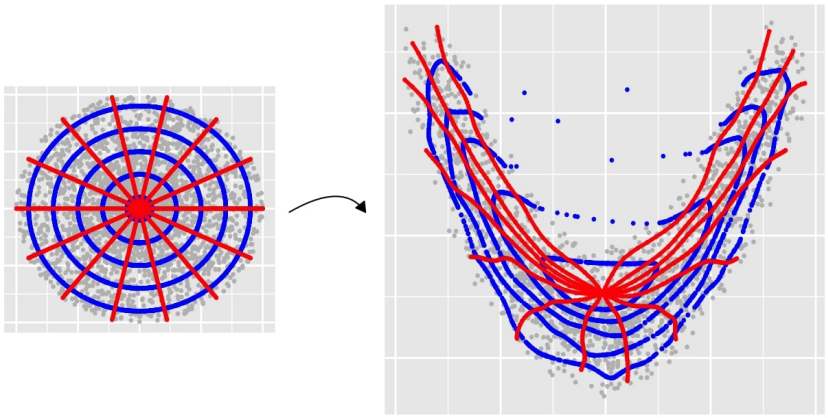


FIGURE 4.2 – (Left) center-outward quantiles of the spherical uniform distribution $\mu = U_d$, and (right) center-outward quantiles of a discrete distribution ν obtained by $\mathbf{Q}_\# \mu = \nu$.

From [87][Section 2], the continuity and invertibility of $\mathbf{Q}$ outside the origin ensures the crucial fact that the quantile contours are closed and nested. It is the notion of center-outward ranks that allows to order points in $\mathcal{X}$ relatively to ν, consistently with Tukey's halfspace depth, as highlighted in [33]. It induces the following weak order.

Definition 4.3. *For* $\nu \in \mathcal{P}(\mathbb{R}^d)$ *and* $x, y \in \mathcal{X}$*, we denote* $x \geq_\mathcal{R} y$ *and say that* y *is deeper than* x *if*

$$\|\mathbf{Q}^{-1}(x)\| \geq \|\mathbf{Q}^{-1}(y)\|.$$

The deeper a point is in $\mathcal{X}$, the more central it is with respect to ν. For a fixed $u \in \mathbb{B}(0,1)\backslash\{0\}$, we can then write $x \geq_\mathcal{R} \mathbf{Q}(u)$ which enables to consider the observations "beyond" $\mathbf{Q}(u)$ in the sense of being less central. Thus, a dataset

$X_1, \cdots, X_n$ can be ordered from the center to the outward, and relabeled to get the order statistics $X_{[1]} \leq_{\mathcal{R}} \cdots \leq_{\mathcal{R}} X_{[n]}$.

Focusing on the *central* areas of a point cloud, $u \mapsto \mathbb{E}[X|X \leq_{\mathcal{R}} Q(u)]$ has been introduced in [85] as a center-outward Lorenz function for X, see Section 4.6. The reverse conditional expectation $\mathbb{E}[X|X \geq_{\mathcal{R}} Q(u)]$ might be thought of as a natural candidate for a superquantile concept. Nevertheless, it cannot be understood as a "typical" observation within *outward* areas. Indeed, consider the following example.

Example 4.4. *Suppose that $\nu = U_d$, so that Q is the identity. For every $u \in \mathbb{B}(0,1)\backslash\{0\}$, $\mathbb{E}[X|X \geq_{\mathcal{R}} Q(u)] = \mathbb{E}\left[X\middle|\|X\| \geq \|u\|\right]$ is the expectation of a symmetric distribution over an annulus centered at the origin. Therefore, one has that $\mathbb{E}[X|X \geq_{\mathcal{R}} Q(u)] = 0$ and it cannot be thought of as a typical extreme observation.*

This is caused by a lack of information in $\mathbb{E}[X|X \geq_{\mathcal{R}} Q(u)]$. In fact, the rank function neglects the directional information of X, which lies in the sign function. In order to overcome this issue, we have to introduce a few notation. For any $u \in \mathbb{B}(0,1)\backslash\{0\}$, let

$$L_u = \left\{ t\frac{u}{\|u\|} : t \in [0,1] \right\} \tag{4.6}$$

be a parametrization of the radius of $\mathbb{B}(0,1)\backslash\{0\}$ whose direction is $u/\|u\|$. The *sign curve* C_u associated to u is the image of L_u by the center-outward quantile function, namely

$$C_u = Q(L_u).$$

Several sign curves are represented in red at the right-hand side of Figure 1.3. Averaging observations less deep that $Q(u)$ along the sign curve C_u induces a "typical" value "beyond" $Q(u)$ in a meaningful multivariate way.

Definition 4.5. *Let $\nu \in \mathcal{P}_1(\mathbb{R}^d)$ with center-outward quantile function Q. The center-outward superquantile function of ν is the function S defined, for any $u \in \mathbb{B}(0,1)\backslash\{0\}$, by*

$$S(u) = \frac{1}{1-\|u\|}\int_{\|u\|}^1 Q\Big(t\frac{u}{\|u\|}\Big)dt,$$

where the above integral is to be understood component-wise.

Remark 4.6 (Consistency with the univariate case)**.** *On the real line, there is only one sign curve $C_1 = \{Q(t); t \in [0,1]\}$. Thus, definition (4.3) can be seen as averaging observations less deep than $Q(\alpha)$, w.r.t. ν, along the sign curve C_1. Nonetheless, note that center-outward quantiles slightly differ from classical quantiles in dimension $d = 1$, because $U([-1,1]) \neq U([0,1])$, even if they carry the same information, [87][Appendix B].*

By the same token, we can define the center-outward expected shortfall function.

Definition 4.7. *Let $\nu \in \mathcal{P}_1(\mathbb{R}^d)$ with center-outward quantile function Q. The center-outward expected shortfall function of ν is the function E defined, for any*

$u \in \mathbb{B}(0,1)\backslash\{0\}$, *by*

$$E(u) = \frac{1}{\|u\|}\int_0^{\|u\|} Q\Big(t\frac{u}{\|u\|}\Big)dt,$$

where the above integral is to be understood component-wise.

Remark 4.8. *For $\nu \in \mathcal{P}_*(\mathbb{R}^d)$, Q is neither defined at the origin nor at the boundary of the unit ball, unless the support of ν is compact. Hence the integrals in $S(u)$ and $E(u)$ are improper and they shall be understood respectively as*

$$\lim_{r\to 1^-}\int_{\|u\|}^{r} Q\Big(t\frac{u}{\|u\|}\Big)dt \qquad \text{and} \qquad \lim_{r\to 0^+}\int_{r}^{\|u\|} Q\Big(t\frac{u}{\|u\|}\Big)dt.$$

However, note that they are convergent as soon as $\nu \in \mathcal{P}_1(\mathbb{R}^d)$. In fact, the necessary assumption is rather the following integrability along sign curves,

$$\int_0^1 \left\| Q(t\frac{u}{\|u\|}) \right\| dt < \infty. \tag{4.7}$$

By the change of variables formula for push-forwards maps and by definition of U_d, $\mathbb{E}_\nu[\|X\|] = \mathbb{E}_{U_d}[\|Q(U)\|] = \mathbb{E}_{(R,\Phi)}[\|Q(R\Phi)\|]$. In other words, denoting by $\mathbb{P}_\mathbb{S}$ the uniform probability measure on the sphere $\mathbb{S}^{d-1} = \{\varphi \in \mathbb{R}^d : \|\varphi\| = 1\}$,

$$\mathbb{E}_\nu[\|X\|] = \int_{\mathbb{S}^{d-1}}\int_0^1 \|Q(t\varphi)\| dt\, d\mathbb{P}_\mathbb{S}(\varphi).$$

Thus, one can see by contradiction that, as soon as $\mathbb{E}_\nu[\|X\|] < \infty$, (4.7) holds for almost all $\varphi \in \mathbb{S}^{d-1}$, thus S and E must be finite almost everywhere.

The following naturally extends Definition 4.2.

Definition 4.9 (Superquantile and expected shortfall regions and contours)**.** *Let $\nu \in \mathcal{P}_1(\mathbb{R}^d)$ with center-outward superquantile function S and expected shortfall E. Then,*
(i) the superquantile (resp. expected shortfall) region $\mathbb{C}^s_\alpha$ (resp. $\mathbb{C}^e_\alpha$) of order $\alpha \in [0,1]$ is the image by S (resp. E) of the ball $\mathbb{B}(0,\alpha)$.
(ii) the superquantile (resp. expected shortfall) contour $\mathcal{C}^s_\alpha$ (resp. $\mathcal{C}^e_\alpha$) of order $\alpha \in [0,1]$ is the boundary of $\mathbb{C}^s_\alpha$ (resp. $\mathbb{C}^e_\alpha$).
(iii) averaged sign curves by E or S are respectively defined by $E(L_u)$ and $S(L_u)$ for any $u \in \mathbb{B}(0,1)\backslash\{0\}$.

The distribution of a point cloud can be split into a central area and a peripheral one, that can be described through the superquantile and expected shortfall functions, as illustrated in the numerical experiments carried out in Section 4.5.

4.2.2 Invariance properties

The two following lemmas are immediate consequences of [74, Lemmas A.7, A.8].

Lemma 4.10. *Assume that $X \in \mathbb{R}^d$ is an integrable random vector. Suppose that $a > 0$, $b \in \mathbb{R}^d$ and $Y = aX + b$. Denote by S_X, S_Y and E_X, E_Y their center-outward superquantile and expected shortfall functions. Then, for $u \in \mathbb{B}(0,1)\backslash\{0\}$,*

$$S_Y(u) = aS_X(u) + b \quad \text{and} \quad E_Y(u) = aE_X(u) + b.$$

Proof. We only detail S_Y, as one can deduce E_Y identically. From [74][Lemma A.7],

$$\begin{aligned} S_Y(u) &= \frac{1}{1-\|u\|}\int_{\|u\|}^1 Q_Y(t\frac{u}{\|u\|})dt, \\ &= \frac{1}{1-\|u\|}\int_{\|u\|}^1 \Big(aQ_X(t\frac{u}{\|u\|}) + b\Big)dt = aS_X(u) + b. \end{aligned}$$ □

Lemma 4.11. *Assume that $X \in \mathbb{R}^d$ is an integrable random vector. Let S_X, S_Y and E_X, E_Y be the center-outward superquantile and expected shortfall functions associated respectively with X and $Y = AX$, for A an orthonormal matrix. Then, for $u \in \mathbb{B}(0,1)\backslash\{0\}$,*

$$S_Y(u) = AS_X(A^T u) \quad \text{and} \quad E_Y(u) = AE_X(A^T u).$$

Proof. Again, we only detail S_Y, as the proof for E_Y is identical. Combining [74][Lemma A.8] with the fact that $\|A^T u\| = \|u\|$ for an orthogonal matrix,

$$\begin{aligned} S_Y(u) &= \frac{1}{1-\|u\|}\int_{\|u\|}^1 Q_Y(t\frac{u}{\|u\|})dt, \\ &= \frac{1}{1-\|u\|}\int_{\|u\|}^1 AQ_X(tA^T\frac{u}{\|u\|})dt = AS_X(A^T u). \end{aligned}$$ □

4.2.3 Main results

Quoting Rockafellar & Royset in [145],

> "the superquantile function [...] is as fundamental to a random variable as the distribution and quantile functions".

This assertion is partially motivated by the fact that the distribution, quantile and superquantile functions are uniquely determined one to another, and by the fact that pointwise convergence of these functions metrizes convergence in distribution. Such properties hold for our integrated concepts and are stated hereafter. First of all, we shall make repeated use of the fact that the center-outward superquantile and expected shortfall functions are two sides of the same coin, that is, for $u \in \mathbb{B}(0,1)\backslash\{0\}$,

$$\int_0^1 Q\Big(t\frac{u}{\|u\|}\Big)dt = \|u\|E(u) + (1-\|u\|)S(u). \tag{4.8}$$

This is a generalization of the immediate property that, for the univariate setting given in the introduction with (4.1) and (4.2), we have for all $\alpha \in]0,1[$,

$$\mathbb{E}[X] = \alpha E(\alpha) + (1-\alpha)S(\alpha).$$

A first main result reads as follows.

Theorem 4.12. *Let $\nu_1, \nu_2 \in \mathcal{P}_1(\mathbb{R}^d)$ with respective center-outward quantile, superquantile and expected shortfall functions denoted by Q_1, S_1, E_1 and Q_2, S_2, E_2. Then, the following are equivalent.*

(i) $\nu_1 = \nu_2$

(ii) $Q_1 = Q_2$ U_d*-a.e.*

(iii) $S_1 = S_2$ U_d*-a.e.*

(iv) $E_1 = E_2$ U_d*-a.e.*

Proof. First of all, it is already known that $(i) \Leftrightarrow (ii)$. Indeed, with the choice of U_d as the reference distribution, McCann's theorem [121] ensures that $(i) \Rightarrow (ii)$. Obviously, $(ii) \Rightarrow (i)$, from the very definition of push-forward measures that is $\nu_1(A) = U_d(Q_1^{-1}(A))$ and $\nu_2(A) = U_d(Q_2^{-1}(A))$. Hereafter, we proceed through the implication loop $(ii) \Rightarrow (iii) \Rightarrow (iv) \Rightarrow (ii)$. Let $\mathbb{P}_{\mathbb{S}}$ be the uniform probability measure on $\mathbb{S}^{d-1} = \{\varphi \in \mathbb{R}^d : \|\varphi\|_2 = 1\}$. Let R and φ be drawn uniformly on $[0,1]$ and $\mathbb{S}^{d-1}$, respectively. Suppose that they are independent, so that their joint distribution is given by $d\mathbb{P}_{(R,\varphi)}(r,\varphi) = d\mathbb{P}_R(r)d\mathbb{P}_{\mathbb{S}}(\varphi)$. There, by definition, if $f : (r,\varphi) \mapsto r\varphi$ on $[0,1] \times \mathbb{S}^{d-1}$, then $U_d = f_{\#}\mathbb{P}_{(R,\varphi)}$. Hence, from the change of variables formula for push-forward measures, for any Borel set $A = f(B \times C)$, for $B \subset [0,1]$, $C \subset \mathbb{S}^{d-1}$,

$$\begin{aligned}\int_A (S_1 - S_2)(u)dU_d(u) &= \int_B \int_C (S_1 - S_2)(r\varphi)d\mathbb{P}_{\mathbb{S}}(\varphi)dr, \\ &= \int_B \int_C \Big(\frac{1}{1-r}\int_r^1 (Q_1 - Q_2)(t\varphi)dt\Big)d\mathbb{P}_{\mathbb{S}}(\varphi)dr. \qquad (4.9)\end{aligned}$$

Nonetheless, for any $r \in B$, by the same change of variables,

$$\int_C \int_r^1 (Q_1 - Q_2)(t\varphi)dtd\mathbb{P}_{\mathbb{S}}(\varphi) = \int_{f(C\times[r,1])} (Q_1 - Q_2)(u)dU_d(u). \qquad (4.10)$$

Note that $U_d(f(C\times[r,1])) = (1-r)\mathbb{P}_{\mathbb{S}}(C)$, which is positive since $r < 1$. Obviously, (4.10) vanishes as soon as $Q_1 = Q_2$ U_d*-a.e.* It implies that (4.9) also vanishes, which justifies that $(ii) \Rightarrow (iii)$. Furthermore, we claim that $(iii) \Rightarrow (iv)$. Indeed,

$$\int_0^1 Q_1(t\frac{u}{\|u\|})dt = \lim_{r\to 0^+} S_1(r\frac{u}{\|u\|}) \quad \text{and} \quad \int_0^1 Q_2(t\frac{u}{\|u\|})dt = \lim_{r\to 0^+} S_2(r\frac{u}{\|u\|}).$$

Consequently, if we assume that $S_1 = S_2$ U_d-*a.e.*, we obtain that

$$\int_0^1 (Q_1 - Q_2)(t\frac{u}{\|u\|})dt = 0. \tag{4.11}$$

Using (4.8), the desired result follows, $(iii) \Rightarrow (iv)$. Finally, assume that $E_1 = E_2$ U_d-a.e. Consequently, for all $r \in]0,1[$ and for all $\varphi \in \mathbb{S}^{d-1}$,

$$\int_0^r Q_1(t\varphi)dt = \int_0^r Q_2(t\varphi)dt.$$

Using that $\int_a^b = \int_0^b - \int_0^a$, for any $0 \leq a \leq b \leq 1$,

$$\int_a^b Q_1(t\varphi)dt = \int_a^b Q_2(t\varphi)dt. \tag{4.12}$$

For any measurable $B \subset \mathbb{S}^{d-1}$, by integrating (4.12) w.r.t. $\mathbb{P}_{\mathbb{S}}$,

$$\int_B \int_a^b Q_1(t\varphi)d\mathbb{P}_R(t)d\mathbb{P}_{\mathbb{S}}(\varphi) = \int_B \int_a^b Q_2(t\varphi)d\mathbb{P}_R(t)d\mathbb{P}_{\mathbb{S}}(\varphi).$$

By use of $U_d = f_{\#}\mathbb{P}_{(R,\varphi)}$, a change of variables above yields, for any $A \subset \mathbb{B}(0,1)\backslash\{0\}$,

$$\int_A Q_1(u)dU_d(u) = \int_A Q_2(u)dU_d(u),$$

and the result follows.

□

Interestingly enough, our proposed integrated quantile functions are both simply related to the Kantorovich potential. Somehow, this development generalizes the work of [148] where the distribution function is related to the univariate superquantile by the way of the surexpectation function, which is nothing more than a particular primitive of the distribution function.

Proposition 4.13. *The center-outward expected shortfall function of $\nu \in \mathcal{P}_1(\mathbb{R}^d)$ satisfies, for any $u \in \mathbb{B}(0,1)\backslash\{0\}$,*

$$\langle E(u), u\rangle = \psi(u). \tag{4.13}$$

Moreover, the center-outward superquantile function of a compactly supported probability measure $\nu \in \mathcal{P}_*(\mathbb{R}^d)$ verifies, for every $u \in \mathbb{B}(0,1)\backslash\{0\}$,

$$\langle S(u), u\rangle = \frac{\|u\|}{1-\|u\|}\Big(\psi(\frac{u}{\|u\|}) - \psi(u)\Big). \tag{4.14}$$

Proof. Fix $u \in \mathbb{B}(0,1)\backslash\{0\}$ and let $f(t) = \psi(tu)$. Then f is a finite and convex function on $]-a, a[$ for some $a > 1$, so that one can apply [146][Corollary 24.2.1].

Combined with *a.e.* differentiability, [146][Theorem 25.3], this yields

$$\psi(u) = \int_0^1 \langle \nabla\psi(tu), u\rangle dt, \tag{4.15}$$

which can be rewritten as

$$\psi(u) = \langle \int_0^1 Q(tu)dt, u\rangle = \langle E(u), u\rangle,$$

from a simple change of variables, leading to our first point (4.13). Moreover, denote for any $u \in \mathbb{B}(0,1)\backslash\{0\}$,

$$f(t) = \psi(t\frac{u}{\|u\|}).$$

This function is finite on $[0,1]$ but not on a larger interval anymore. But, because $\nu \in \mathcal{P}_*(\mathbb{R}^d)$, f is differentiable everywhere on $]0,1[$, and we have by the chain rule formula,

$$f'(t) = \langle Q(t\frac{u}{\|u\|}), \frac{u}{\|u\|}\rangle.$$

Since $f'(t)$ is a non-decreasing function from $\mathbb{R}$ to $[-\infty, +\infty]$ finite at $t = \|u\|$, [146][Theorem 24.2] ensures that the function F, defined for all $x \in \mathbb{R}$, by

$$F(x) = \int_{\|u\|}^x f'(t)dt,$$

is a well-defined closed proper convex function on $\mathbb{R}$. We emphasize that $F(x)$ takes infinite values as soon as $x > 1$, while, for $x = 1$, F is well-defined as a Lebesgue integral, or as a limit of Riemann integrals, as explained in the proof of [146][Theorem 24.2], and it may take finite or infinite values. In addition, the latter theorem tells us that $F(x) = f(x) + \alpha$ for some $\alpha \in \mathbb{R}$ everywhere. But $F(\|u\|) = 0$ and $f(\|u\|) = \psi(u)$, so $F(x) = f(x) - \psi(u)$. Assuming that the support $\mathcal{X}$ of ν is compact, ψ must be lipschitz continuous, [122][Lemma 2], a fortiori bounded and the subdifferential $\partial\psi(K)$ must also be bounded, see *e.g.* [147]. As a byproduct, $\psi(u/\|u\|)$ and $Q(u/\|u\|)$ are finite. In this case, $F(1) = f(1) - \psi(u)$ is finite and can be rewritten as

$$F(1) = \int_{\|u\|}^1 \langle Q(t\frac{u}{\|u\|}), \frac{u}{\|u\|}\rangle dt = \psi(\frac{u}{\|u\|}) - \psi(u),$$

which implies (4.14), completing the proof of Proposition 4.13.

□

Thanks to this relation between the potential ψ whose gradient gives Q and the center-outward expected shortfall function, we are now able to characterize the convergence in distribution for a sequence of random vectors through superquantiles and expected shortfalls. For that purpose, we rely on existing results, [74, 87], on the

relation between convergence in distribution and center-outward quantile functions.

Lemma 4.14. *Let (X, X_n) be a sequence of random vectors with distributions ν and ν_n respectively, in $\mathcal{P}_*(\mathbb{R}^d)$. Then,*

$$X_n \xrightarrow{\mathcal{L}} X \quad \Leftrightarrow \quad \forall u \in \mathbb{B}(0,1),\ \lim_{n\to\infty} \psi_n(u) = \psi(u), \tag{4.16}$$

$$\Leftrightarrow \quad \lim_{n\to\infty} \sup_{u\in K} |\psi_n(u) - \psi(u)| = 0, \tag{4.17}$$

$$\Leftrightarrow \quad \lim_{n\to\infty} \sup_{u\in K} \|Q_n(u) - Q(u)\| = 0, \tag{4.18}$$

for every compact $K \subset \mathbb{B}(0,1)\backslash\{0\}$. In fact, uniform convergence of ψ_n towards ψ even holds on every compact $K \subset \mathbb{B}(0,1)$.

Proof. On the one hand, assume that (X_n) converges in distribution to X. Then, the right-hand side of (4.18) is the main result of [74][Theorem 4.1] when Q_n and Q are homeomorphisms between convex sets, with uniform convergence on any compact subset. But, with reference distribution U_d, center-outward quantile maps are not, because of the discontinuity at the origin, see [87][Remark 3.4]. Nonetheless, as highlighted in the proof of [87][Proposition 3.3], the assumption that $X_n \xrightarrow{\mathcal{L}} X$ for $\nu_n, \nu \in \mathcal{P}_*(\mathbb{R}^d)$ is sufficient to apply Theorem 2.8 in [12]. As a consequence, $\psi_n(u)$ converges towards $\psi(u)$ for every $u \in \mathbb{B}(0,1)$, that is the right-hand side of (4.16). Being finite and convex functions, [146][Theorem 10.8] ensures that such pointwise convergence implies the right-hand side of (4.17). Since ψ_n and ψ are differentiable on $\mathbb{B}(0,1)\backslash\{0\}$ as soon as $\nu_n, \nu \in \mathcal{P}_*(\mathbb{R}^d)$, one can apply [146][Theorem 25.7] on any open and convex set $K' \subset \mathbb{B}(0,1)\backslash\{0\}$. This gives us the uniform convergence of Q_n towards Q for any compact K included in an open and convex set $K' \subset \mathbb{B}(0,1)\backslash\{0\}$. To extend this to any compact K, we use that, by compacity, one can extract from the open cover $\{\mathbb{B}(x,\delta); x \in K\}$ a finite cover. Consequently, it exists $N \in \mathbb{N}$ and $x_1, \cdots, x_N \in K$ such that

$$\sup_{u\in K} \|Q_n(u) - Q(u)\| \leq \sum_{k=1}^{N} \sup_{u\in\mathbb{B}(x_k,\delta)} \|Q_n(u) - Q(u)\|. \tag{4.19}$$

Here, the closure of each ball $\mathbb{B}(x_k,\delta)$ is a compact set, which, because of the choice of δ, is a subset of some open and convex set $K' \subset \mathbb{B}(0,1)\backslash\{0\}$. But we already know that Q_n uniformly converges towards Q on such a set, thus the right-hand side of (4.19) vanishes when $n \to 0$, which yields the right-hand side of (4.18). It only remains to prove that $X_n \xrightarrow{\mathcal{L}} X$ as a direct consequence. This last claim relies on the Portmanteau theorem which says that (X_n) converges in distribution to X iff for any bounded and continuous function f,

$$\lim_{n\to+\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]. \tag{4.20}$$

However, we clearly have for any bounded and continuous function f,

$$\mathbb{E}[f(X_n)] = \int_{\mathbb{R}^d} f(x)d\nu_n(x) = \int_{\mathbb{B}(0,1)} f(Q_n(u))dU_d(u).$$

using the change of variables $\nu_n = Q_{n\#}U_d$. Hence, as $f \circ Q_n$ is uniformly bounded, the dominated convergence theorem leads to (4.20). □

Theorem 4.15. *Let (X, X_n) be a sequence of random vectors with distributions ν and ν_n respectively, in $\mathcal{P}_*(\mathbb{R}^d)$. Then,*

$$X_n \xrightarrow{\mathcal{L}} X \quad \Leftrightarrow \quad \forall u \in \mathbb{B}(0,1)\backslash\{0\}, \lim_{n\to+\infty} E_n(u) = E(u), \tag{4.21}$$

$$\Leftrightarrow \quad \lim_{n\to\infty} \sup_{u\in K} \|E_n(u) - E(u)\| = 0, \tag{4.22}$$

for every compact $K \subset \mathbb{B}(0,1)\backslash\{0\}$.

Proof. On the one hand, assume that (X_n) converges in distribution to X. Then, it follows from (4.18) that (Q_n) converges uniformly to Q on any compact $K \subset \mathbb{B}(0,1)\backslash\{0\}$. For any $u \in \mathbb{B}(0,1)\backslash\{0\}$, we have from Definition 4.7 that

$$\|E_n(u) - E(u)\| \leq \frac{1}{\|u\|} \int_0^{\|u\|} R_n(t\frac{u}{\|u\|})dt,$$

where $R_n(v) = \|Q_n(v) - Q(v)\|$. Fix a compact $K \subset \mathbb{B}(0,1)\backslash\{0\}$ such that for every $u \in K$, $C < \|u\| < D$ for some positive constants C, D. Then, we clearly have

$$\sup_{u\in K} \|E_n(u) - E(u)\| \leq \frac{1}{C} \int_0^D \sup_{u\in K} R_n(t\frac{u}{\|u\|})dt.$$

Consequently, for any $\xi \in \mathbb{R}$ such that $0 \leq \xi < D$,

$$\sup_{u\in K} \|E_n(u) - E(u)\| \leq \frac{1}{C} \left(\int_0^\xi \sup_{u\in K} R_n(t\frac{u}{\|u\|})dt + \int_\xi^D \sup_{u\in K} R_n(t\frac{u}{\|u\|})dt \right). \tag{4.23}$$

On the one hand, for all $t \in [\xi, D]$ and $u \in K$, $tu/\|u\|$ lies inside a compact $K' \subset \mathbb{B}(0,1)\backslash\{0\}$, so that the second term in (4.23) satisfies

$$\int_\xi^D \sup_{u\in K} R_n(t\frac{u}{\|u\|})dt \leq (D-\xi) \sup_{v\in K'} R_n(v) \leq \sup_{v\in K'} R_n(v), \tag{4.24}$$

which vanishes when $n \to +\infty$. On the other hand,

$$\int_0^\xi \sup_{u\in K} R_n(t\frac{u}{\|u\|})dt \leq \xi \sup_{v\in\overline{\mathbb{B}}(0,D)} R_n(v). \tag{4.25}$$

We now claim that $\sup_{v\in\overline{\mathbb{B}}(0,D)} R_n(v)$ is bounded by a constant. Recall that, for

any $v \in \overline{\mathbb{B}}(0,D)$, $Q_n(v)$ and $Q(v)$ belong to the subdifferentials $\partial\psi_n(v)$ and $\partial\psi(v)$, by definition. Because $S = \overline{\mathbb{B}}(0,D)$ is a compact subset of the open unit ball, and because, for any $u \in S$, $\psi_n(u)$ is convergent and a fortiori bounded, see Lemma 4.14, we are in position to apply [146][Theorem 10.6]. It directly implies that the sequence (ψ_n) is uniformly bounded and equi-Lipschitz on S. However, it is well-known that any L-lipschitz convex function ϕ on S must have a bounded subdifferential $\partial\phi(S)$, see *e.g.* [147]. Indeed, for any $u \in S$ and any $\xi \in \partial\phi(S)$, consider $\delta > 0$ such that $y = u - \delta\xi$ belongs to S. By definition of $\partial\phi(S)$ and the lipschitz constant L,

$$L\delta\|\xi\| \geq \phi(y) - \phi(u) \geq \langle \xi, y - u\rangle = \delta\|\xi\|^2.$$

This immediately gives us the existence of a uniform bound on the family of subdifferentials $(\partial\psi_n(S))_n$. Because $\partial\psi(S)$ is also bounded, by lipschitz regularity of ψ on S,

$$\sup_{v\in\overline{\mathbb{B}}(0,D)} R_n(v) \leq \sup_{v\in\overline{\mathbb{B}}(0,D)} \|Q_n(v)\| + \|Q(v)\| \leq M < +\infty. \tag{4.26}$$

Therefore the right-hand side of (4.25) can be bounded by ξM for M given in (4.26). Then it follows from (4.23) and (4.24) that, for any $\xi \in [0, D[$, it exists $n_1 \in \mathbb{N}$ such that, for all $n \geq n_1$,

$$\sup_{u\in K} \|E_n(u) - E(u)\| \leq \frac{1}{C}\Big(\xi M + \sup_{v\in K'} R_n(v)\Big).$$

Thus, for any $\xi \in [0, D[$,

$$\lim_{n\to+\infty} \sup_{u\in K} \|E_n(u) - E(u)\| \leq \frac{\xi M}{C}.$$

which leads to the right-hand side of (4.22) as ξ goes to zero. In addition, the right-hand side of (4.22) immediately implies the right-hand side of (4.21). On the other hand, assume that the right-hand side of (4.21) holds. Then, we obtain from (4.13) that $\forall u \in \mathbb{B}(0,1)$,

$$\lim_{n\to+\infty} \psi_n(u) = \psi(u).$$

Thus, the desired result follows from Lemma 4.14, which completes the proof of Theorem 4.15. □

Our last result requires some uniform integrability assumption on the sequence (X_n).

Theorem 4.16. *Let (X, X_n) be a sequence of random vectors with distributions ν and ν_n respectively, in $\mathcal{P}_*(\mathbb{R}^d)$. In addition, suppose that there exists a random variable Z greater than 1 such that $\mathbb{E}[Z\ln(Z)] < +\infty$ and for all $n \in \mathbb{N}$,*

$$\|X_n\| \leq Z \qquad a.s. \tag{4.27}$$

Then,

$$X_n \xrightarrow{\mathcal{L}} X \Leftrightarrow \lim_{n\to\infty} S_n(u) = S(u), \tag{4.28}$$

for almost all $u \in \mathbb{B}(0,1)$.

Remark 4.17. *One can observe that if the support of every* ν_n *is bounded by some constant* M, (4.27) *is no longer needed. Under this restrictive assumption, there is no mass going out to infinity in any direction and one can easily check that*

$$X_n \xrightarrow{\mathcal{L}} X \Leftrightarrow \lim_{n\to\infty} \sup_{u\in K} \|S_n(u) - S(u)\| = 0,$$

for every compact $K \subset \mathbb{B}(0,1)\backslash\{0\}$.

Proof. On the one hand, assume that (X_n) converges in distribution to X. We clearly have from (4.27) that, if $\Phi(x) = x\ln(x)$, then

$$\mathbb{E}[\sup_{n\in\mathbb{N}} \Phi(X_n)] \leq \mathbb{E}[Z\ln(Z)] < +\infty.$$

Hence, using the same change of variables as in the proof of Theorem 4.12, we obtain that

$$\mathbb{E}[\sup_{n\in\mathbb{N}} \Phi(X_n)] = \int_{\mathbb{S}^{d-1}} \int_0^1 \sup_{n\in\mathbb{N}} \|Q_n(t\varphi)\| \ln(\|Q_n(t\varphi)\|) dt d\mathbb{P}_{\mathbb{S}}(\varphi) < +\infty, \tag{4.29}$$

where we recall that $\mathbb{S}^{d-1} = \{\varphi \in \mathbb{R}^d : \|\varphi\|_2 = 1\}$. Consequently, we deduce from (4.29) that for almost all $\varphi \in \mathbb{S}^{d-1}$,

$$\sup_{n\in\mathbb{N}} \int_0^1 \|Q_n(t\varphi)\| \ln(\|Q_n(t\varphi)\|) dt \leq \int_0^1 \sup_{n\in\mathbb{N}} \|Q_n(t\varphi)\| \ln(\|Q_n(t\varphi)\|) dt < +\infty. \tag{4.30}$$

It clearly leads to the uniform integrability of (Q_n) along sign curves, see [26][Theorem 4.5.9]. However, we already saw from Lemma 4.14 that the convergence in distribution of (X_n) to X implies the pointwise convergence of Q_n to Q on $\mathbb{B}(0,1)\backslash\{0\}$. Therefore, it follows from the Lebesgue-Vitali theorem [26][Theorem 4.5.4] that for almost all $\varphi \in \mathbb{S}^{d-1}$,

$$\lim_{n\to+\infty} \int_0^1 Q_n(t\varphi) dt = \int_0^1 Q(t\varphi) dt,$$

which means that for almost all $u \in \mathbb{B}(0,1)\backslash\{0\}$,

$$\lim_{n\to+\infty} \int_0^1 Q_n\Big(t\frac{u}{\|u\|}\Big) dt = \int_0^1 Q\Big(t\frac{u}{\|u\|}\Big) dt. \tag{4.31}$$

Hereafter, we already saw from (4.8) that

$$S_n(u)-S(u)=\frac{1}{1-\|u\|}\left(\int_0^1 (Q_n-Q)\Big(t\frac{u}{\|u\|}\Big)dt-\|u\|\Big(E_n(u)-E(u)\Big)\right). \quad (4.32)$$

Finally, the right-hand side of (4.28) follows from (4.21) together with (4.31) and (4.32). On the other hand, assume that the right-hand side of (4.28) holds. We have for almost all $u\in\mathbb{B}(0,1)\backslash\{0\}$,

$$\lim_{n\to+\infty}\int_{\|u\|}^1 Q_n\Big(t\frac{u}{\|u\|}\Big)dt=\int_{\|u\|}^1 Q\Big(t\frac{u}{\|u\|}\Big)dt. \quad (4.33)$$

Hence, we obtain from (4.33) that for almost every $r\in]0,1[$ and $\varphi\in\mathbb{S}^{d-1}$,

$$\lim_{n\to+\infty}\int_r^1 Q_n(t\varphi)dt=\int_r^1 Q(t\varphi)dt. \quad (4.34)$$

Obviously, for almost all $r\in]0,1[$ and almost all $\varphi\in\mathbb{S}^{d-1}$,

$$\begin{aligned}\lim_{n\to+\infty}\int_0^1 (Q_n-Q)(t\varphi)dt&=\lim_{n\to+\infty}\left(\int_0^r (Q_n-Q)(t\varphi)dt+\int_r^1 (Q_n-Q)(t\varphi)dt\right),\\&=\lim_{n\to+\infty}\int_0^r (Q_n-Q)(t\varphi)dt.\end{aligned}$$

In addition, this integral is always finite since (4.7) holds for integrable probability measures. As the left-hand side does not depend on r,

$$\lim_{n\to+\infty}\int_0^1 (Q_n-Q)(t\varphi)dt=\lim_{r\to 0}\lim_{n\to+\infty}\int_0^r (Q_n-Q)(t\varphi)dt. \quad (4.35)$$

From the uniform integrability of (Q_n) along sign curves, we obtain that

$$\lim_{r\to 0}\lim_{n\to+\infty}\int_0^r \|Q_n(t\varphi)-Q(t\varphi)\|dt=0.$$

Hence, by use of (4.35), we find that

$$\lim_{n\to+\infty}\int_0^1 Q_n(t\varphi)dt=\int_0^1 Q(t\varphi)dt.$$

It ensures that for almost all $u\in\mathbb{B}(0,1)\backslash\{0\}$,

$$\lim_{n\to+\infty}\int_0^1 Q_n\Big(t\frac{u}{\|u\|}\Big)dt=\int_0^1 Q\Big(t\frac{u}{\|u\|}\Big)dt. \quad (4.36)$$

Consequently, we deduce from (4.32) and (4.36) that for a.e. $u\in\mathbb{B}(0,1)\backslash\{0\}$,

$$\lim_{n\to+\infty}E_n(u)=E(u) \quad (4.37)$$

Finally, it follows from (4.37) together with (4.21) that (X_n) converges in distribution to X, which completes the proof of Theorem 4.16. □

4.3 A class of reference measures

Until now, the concepts of MK superquantiles and expected shortfalls strongly rely on the choice of the reference distribution, and one may question the choice of U_d. This issue is discussed hereafter, with explicit MK quantile functions for generalized gamma models, by calibrating the reference distribution. Ultimately, one obtains nested regions indexed by their probability content. The choice of the reference distribution μ is a major tool to adapt MK quantiles, ranks and signs to any task encountered. Notably, the uniform distribution over the unit hypercube has advantages concerning marginal independence, [50, 74], whereas orthogonal invariance only holds with a spherical reference distribution, [74, 85]. U_d plays a specific role for the interpretation of quantile regions indexed by a probability content level. This section deals with an alternative class of reference measures defined by α-level sets on the p-unit ball, that preserve the interpretability of U_d. We also consider the restriction of U_d to $\mathbb{R}^d_+$, with a *left-to-right* ordering that might be preferred in the subsequent risk applications to the *center-outward* one. Denote by $\mathbb{S}^{d,p}$ the unit sphere $\mathbb{S}^{d,p} = \{\varphi \in \mathbb{R}^d : \|\varphi\|_p = 1\}$ with

$$\|\varphi\|_p^p = \sum_{k=1}^{d} |\varphi_k|^p,$$

and $\mathbb{S}^{d,p}_+ = \mathbb{S}^{d,p} \cap \mathbb{R}^d_+$. Moreover, let q be the Hölder conjugate of p, given by

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Definition 4.18. *The q-spherical conjugate distribution $U_{d,q}$ is defined as the product $R\Phi$ between two independent random variables R and Φ where*

- *R has uniform distribution on $[0,1]$,*
- *the distribution of $\Phi \in \mathbb{S}^{d,q}$ is given by $\Phi = \Psi^{\otimes(p-1)}$ for Ψ uniformly drawn on $\mathbb{S}^{d,p}$, and $\otimes$ the component-wise exponent.*

The restriction of $U_{d,q}$ to the convex cone $\mathbb{R}^d_+$ is denoted by $U^+_{d,q}$.

With reference measure $U_{d,q}$, the MK quantile function Q_q can always be defined almost everywhere from a convex potential ψ by Definition 1.2, for which we can always assume that $\psi(0) = 0$. At any point u of non differentiability of ψ, one can still define $Q_q(u)$ as the average of the subdifferential $\partial\psi(u)$, so that Q_q is defined everywhere on the q-unit ball. The interpretability of the quantile concepts remains the same among this class, as the $U_{d,q}$-probability of the q-ball of radius $\alpha \in [0,1]$

is α,

$$\mathbb{P}(\|R\Phi\|_q \leq \alpha) = \mathbb{P}(R \leq \alpha) = \alpha.$$

Crucially for us, our definitions and properties of superquantiles and expected shortfalls naturally adapt. Definition 4.5 as well as Definition 4.7 extend, for all $u \in \mathbb{B}(0,1)$, to

$$S_q(u) = \frac{1}{1-\|u\|_q} \int_{\|u\|_q}^1 Q_q(t\frac{u}{\|u\|_q})dt$$

and

$$E_q(u) = \frac{1}{\|u\|_q} \int_0^{\|u\|_q} Q_q(t\frac{u}{\|u\|_q})dt.$$

We now develop on how to sample uniformly on p-spheres $\mathbb{S}^{d,p}$. Let Γ stand for the Euler Gamma function and denote by $\mathcal{L}_p$ the probability distribution with density function

$$f_p(x) = \frac{p}{\Gamma(p^{-1})} \exp(-x^p) I_{\mathbb{R}_+}(x).$$

A random variable X on $\mathbb{R}_+$ is drawn from $\mathcal{L}_p$ if and only if X^p follows a Gamma distribution with shape and scale parameters $1/p$ and 1. The following lemma indicates how to sample uniformly on p-spheres when $p > 0$.

Lemma 4.19 ([14, 152])**.** *For any real $p > 0$, the components of a random vector $X \in \mathbb{R}_+^d$ are independent with distribution $\mathcal{L}_p$ if and only if the random variables $\|X\|_p$ and $\|X\|_p^{-1}X$ are independent where*

- *$\|X\|_p^{-1}X$ is uniformly distributed on $\mathbb{S}_+^{d,p}$,*
- *$\|X\|_p^p$ has Gamma distribution with shape and scale parameters d/p and 1.*

For $p = 2$, $\mathcal{L}_p$ corresponds to $|Z|$ where Z is drawn from a $\mathcal{N}(0,1)$ distribution. The MK distribution function with reference U_d (or $U_{d,2}$) has been obtained in [33][Section 2.4]. Here, we use similar arguments to find gradient-of-convex maps for our alternative class of reference measures. We emphasize that such maps are invariant to shifts, see [74][Lemma A.7], whereas invariance to rotations requires a spherically symmetric reference measure, [74][Lemma A.8].

Proposition 4.20. *Suppose that the components of a random vector $X \in \mathbb{R}_+^d$ are independent with $\mathcal{L}_p$ distribution for $p > 1$, and let $q = p/(p-1)$. Then, the MK distribution function with respect to $U_{d,q}^+$ has the explicit formulation*

$$F_q(x) = (\|x\|_p^{-1}x)^{\otimes(p-1)} G(\|x\|_p), \tag{4.38}$$

where $\otimes$ stands for the component-wise exponent and G is the univariate distribution function of $\|X\|_p$. Its inverse, the MK quantile function, is given, for all $u \in \mathbb{R}_+^d$ such that $\|u\|_q \leq 1$, by

$$Q_q(u) = (\|u\|_q^{-1}u)^{\otimes(q-1)} G^{-1}(\|u\|_q). \tag{4.39}$$

The MK superquantile and expected shortfall functions are respectively given, for all $u \in \mathbb{R}^d_+$ *such that* $\|u\|_q \leq 1$*, by*

$$S_q(u) = (\|u\|_q^{-1}u)^{\otimes(q-1)}\overline{S}(\|u\|_q), \tag{4.40}$$

$$E_q(u) = (\|u\|_q^{-1}u)^{\otimes(q-1)}\overline{E}(\|u\|_q), \tag{4.41}$$

where $\overline{S}$ *and* $\overline{E}$ *are the univariate superquantile and expected shortfall functions associated with the distribution of* $\|X\|_p$*.*

Proof. Denote by G the probability distribution function of $\|X\|_p$. Moreover, for $z \in \mathbb{R}$, let

$$\Psi(z) = \int_{-\infty}^{z} G(t)dt,$$

and $\varphi(x) = \Psi(\|x\|_p)$. Then, φ is convex by the composition between the non-decreasing function Ψ and $\|\cdot\|_p$, both convex, see [145][Theorem 5.1]. In addition, $\nabla\varphi(x) = F_q(x)$, which means that F_q is the gradient of a convex function. It remains to show that $F_q(X)$ follows the distribution $U^+_{d,q}$. From Lemma 4.19, $F_q(X)$ is the product of two independent random variables, $G(\|X\|_p)$ and $(\|X\|_p^{-1}X)^{\otimes(p-1)}$. On the one hand, $G(\|X\|_p)$ is uniformly distributed on $[0,1]$, by definition of G. On the other hand, the distribution of $\|X\|_p^{-1}X$ is uniform on $\mathbb{S}^{d,p}_+$, which implies (4.38). In order to compute $Q_q = F_q^{-1}$, it remains to invert (4.38). If $u = (\|x\|_p^{-1}x)^{\otimes(p-1)}G(\|x\|_p)$,

$$\|u\|_q^q = G(\|x\|_p)^q \sum_{k=1}^{d} \left(\frac{x_k}{\|x\|_p}\right)^p.$$

Thus, $\|u\|_q = G(\|x\|_p)$ which yields $\|u\|_q^{-1}u = (\|x\|_p^{-1}x)^{\otimes(p-1)}$. This rewrites

$$\|u\|_q^{-1}u\|x\|_p^{p-1} = x^{\otimes(p-1)},$$

where $\|x\|_p^{p-1} = G^{-(p-1)}(\|u\|_q)$, so

$$\|u\|_q^{-1}uG^{-(p-1)}(\|u\|_q) = x^{\otimes(p-1)}.$$

Finally, (4.39) follows by applying the exponent $q - 1 = 1/(p-1)$, which implies (4.40) and (4.41), completing the proof of Proposition 4.20.

□

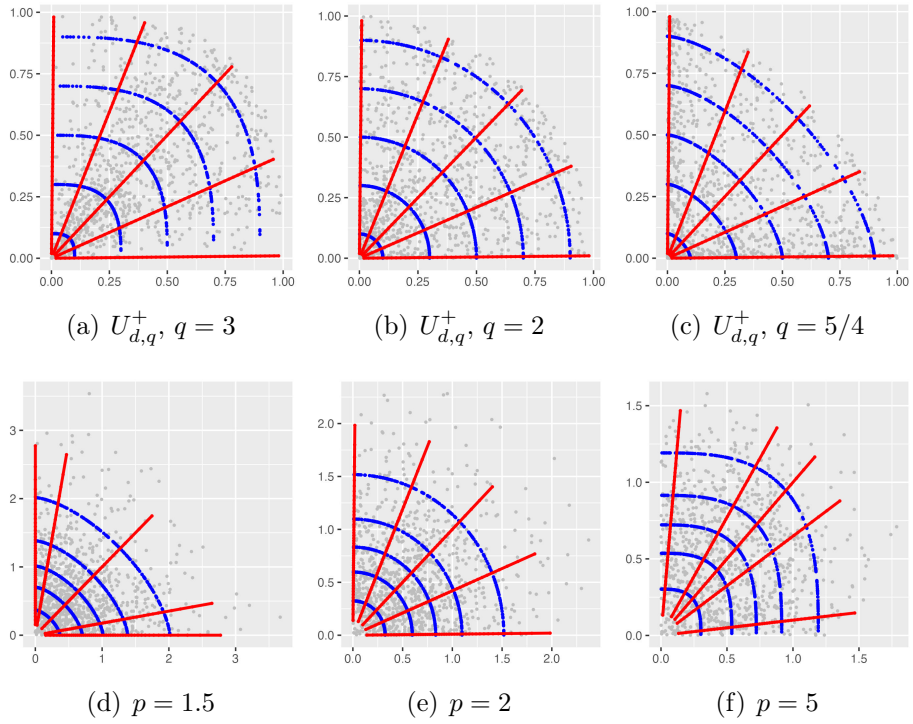
Remark 4.21. *It is well-known that in the special case* $p = 2$*, the distribution of* $(\|X\|_p^{-1}X)^{\otimes(p-1)}$ *is uniform on* $\mathbb{S}^{d,q}_+$ *where* $q = p = 2$*, see [33]. One may wonder if this property holds true for other choices of* $p > 1$*. This is actually not the case as we shall see now. As* q *is the Hölder conjugate of* p*, we clearly have* $\|X\|_p^{p-1} =$

$\|X^{\otimes(p-1)}\|_q$ *which implies that*

$$\left(\frac{X}{\|X\|_p}\right)^{\otimes(p-1)} = \frac{X^{\otimes(p-1)}}{\|X^{\otimes(p-1)}\|_q}.$$

All the components of X *share the same* $\mathcal{L}_p$ *distribution. Consequently, each* X_i^p *has a Gamma distribution with shape and scale parameters* $1/p$ *and* 1*. It implies that* $X_i^{p-1} = X_i^{p/q}$ *is the power* $1/q$ *of a Gamma distribution. We immediately deduce from Lemma 4.19 that, as soon as* $p \neq 2$*,* $(\|X\|_p^{-1}X)^{\otimes(p-1)}$ *is not uniformly distributed on* $\mathbb{S}_+^{d,q}$*.*

Figure 4.3 illustrates Proposition 4.20. Each column contains the reference measure $U_{d,q}^+$ for $q = p/(p-1)$ in the first line, and the associated distribution with *i.i.d.* components $X_i \sim \mathcal{L}_p$ below. Reference contours of orders $0.25, 0.5, 0.75$ are represented as well as the explicit MK quantile contours. Whereas the center-outward ordering intrinsic to U_d is natural for elliptical models, [33][Section 2.4], the left-to-right ordering may be more relevant under generalized Gamma models in $\mathbb{R}_+^d$.



(a) $U_{d,q}^+$, $q = 3$ (b) $U_{d,q}^+$, $q = 2$ (c) $U_{d,q}^+$, $q = 5/4$

(d) $p = 1.5$ (e) $p = 2$ (f) $p = 5$

FIGURE 4.3 – (First line) Reference measures $U_{d,q}^+$ for indices q conjugate with p. (Second line) Generalized Gamma distributions $\mathcal{L}_p$ and explicit quantile contours and sign curves.

4.4 Multivariate values and vectors at risk

Considering several univariate risks separately leads to neglecting the correlation structure between the components of a random vector. Hence, studying a notion of risk in a multivariate way is of major importance. We refer to [59][Section 2.1] for motivations for intrinsically multivariate risk analysis problems. The risk framework considers a vector of dependent losses $X \in \mathbb{R}^d_+$, where each component is a positive measure whose unit may be different from one to another. Naturally, the larger the value of a component, the greater the associated risk. On the one hand, we argue that real-valued risk measures are not sufficient in such setting. With real-valued random variables, real-valued risk measures are able to catch all the information needed. But, with the increase of the dimension, one may need vectors in $\mathbb{R}^d$ to get the directional information of the multivariate tails. With this in mind, we make here a proposal for vector-valued risk measures, namely *Vectors-at-Risk* and *Conditional-Vectors-at-Risk*, which aim to sum up the relevant information contained in the center-outward quantiles and superquantiles. Related to these vector-valued measures, we define *multivariate values-at-risk* and *multivariate conditional-values-at-risk* which are measures of the form of $\rho : \mathbb{R}^d_+ \to \mathbb{R}$ able to compare X and Y by $\rho(X) > \rho(Y)$. Importantly, the computation of these real-valued measures gives the vector-valued ones, hitting two targets with one shot without added complexity. On the other hand, a multivariate extension of the concepts of VaR and CVaR shall have the same interpretation as in $\mathbb{R}$. In dimension $d = 1$, the VaR of some risk at a level α is simply the quantile of order α. Accordingly, the CVaR is the superquantile of order α. These are to be understood respectively as the worst risk encountered with ν-probability α, and as the averaged risk beyond this quantile. Being an observation from the underlying distribution, it shall be vector-valued in our multivariate framework.

By construction, the MK quantile contour of order $\alpha \in [0, 1]$ contains the points having the most outward position with ν-probability α. Hence, we argue that a Vector-at-Risk of order α should belong to this contour. Even if it is always adapted to the geometry of ν, a MK quantile contour can either describe a *central* or a *bottom-left* area, by changing the reference measure. We explore both these possibilities, leaving it to the reader to decide which tool to adopt. In what follows, we rely on U_d and U_d^+, its restriction to $\mathbb{R}^d_+$. Of course, other p-spherical uniform distributions could be used, but we do not know if this would induce any benefit. Starting from the choice between *central* or *bottom-left* areas for MK quantile contours, the worst vectors of losses are the furthest from the origin in our context. Then, we suggest to select points from the center-outward quantile contour of order α with maximal norm, using Definition 4.2. *Conditional-Vectors-at-Risk* (CVaRs) are defined in the same way, but considering the center-outward superquantiles, with Definition 4.9.

Definition 4.22. *The Vector-at-Risk at level α is defined by*

$$VaR_\alpha(X) \in \operatorname{argsup}\{\ \|X\|_1\ ;\ X \in \mathcal{C}_\alpha\ \}.$$

Similarly, the Conditional-Vector-at-Risk is defined by

$$CVaR_\alpha(X) \in \operatorname{argsup}\{\ \|X\|_1\ ;\ X \in \mathcal{C}^s_\alpha\ \}.$$

The choice of $\|\cdot\|_1$ in the above definition depends on how to compare different risks. We emphasize that this choice is very distinct to the one of the reference measure : this consideration takes place once the contours are fixed. Without added information, we believe that two observations of same 1-norm shall be considered of same importance. Intuitively, when comparing multivariate risks, the risks per component add up. For example, if

$$x_1 = \begin{pmatrix}1\\0\end{pmatrix} \quad \text{and} \quad x_2 = \begin{pmatrix}0.5\\0.5\end{pmatrix}$$

belong to the same quantile contour of level α, choosing $\|\cdot\|_2$ instead of $\|\cdot\|_1$ in Definition 4.22 would induce to consider that x_1 is worst than x_2. This choice is coherent with the common practice of computing univariate VaR and CVaR on the random variable $\|X\|_1$ in financial applications, but a major difference is that it encodes the multivariate joint probability of the vector X before applying the sum. Thus, it takes into account the correlations, while providing more information, because typical values for component-wise risks can be retrieved through our Vectors-at-Risk. With this in mind, the Vector-at-Risk of order α is to be understood as the "worst" risk encountered with probability α, whereas the Conditional-Vector-at-Risk is the average risk beyond this "worst" observation, where the notion of "worst" is characterized here by $\|\cdot\|_1$ and the focus is either on central or on bottom-left areas. Thus, this generalizes the understanding of univariate VaR and CVaR. We displayed in Figure 4.4 the VaR_α and CVaR_α with quantile and superquantile contours estimated via EOT for $\varepsilon = 10^{-3}$. Each red point is the worst observation inside the corresponding quantile or superquantile regions.

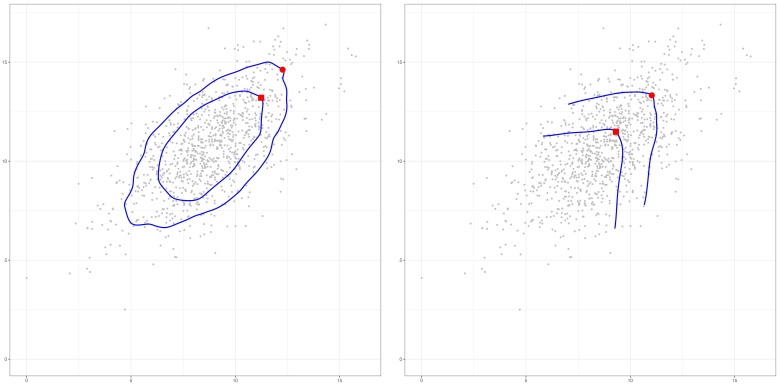
Intuitively, Definition 4.22 already gives a real-valued risk measure. Note that the set $\{\ \|X\|_1\ ;\ X \in \mathcal{C}_\alpha\ \}$ is the same as $\{\ \|Q_X(u)\|_1\ ;\ \|u\|_2 = \alpha\ \}$ for both the reference distribution U_d and its restriction U_d^+ to $\mathbb{R}^d_+$.

Definition 4.23. *The multivariate value-at-risk at level α is defined as*

$$\rho^Q_\alpha(X) = \sup_u\{\ \|Q_X(u)\|_1\ ;\ \|u\|_2 = \alpha\ \}.$$

Similarly, the multivariate conditional value-at-risk is defined as

$$\rho^S_\alpha(X) = \sup_u\{\ \|S_X(u)\|_1\ ;\ \|u\|_2 = \alpha\ \}.$$



(a) Center-outward reference U_d (b) Left-to-right reference U_d^+

FIGURE 4.4 – Gaussian distribution with VaR_α and CVaR_α in red and associated MK quantile and superquantile contours in blue, for $\alpha = 0.5$.

These *multivariate (conditional) values-at-risk* indicate on the spreadness of the point cloud drawn from a vector of losses, and contain information about its centrality and the correlations between its components. Empirical experiments have been conducted in Section 4.5.1 where they are compared with the *maximal correlation risk measure* from [15].

4.5 Numerical experiments.

In the experiments of the next section, we use the entropic map as a regularized empirical quantile map. Plugging Q_ε into Definitions 4.5 and 4.7 induces the entropic analogs

$$S_\epsilon(u) = \frac{1}{1 - \|u\|} \int_{\|u\|}^1 Q_\varepsilon\Big(t \frac{u}{\|u\|}\Big) dt \tag{4.42}$$

and

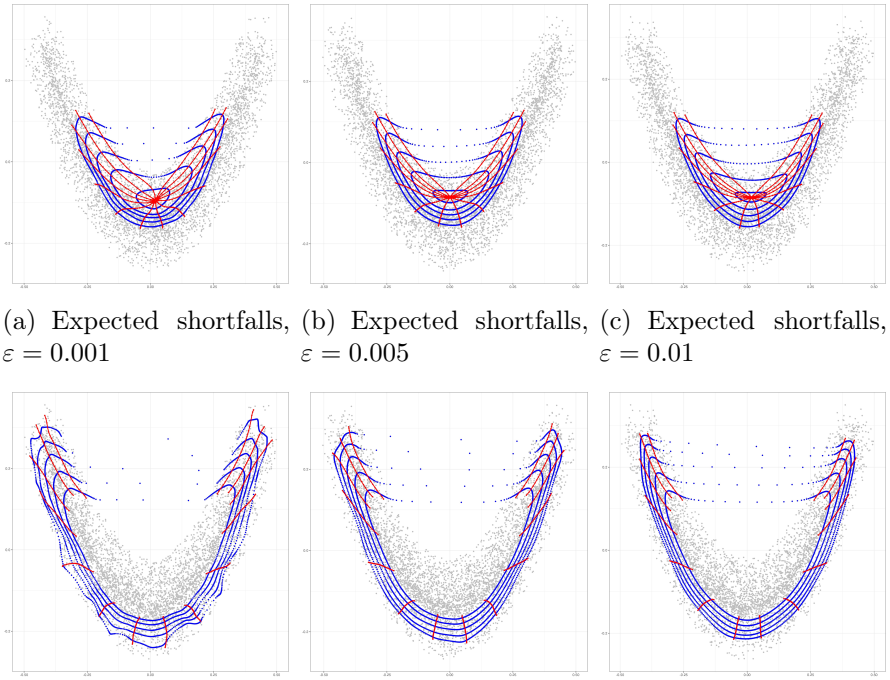
$$E_\epsilon(u) = \frac{1}{\|u\|} \int_0^{\|u\|} Q_\varepsilon\Big(t \frac{u}{\|u\|}\Big) dt. \tag{4.43}$$

In order to solve EOT between the empirical measure $\widehat{\nu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ and the reference distribution μ, we use the stochastic Robbins-Monro algorithm taken from [17, 71]. The resulting estimator is $\widetilde{Q}_\varepsilon^n$, defined in (2.2). Empirical counterparts $\widehat{S}_{\varepsilon,n}$ and $\widehat{E}_{\varepsilon,n}$ are obtained by plug-in estimators, replacing Q_ε by $\widetilde{Q}_\varepsilon^n$ within definitions (4.42) and (4.43). The regularization strength is taken as $\varepsilon = 10^{-3}$ and the integrals in (4.42) are estimated by Riemann sums.

4.5.1 Empirical study on simulated data

Descriptive plots for expected shortfalls and superquantiles

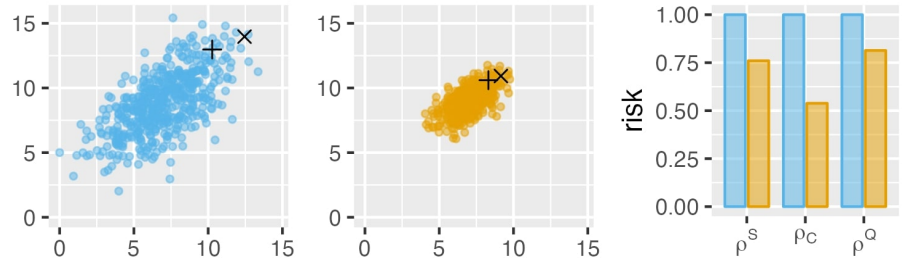
Descriptive plots associated with Definition 4.9 are given in Figure 4.5, to describe all the information contained in our new concepts. The continuous reference distribution U_d is transported to the discrete banana-shaped measure ν with support of size $n = 5000$, via our empirical center-outward expected shortfall (resp. superquantile) function. Note how the data splits into a central area and a periphery area by the range of points covered by both maps. These are satisfying estimators for the well-suited $\varepsilon = 10^{-3}$, that is to say the first column of Figure 4.5. As the regularization parameter ε for E_ε and S_ε grows, the contours concentrate around the mean vector of ν, which is a known feature of entropic optimal transport. One can observe that our regularized approach yields smooth interpolation between image points. For visualization purposes, the red points of averaged sign curves are linked by straight paths. The blue points are not, to illustrate how the contours capture the empty space in the middle of the non convex point cloud.



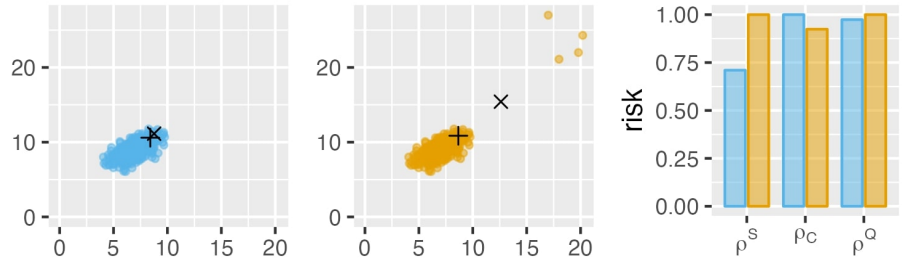
(a) Expected shortfalls, $\varepsilon = 0.001$ (b) Expected shortfalls, $\varepsilon = 0.005$ (c) Expected shortfalls, $\varepsilon = 0.01$

(d) Superquantiles, $\varepsilon = 0.001$ (e) Superquantiles, $\varepsilon = 0.005$ (f) Superquantiles, $\varepsilon = 0.01$

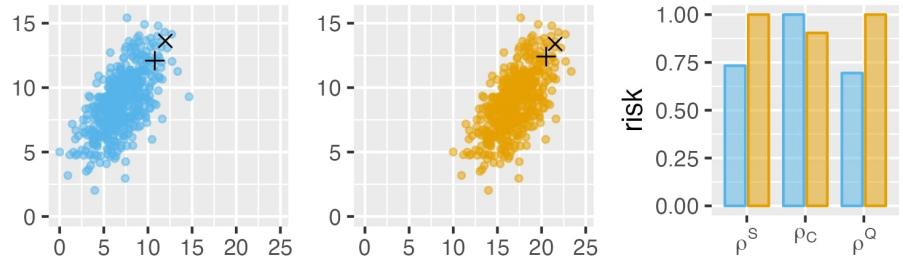
FIGURE 4.5 – Center-outward expected shortfall and superquantile contours of levels α in $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ in blue and averaged sign curves C_u in red.



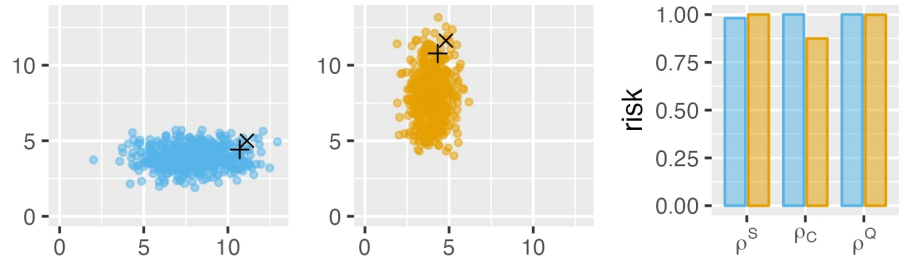
(a) Same Gaussians, reduced covariance matrix.



(b) Same Gaussians, added outliers.



(c) Same Gaussians with shift.



(d) Gaussians pointed in the vertical or horizontal direction

FIGURE 4.6 – VaRs (+) and CVaRs (×) on toy examples. Third column : real-valued risk measurements for the point clouds of the same line.

Risk measurements on toy examples

In dimension $d = 2$, an empirical distribution can be represented with a scatter plot and the riskiest observations are visible with the naked eye : they are located furthest from the origin. From this principle, we selected easy-to-handle situations to evaluate our risk measures in Figure 4.6. Each row refers to a situation where a blue scatter plot (first column) is to be compared with an orange one (second column). For each situation, our *(Conditional) Vectors-at-Risk* of order $\alpha = 0.75$ is illustrated on each scatter plot. Moreover, the associated ρ_α^Q and ρ_α^S are computed and compared to the maximal correlation risk mesure ρ_C from [15] in the third column. The higher the bar, the riskier the corresponding vector of losses. In view of their comparison, the measurements $\rho \in \{\rho_\alpha^Q, \rho_\alpha^S, \rho_C\}$ are rescaled. For Y_1 the distribution of the blue scatter plot and Y_2 the orange one, one considers, for the height of the bars,

$$\rho(Y_1)/\max\Big(\rho(Y_1), \rho(Y_2)\Big) \quad \text{and} \quad \rho(Y_2)/\max\Big(\rho(Y_1), \rho(Y_2)\Big).$$

For each situation, our VaR and CVaR provide typical observations in the multivariate tails. The *maximal-correlation risk measure* ρ_C performs as well as expected, while our CVaR succeeds in more situations. Indeed, ρ_C benefits from several theoretical properties but only measures the riskiness of $X - \mathbb{E}(X)$, hence it neglects the shift effects.

Figure 4.6(a) contains two Gaussian distributions with identical mean vectors and covariance matrices related through the multiplication by a positive real. This is well tackled by each real-valued risk measurement. The existence of more outliers must be taken into account similarly, as in figure 4.6(b), where the two scatter plots originate from the same underlying distribution, but some outliers are added to the orange one. The CVaR is much more sensitive to these outliers than the VaR, as in dimension $d = 1$. Note that ρ_C ignores the outliers and leads to the wrong decision in the sense than the blue distribution is considered as the riskiest one. In the situation of Figure 4.6(c), the orange scatterplot is identical to the blue one, but shifted to the right side, so that it must be the riskiest. As ρ_C ignores this shift, it induces the wrong decision. Finally, in the last example of Figure 4.6(d), making a decision on the relative risk between the underlying distributions requires a preference for one of the components. Here, the two situations reveal risks of same intensity but directed towards different directions. To inform on the underlying directional information, vector-valued risk measures such as our VaRs and CVaRs are needed in complement.

4.5.2 Risk measurements on wind gusts data

In this section, we illustrate our new multivariate risk measures on the analysis of a real dataset provided by the ExtremalDep R package, [16], and dedicated to the study of strong wind gusts. This dataset has previously been studied in [56, 76, 117]

in a context of risk measurement. The three variables are hourly wind gust (WG) in meters per second, wind speed (WS) in meters per second, and air pressure at sea level (DP) in millibars, recorded at Parcay-Meslay (France) between July 2004 and July 2013. We consider the 1450 weekly maximum of each measurement. Because the variables are of different nature, it is the precise framework where multivariate risk analysis is useful, rather than the aggregation of several variables.

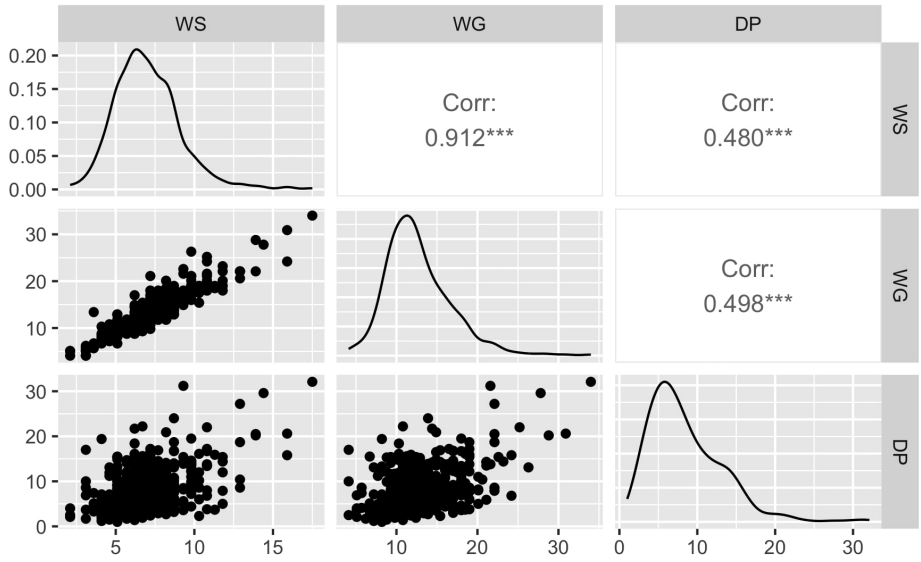
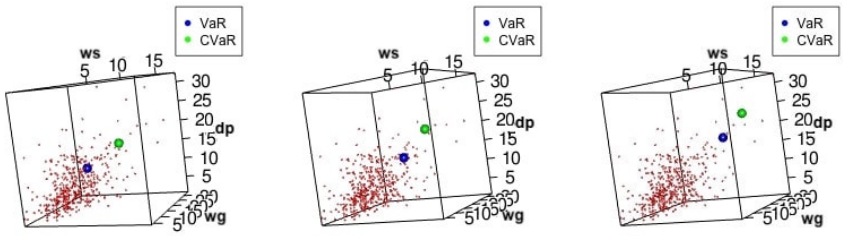


FIGURE 4.7 – Three dimensional wind gust data set.

Figure 4.7 represents our three-dimensional dataset with pair scatterplots under the diagonal and Pearson correlation values above. The diagonal represents empirical density functions of each variable. Upper-right dependence can be observed and has physical explanations. Strong wind gusts occur with stormy weather, during which strong wind speed and high air pressure are frequently recorded.



(a) Risk level 0.25 (b) Risk level 0.5 (c) Risk level 0.75

FIGURE 4.8 – Vectors-at-Risk and Conditional-Vectors-at-Risk.

Figure 4.8 represents the three dimensional empirical distribution in red together with our vectorial risk measures. With the increase of the dimension, such representative plots are no longer convenient. Rather, these measurements can be retrieved as in Table 4.1.

	$VaR_{0.25}$	$VaR_{0.5}$	$VaR_{0.75}$
WS	8.21	9.98	12.36
WG	15.06	18.45	21.58
DP	11.92	13.65	17.84

	$CVaR_{0.25}$	$CVaR_{0.5}$	$CVaR_{0.75}$
WS	11.40	12.37	14.26
WG	21.28	22.43	26.24
DP	16.68	19.86	23.15

TABLE 4.1 – Components of (Conditional) Vectors-at-Risk for several risk levels.

This summarizes the targeted information contained in the dataset. For instance, with a given probability 0.25, 0.5 or 0.75, one shall expect, at worst, wind gusts of respective speed 15.06, 18.45, 21.58 m/s. With respectively same probability, averaged observations beyond these worst cases shall lie around 21.28, 22.43, 26.24 m/s. As in [117], we interpret these results thanks to the Beaufort scale. Note that this scale does not capture the speed of wind gusts, as it usually averages over 10 minutes, by convention. Values between 13.9 and 17.1 m/s can be considered as high winds. Strong winds begin with 17.2 m/s, with severely strong winds above 20.7 m/s up to 24.4 m/s. Severely strong winds can cause slight structural damage, but less than storms for values around 24.5-28.4 m/s. Above and up to 32.6 m/s, violent storms are very rarely experienced and cause widespread damage. Wind speeds greater than 32.7 m/s correspond to hurricanes. Thanks to Table 4.1, with probability 0.75, the worst scenarios in Parcay-Meslay for wind gusts are severely strong winds. Moreover, along the tail events corresponding to 25% of occurrences, one shall expect wind gusts of same speed as storms. Also, to illustrate the fact that considering univariate measures leads to underestimating the risk, we display in Figure 4.2 traditional univariate Values-at-Risk of WG, the wind gusts. For example, its median is 11.80, (strong breeze), and must be compared with 18.45, (strong winds), the second coordinate of $VaR_{0.5}$.

Min	1st quartile	Median	3rd quartile	Max
4.10	9.80	11.80	14.90	34

TABLE 4.2 – Univariate quantiles of our variable WG.

Such differences have statistical foundings. The median depends on the *univariate* empirical distribution of WG to describe a probability of 1/2. Conversely, our multivariate $VaR_{0.5}$ encodes the *multivariate* joint probability of the whole point cloud (WG,WS,DP). This can be summarized by the fact that univariate risk measures *neglect the correlations*, which legitimates the use of multivariate risk analysis.

4.6 On the class of integrated quantile functions

We end this chapter with bibliographical notes. The role of integrated quantile functions in dimension $d = 1$ has been highlighted in [79]. Obviously, this chapter

belongs to the line of works trying to extend such functions to the setting $d > 1$. In Section 4 of [85], two different approaches are discussed, in view of generalizing

$$\alpha \mapsto \int_0^\alpha Q(t)dt. \tag{4.44}$$

Hereafter, we bridge their concepts with ours, namely Definition 4.7 that extends (4.44). Recall that $\mathbb{S}^{d-1} = \{\varphi \in \mathbb{R}^d : \|\varphi\|_2 = 1\}$ and $\mathbb{P}_S$ is the uniform probability measure on $\mathbb{S}^{d-1}$. There, the (absolute) center-outward Lorenz function from [85][Definition 4] writes as

$$L_{\mathbf{X}\pm} : \alpha \mapsto \mathbb{E}[X\mathbb{1}_{X\in\mathbb{C}_\alpha}] = \int_{\mathbb{S}^{d-1}} \alpha E(\alpha\varphi)d\mathbb{P}_S(\varphi). \tag{4.45}$$

This being said, we believe that our proposed concepts of integration along sign curves contain more information, namely directional. When targeting the contributions of central regions to the expectation, $L_{\mathbf{X}\pm}$ provides meaningful concepts of Lorenz curves, but this approach is insufficient for superquantiles and multivariate tails, as illustrated in Example 4.4.

Another important generalization of (4.44) is the one from [60] about multivariate Lorenz curves. A main difference between the concepts of [60] and [85] is the reference distribution, either the uniform on the unit hypercube or the spherical uniform. Here, Lorenz curves aim to visualize inequalities within a given population. In this context, one could focus either on the contribution of middle classes as in [85], or on the one of the bottom of the population as in [60], in a way that these works are in fact complementary. Furthermore, an (absolute) center-outward Lorenz potential function is proposed in [85][Definition 6], $\alpha \mapsto \mathbb{E}_\varphi[\psi(\alpha\varphi)]$, that is real valued for $\varphi \in \mathbb{S}^{d-1}$. This is quite natural because, at the core of the univariate considerations of [79], one can retrieve the property that the quantile function is the gradient of a convex potential, that is the definition of the center-outward quantile function Q. Also, it has a physical interpretation, with a measurement of the *work* of the quantile function. There, one might note that our center-outward expected shortfall function draws a connection between [85][Definition 4] and [85][Definition 6], through (4.45) and (4.13). Somehow, this strengthens the relation between $L_{\mathbf{X}\pm}$ and the potential function ψ. Studying deeper such connections between existing notions might lead to interesting results, but is left for further work.

Conclusion

In this thesis, we studied the notion of Monge-Kantorovich quantiles and related objects, from their estimation to their application for data analysis. We motivated the use of entropic regularization to obtain a smooth and monotone estimator of the MK quantile function, that can be defined without assuming finite second-order moments. On the computational side, a convenient rewriting allowed to formulate a stochastic algorithm that converges almost surely towards the solution of *continuous* EOT. This algorithm lies in an infinite dimensional Banach space of Fourier coefficients, which makes it difficult to analyze. Moreover, our approach was extended for spherical data, by replacing Fourier coefficients in $\mathbb{R}^d$ by spherical harmonics on the hypersphere $\mathbb{S}^{d-1}$. By the way, it has been showed that the entropic map can be generalized on $\mathbb{S}^{d-1}$, building on the particular form of Riemannian gradients on $\mathbb{S}^{d-1}$. In addition, we proposed a multivariate definition for the superquantile and expected shortfall functions, based on optimal transportation ideas. On the real line, these integrated quantile functions characterize distributions and their weak convergence, a fact that continues to hold for our multivariate extensions. Our new definitions come with multivariate extensions of value-at-risk and conditional-value-at-risk, that are fundamental measurements of the riskiness of a point cloud.

Here are some future research perspectives.

A broader use for our stochastic algorithm ?

Our stochastic algorithm introduced in Chapter 2 is of particular interest when the number of observations n is large and prevents the storage of the cost matrix. A major appealing feature is that the computational cost at each iteration is independent from n and only depends on the size of a regular grid in $[0,1]^d$ fixed by the user. Nonetheless, the size of such a grid grows exponentially with the dimension, that is a main bottleneck of our algorithm. Because this grid especially underlines our implementation of the Fast Fourier Transform, one can hope that more sophisticated transforms could alleviate this, see for instance [141]. The extension to other orthonormal basis may also be pursued, like we did with spherical harmonics on $\mathbb{S}^2$ in Chapter 3, where, again, the limitation with respect to the dimension is due to memory constraints. One might also think of non orthonormal expansions, such as dictionaries, [168], that may preserve the convexity of the objective function, being linear in the parameters.

It may also be applied beyond the context of MK quantiles. For instance, in the linear optimal transport framework, [174], one also looks for Monge maps from a fixed reference distribution μ. The aim is to construct Hilbertian embeddings of probability distributions, typically on images in $[0,1]^2$ with a number of pixels n potentially high. Note that the same objective can be investigated via regularized potentials instead of Monge maps, [9]. In this setting, the reference measure μ is typically an average of given images, which questions our restriction $\mu = \text{Unif}([0,1]^d)$. But the knowledge of the reference measure μ is mostly required to ensure that the optimal potential verifies $\mathbf{u}_\varepsilon : [0,1]^d \to \mathbb{R}$, and for computing integrals with respect to μ. Thus, one could in fact replace $\mu = \text{Unif}([0,1]^d)$ by other measures supported on $[0,1]^d$.

About superquantiles ?

Perspectives related to our multivariate notions of superquantile and expected shortfall functions are both theoretical and practical.

Firstly, regularity assumptions for the weak convergence of Theorems 4.15 and 4.16 might be alleviated, as they are stronger than the ones on the real line [145]. It would also be interesting to question limit distributions of empirical Vectors-at-Risk and Conditional-Vectors-at-Risk, in light with recent results from [150] for the maximum tail correlation. In a broader sense, the convergence of empirical entropic optimal transport towards its population counterpart is an active field, [13, 78, 123, 140], and it remains to study to what extent existing results adapt to the convergence of $\widehat{S}_{\varepsilon,n}$ and $\widehat{E}_{\varepsilon,n}$.

Besides, one might note that our estimation of MK superquantiles and expected shortfalls can be expensive as it relies on Riemann sums. In $\mathbb{R}$, these objects have different characterizations that allow their estimation by stochastic algorithms, [21]. It would be interesting to study if such equivalent formulations are possible for our multivariate concepts, in view of similar numerical procedures.

Extending our definitions to extreme quantile levels and building related procedures is another line of study. For instance, [171][Theorem 4.3] ensures the existence of a MK quantile map which is stable as moving further to the tail contours, which is not the case with the spherical uniform U_d. Our proposed center-outward superquantiles can be defined with other spherical reference measures, thus one can also calibrate these superquantiles as the work of [171] suggests.

In addition, the many applications of the univariate superquantile function in other fields than risk measurement, such as superquantile regression, [145], or optimization with a superquantile loss, [105], appeal for further work, even more so with Theorems 4.15 and 4.16. Superquantiles-based statistical tests of Goodness-of-Fit may be constructed, where one expects advantages with respect to the robustness. Also, our risk measures appeal for being used in practical settings, as done in [65].

A

Annex for Chapter 2

A.1 Additional proofs

A.1.1 Proof of Proposition 2.5

We first state a result about Fréchet differentiation under Lebesgue integrals, that follows from [166, Lemma A.2], and which extends well-known results on the differentiation of integral functionals. For the proof of a similar result, we also refer to the unpublished note [97].

Lemma A.1 (Leibniz's rules of Fréchet differentiation)**.** *Let $(\Theta, \|\cdot\|)$ be an infinite dimensional Banach space and σ a finite measure on a measurable space $\mathbb{T}$. Let $\theta_0 \in \Theta$ and denote by $B(\theta_0, R) \subset \Theta$ the ball of center θ_0 and radius R. Consider a function $f : \Theta \times \mathbb{T} \to \mathbb{R}$ that is Fréchet differentiable at θ_0 (for every $t \in \mathbb{T}$), and suppose that there exists $K \in L^1(\sigma)$ such that, for all $\theta_1, \theta_2 \in B(\theta_0, R)$ and all $t \in \mathbb{T}$,*

$$|f(\theta_1, t) - f(\theta_2, t)| \leq K(t)\|\theta_1 - \theta_2\|.$$

Then, the integral functional $F : \Theta \to \mathbb{R}$ defined by $F(\theta) = \int_{\mathbb{T}} f(\theta, t)d\sigma(t)$ is Fréchet differentiable at θ_0 and

$$DF(\theta_0) = \int_{\mathbb{T}} D_\theta f(\theta_0, t)d\sigma(t),$$

where $D_\theta f(\theta_0, t)$ denotes the Fréchet derivative of $\theta \mapsto f(\theta, t)$ at θ_0.

In what follows, we will apply Lemma A.1 with $\Theta = \bar{\ell}_1(\Lambda)$, $\mathbb{T} = \mathcal{X}$ and $\sigma = \mu$ to obtain the expression of the Fréchet differential of H_ε. Let us first prove that,

for every $y \in \mathcal{Y}$, the function $h_\varepsilon(\cdot, y) : \bar{\ell}_1(\Lambda) \to \mathbb{R}$ defined in (2.5) is Fréchet differentiable. To this end, we introduce the function $g_y(\cdot, x) : \bar{\ell}_1(\Lambda) \to \mathbb{R}$ defined as

$$g_y(\theta, x) = \frac{1}{\varepsilon}\left(\sum_{\lambda\in\Lambda} \theta_\lambda \phi_\lambda(x) - c(x, y)\right) \tag{A.1}$$

and $G_y(\theta) = \int_{\mathcal{X}} \exp(g_y(\theta, x))d\mu(x)$. In this way, one has that $h_\varepsilon(\theta, y) = \varepsilon \log G_y(\theta) + \varepsilon$. For every $x \in \mathcal{X}$, the function $\theta \mapsto \exp(g_y(\theta, x))$ is clearly Fréchet differentiable and, for $\tau \in \bar{\ell}_1(\Lambda)$,

$$D_\theta \exp(g_y(\theta, x))[\tau] = \frac{1}{\varepsilon} \sum_{\lambda\in\Lambda} \exp(g_y(\theta, x))\phi_\lambda(x)\tau_\lambda. \tag{A.2}$$

Moreover, it is a bounded linear operator from $\bar{\ell}_1(\Lambda)$ to $\mathbb{R}$. In what follows, we identify this operator to the infinite-dimensional vector

$$D_\theta \exp(g_y(\theta, x)) = \frac{1}{\varepsilon} \exp(g_y(\theta, x)) \left(\overline{\phi_\lambda(x)}\right)_{\lambda\in\Lambda}.$$

From now on, let $\theta_0 \in \bar{\ell}_1(\Lambda)$ and $R > 0$. Then, for any $\theta_1, \theta_2 \in B(\theta_0, R)$, the mean value theorem for functions defined on a Banach space implies that

$$|\exp(g_y(\theta_1, x)) - \exp(g_y(\theta_2, x))| \le \sup_{\theta\in B(\theta_0,R)} \|D_\theta \exp(g_y(\theta, x))\|_{op} \|\theta_1 - \theta_2\|_{\ell_1}, \tag{A.3}$$

where the operator norm of $D_\theta \exp(g_y(\theta, x))$ is defined as

$$\|D_\theta \exp(g_y(\theta, x))\|_{op} = \sup_{\|\tau\|_{\ell_1}\le 1} |D_\theta \exp(g_y(\theta, x))[\tau]|.$$

Since

$$|D_\theta \exp(g_y(\theta, x))[\tau]| = \left|\frac{1}{\varepsilon} \sum_{\lambda\in\Lambda} \exp(g_y(\theta, x))\phi_\lambda(x)\tau_\lambda\right| \le \frac{1}{\varepsilon} \exp(g_y(\theta, x)) \sum_{\lambda\in\Lambda} |\tau_\lambda|,$$

one has that, for any $\theta \in \bar{\ell}_1(\Lambda)$,

$$\|D_\theta \exp(g_y(\theta, x))\|_{op} \le \frac{1}{\varepsilon} \exp(g_y(\theta, x)) \le \frac{1}{\varepsilon} \exp\left(\frac{\sum_{\lambda\in\Lambda} |\theta_\lambda| + c(x, y)}{\varepsilon}\right) \tag{A.4}$$

Consequently, let

$$K_y(x) = \frac{1}{\varepsilon} \exp\left(\frac{c(x, y)}{\varepsilon}\right) \sup_{\theta\in B(\theta_0,R)} \exp\left(\frac{\|\theta\|_{\ell_1}}{\varepsilon}\right).$$

It follows from (A.3) and (A.4) that, for all $\theta_1, \theta_2 \in B(\theta_0, R)$ and $x \in \mathcal{X}$,

$$|\exp(g_y(\theta_1, x)) - \exp(g_y(\theta_2, x))| \le K_y(x)\|\theta_1 - \theta_2\|_{\ell_1}.$$

Because μ is compactly supported, it is obvious that, for all $y \in \mathcal{Y}$, the function K_y belongs to $L^1(\mu)$, and therefore, by Lemma A.1, we conclude that G_y and $h_\varepsilon(\cdot, y)$ are Fréchet differentiable and that the linear operator $D_\theta h_\varepsilon(\theta, y)$ is identified as an element of $\bar{\ell}_\infty(\Lambda)$ given by

$$D_\theta h_\varepsilon(\theta, y) = \left(\frac{\int_{\mathcal{X}} \overline{\phi_\lambda(x)} \exp(g_y(\theta, x)) d\mu(x)}{\int_{\mathcal{X}} \exp(g_y(\theta, x)) d\mu(x)} \right)_{\lambda \in \Lambda}. \tag{A.5}$$

Similarly, to prove that the function $H_\varepsilon(\theta) = \int_{\mathcal{Y}} h_\varepsilon(\theta, y) d\nu(y)$ is Fréchet differentiable, it is sufficient to bound the operator norm of $D_\theta h_\varepsilon(\theta, y)$ for $\theta \in B(\theta_0, R)$. Recalling that

$$F_{\theta,y}(x) = \frac{\exp\left(g_y(\theta, x)\right)}{\int_{\mathcal{X}} \exp\left(g_y(\theta, x)\right) d\mu(x)},$$

we remark that, for any $\tau \in \bar{\ell}_1(\Lambda)$,

$$|D_\theta h_\varepsilon(\theta, y)[\tau]| = \left| \sum_{\lambda \in \Lambda} \int_{\mathcal{X}} F_{\theta,y}(x) \phi_\lambda(x) d\mu(x) \tau_\lambda \right| \leq \int_{\mathcal{X}} F_{\theta,y}(x) d\mu(x) \sum_{\lambda \in \Lambda} |\tau_\lambda| = \|\tau\|_{\ell_1}.$$

Therefore, $\|D_\theta h_\varepsilon(\theta, y)\|_{op} \leq 1$ which proves inequality (2.20). It also means that $D_\theta h_\varepsilon(\theta, y)$ can be identified as an element of $\bar{\ell}_\infty(\Lambda)$. Thus, arguing as previously, that is by combining the mean value theorem with Lemma A.1, we obtain that $H_\varepsilon(\theta)$ is Fréchet differentiable with

$$DH_\varepsilon(\theta) = \int_{\mathcal{Y}} D_\theta h_\varepsilon(\theta, y) d\nu(y) = \left(\int_{\mathcal{Y}} \frac{\partial h_\varepsilon(\theta, y)}{\partial \theta_\lambda} d\nu(y) \right)_{\lambda \in \Lambda}$$

which can also be identified as an element of $\bar{\ell}_\infty(\Lambda)$ such that $\|DH_\varepsilon(\theta)\|_{op} = \|DH_\varepsilon(\theta)\|_{\ell_\infty}$ satisfies inequality (2.22) by combining inequality (2.20) together with the fact that ν is a probability measure. This achieves the proof of Proposition 2.5.

□

A.1.2 Proof of Proposition 2.6

First, let us recall that, for $(\Theta, \|\cdot\|)$ a given Banach space, a function $f : \Theta \to \mathbb{R}$ is twice Fréchet differentiable if Df is Fréchet differentiable. In this case, the second order Fréchet derivative of f at θ_0 is denoted by $D^2 f(\theta_0)$ and it is identified as an element of $L(\Theta \times \Theta, \mathbb{R})$ the set of continuous bilinear mapping from $\Theta \times \Theta$ to $\mathbb{R}$. Moreover, the operator norm of $D^2 f(\theta_0)$ is defined as

$$\|D^2 f(\theta_0)\|_{op} = \sup_{\|\theta\|_{\ell_1} \leq 1, \|\theta'\|_{\ell_1} \leq 1} |D^2 f(\theta_0)[\theta, \theta']|.$$

To derive the expression of the second order Fréchet derivative of the functions $h_\varepsilon(\cdot, y)$ and H_ε, we use similar arguments to those in the proof of Proposition 2.5. First, recall from (A.5) that the Fréchet derivative $D_\theta h_\varepsilon(\theta, y)$ is the linear operator defined as

$$D_\theta h_\varepsilon(\theta, y) = \left(\int F_{\theta,y}(x)\overline{\phi_\lambda(x)} d\mu(x) \right)_{\lambda \in \Lambda} = \int \psi_y(x, \theta) d\mu(x)$$

where $\psi_y(x, \theta) : \bar{\ell}_1(\Lambda) \to \mathbb{R}$ is the linear operator

$$\psi_y(x, \theta)[\tau] = \sum_{\lambda \in \Lambda} F_{\theta,y}(x)\phi_\lambda(x)\tau_\lambda = \sum_{\lambda \in \Lambda} F_{\theta,y}(x)\overline{\phi_\lambda(x)}\overline{\tau_\lambda}.$$

As a standard strategy, we aim to derive this with respect to θ. From (A.1), one has that

$$F_{\theta,y}(x) = \frac{\exp(g_y(\theta, x))}{G_y(\theta)}.$$

Therefore, using (A.2) combined with the differentiability of $G_y(\theta)$,

$$D_\theta F_{\theta,y}(x)[\tau] = \frac{1}{\varepsilon} \left(\sum_{\lambda \in \Lambda} \tau_\lambda \phi_\lambda(x) F_{\theta,y}(x) - F_{\theta,y}(x) \int_{\mathcal{X}} \sum_{\lambda \in \Lambda} \tau_\lambda \phi_\lambda(z) F_{\theta,y}(z) d\mu(z) \right). \tag{A.6}$$

Thus, the mapping $\theta \mapsto \psi_y(x, \theta)[\tau]$ is clearly Fréchet differentiable, and its Fréchet derivative can be identified as the following symmetric bilinear mapping from $\bar{\ell}_1(\Lambda) \times \bar{\ell}_1(\Lambda)$ to $\mathbb{R}$

$$\begin{aligned} D_\theta \psi_y(x, \theta)[\tau, \tau'] &= \frac{1}{\varepsilon} \sum_{\lambda, \lambda' \in \Lambda} \tau'_{\lambda'} \overline{\tau_\lambda} \left(\phi_{\lambda'}(x)\overline{\phi_\lambda(x)} F_{\theta,y}(x) - \phi_{\lambda'}(x) F_{\theta,y}(x)\overline{\phi_\lambda(x)} F_{\theta,y}(x) \right) \\ &= \frac{1}{\varepsilon} \sum_{\lambda, \lambda' \in \Lambda} \tau'_{\lambda'} \overline{\tau_\lambda} \phi_{\lambda'}(x)\overline{\phi_\lambda(x)} F_{\theta,y}(x) - \sum_{\lambda \in \Lambda} \tau'_\lambda \phi_\lambda(x) F_{\theta,y}(x) \overline{\sum_{\lambda \in \Lambda} \tau_\lambda \phi_\lambda(x) F_{\theta,y}(x)}. \end{aligned}$$

We now compute an upper bound for the norm of this linear operator. One can observe that, for $\tau = \tau'$,

$$|D_\theta \psi_y(x, \theta)[\tau, \tau]| \leq \frac{1}{\varepsilon} F_{\theta,y}(x) \|\tau\|_{\ell_1}^2, \tag{A.7}$$

thanks to the elementary fact that

$$\sum_{\lambda \in \Lambda} \tau_\lambda \phi_\lambda(x) F_{\theta,y}(x) \overline{\sum_{\lambda \in \Lambda} \tau_\lambda \phi_\lambda(x) F_{\theta,y}(x)} = \left| \sum_{\lambda \in \Lambda} \tau_\lambda \phi_\lambda(x) F_{\theta,y}(x) \right|^2 \geq 0. \tag{A.8}$$

Then, using the equality,

$$4 D_\theta \psi_y(x, \theta)[\tau, \tau'] = D_\theta \psi_y(x, \theta)[\tau + \tau', \tau + \tau'] - D_\theta \psi_y(x, \theta)[\tau - \tau', \tau - \tau']$$

combined with the upper bound (A.7), we obtain that

$$4|D_\theta \psi_y(x,\theta)[\tau,\tau']| \leq \frac{1}{\varepsilon} F_{\theta,y}(x) \left(\|\tau+\tau'\|_{\ell_1}^2 + \|\tau-\tau'\|_{\ell_1}^2 \right).$$

Therefore, we immediately obtain that

$$\sup_{\|\tau\|_{\ell_1}\leq 1, \|\tau'\|_{\ell_1}\leq 1} |D_\theta \psi_y(x,\theta)[\tau,\tau']| \leq \frac{2}{\varepsilon} F_{\theta,y}(x).$$

Consequently, we may proceed as in the proof of Proposition 2.5 to obtain that $h_\varepsilon(\cdot, y)$ is twice Fréchet differentiable and that its second Fréchet derivative is the following symmetric bilinear mapping from $\bar{\ell}_1(\Lambda) \times \bar{\ell}_1(\Lambda)$ to $\mathbb{R}$

$$\begin{aligned} D_\theta^2 h_\varepsilon(\theta,y)[\tau,\tau'] = {} & \frac{1}{\varepsilon} \sum_{\lambda'\in\Lambda} \sum_{\lambda\in\Lambda} \tau'_{\lambda'} \overline{\tau_\lambda} \int_{\mathcal{X}} \phi_{\lambda'}(x) \overline{\phi_\lambda}(x) F_{\theta,y}(x) d\mu(x) \\ & - \frac{1}{\varepsilon} \left(\sum_{\lambda\in\Lambda} \tau'_\lambda \int_{\mathcal{X}} \phi_\lambda(x) F_{\theta,y}(x) d\mu(x) \right) \overline{\left(\sum_{\lambda\in\Lambda} \tau_\lambda \int_{\mathcal{X}} \phi_\lambda(x) F_{\theta,y}(x) d\mu(x) \right)}. \end{aligned}$$

Note that, for $\tau = \tau'$, an application of Jensen's inequality with respect to the probability measure $F_{\theta,y}(x)d\mu(x)$ implies that $D_\theta^2 h_\varepsilon(\theta,y)[\tau,\tau] \geq 0$. Moreover, it follows once again from (A.8) together with the elementary fact that $\int_{\mathcal{X}} F_{\theta,y} d\mu = 1$, that

$$D_\theta^2 h_\varepsilon(\theta,y)[\tau,\tau] \leq \frac{1}{\varepsilon} \|\tau\|_{\ell_1}^2. \tag{A.9}$$

Hereafter, we deduce from the equality

$$4 D_\theta^2 h_\varepsilon(\theta,y)[\tau,\tau'] = D_\theta^2 h_\varepsilon(\theta,y)[\tau+\tau',\tau+\tau'] - D_\theta^2 h_\varepsilon(\theta,y)[\tau-\tau',\tau-\tau'],$$

the positivity of $D_\theta^2 h_\varepsilon(\theta,y)[\tau-\tau',\tau-\tau']$ and inequality (A.9), that

$$4|D_\theta^2 h_\varepsilon(\theta,y)[\tau,\tau']| \leq D_\theta^2 h_\varepsilon(\theta,y)[\tau+\tau',\tau+\tau'] \leq \frac{1}{\varepsilon} \|\tau+\tau'\|_{\ell_1}^2.$$

It ensures that

$$\|D_\theta^2 h_\varepsilon(\theta,y)\|_{op} = \sup_{\|\tau\|_{\ell_1}\leq 1, \|\tau'\|_{\ell_1}\leq 1} |D_\theta^2 h_\varepsilon(\theta,y)[\tau,\tau']| \leq \frac{1}{\varepsilon},$$

which proves inequality (2.24).

Finally, combining the above upper bound on $\|D_\theta^2 h_\varepsilon(\theta,y)\|_{op}$ and using again an adaptation of Lemma A.1 to obtain a Leibniz's formula for the second order Fréchet differentiation under the integral sign, one can prove that $H_\varepsilon(\theta)$ is twice Fréchet differentiable by integrating $D_\theta^2 h_\varepsilon(\theta,y)$ with respect to $d\nu(y)$, which implies that $D^2 H_\varepsilon(\theta)$ is the linear operator defined by (2.25). Moreover, the upper bound (2.26) follows from inequality (2.24) and the fact that ν is a probability measure, which completes the proof of Proposition 2.6. □

A.1.3 Proof of Proposition 2.7

For $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and for $(\theta^{(1)}, \theta^{(2)}) \in \bar{\ell}_1(\Lambda) \times \bar{\ell}_1(\Lambda)$, denote

$$e_1(x,y) = \exp\left(\frac{\sum_{\lambda\in\Lambda} \theta_\lambda^{(1)}\phi_\lambda(x) - c(x,y)}{\epsilon}\right)$$

and

$$e_2(x,y) = \exp\left(\frac{\sum_{\lambda\in\Lambda} \theta_\lambda^{(2)}\phi_\lambda(x) - c(x,y)}{\epsilon}\right).$$

We have, for all $0 < t < 1$, and for a fixed $y \in \mathcal{Y}$, that

$$\begin{aligned} t h_\varepsilon(\theta^{(1)}, y) + (1-t) h_\varepsilon(\theta^{(2)}, y) &= \varepsilon t \log \int_{\mathcal{X}} e_1(\cdot, y) d\mu + (1-t)\log \int_{\mathcal{X}} e_2(\cdot, y) d\mu + \varepsilon \\ &= \varepsilon \log\left(\left(\int_{\mathcal{X}} e_1(\cdot, y) d\mu\right)^t \left(\int_{\mathcal{X}} e_2(\cdot, y) d\mu\right)^{1-t}\right) + \varepsilon. \end{aligned} \tag{A.10}$$

Hereafter, applying Hölder's inequality to $f(x) = e_1^t(x,y)$ and $g(x) = e_2^{1-t}(x,y)$ with Hölder conjugates $p = 1/t$ and $q = 1/(1-t)$, we obtain that

$$\int_{\mathcal{X}} f(x)g(x)\, d\mu(x) \leq \left(\int_{\mathcal{X}} e_1(x,y) d\mu(x)\right)^t \left(\int_{\mathcal{X}} e_2(x,y) d\mu(x)\right)^{1-t}. \tag{A.11}$$

However, one can observe that

$$\int_{\mathcal{X}} fg\, d\mu = \int_{\mathcal{X}} \exp\left(\frac{t\sum_{\lambda\in\Lambda} \theta_\lambda^{(1)}\phi_\lambda(x) + (1-t)\sum_{\lambda\in\Lambda} \theta_\lambda^{(2)}\phi_\lambda(x) + c(x,y)}{\varepsilon}\right) d\mu(x),$$

which ensures that

$$\varepsilon \log \int_{\mathcal{X}} f(x)g(x)\, d\mu(x) + \varepsilon = h_\varepsilon(t\theta^{(1)} + (1-t)\theta^{(2)}, y).$$

Hence, combining the above equality with (A.10) and (A.11), we obtain that

$$h_\varepsilon(t\theta^{(1)} + (1-t)\theta^{(2)}, y) \leq t h_\varepsilon(\theta^{(1)}, y) + (1-t) h_\varepsilon(\theta^{(2)}, y),$$

which proves the convexity of $\theta \mapsto h_\varepsilon(\theta, y)$. Since $H_\varepsilon(\theta) = \mathbb{E}\left[h_\varepsilon(\theta, Y)\right]$, we also obtain the convexity of the function H_ε. Furthermore, assume that Hölder's inequality (A.11) becomes an equality, which means that the functions f^p and g^q are linearly dependent in $L^1(\mu)$. This would mean that it exists $\beta_y > 0$ such that $e_1(x,y) = \beta_y e_2(x,y)$ for all $x \in \mathcal{X}$. Applying the logarithm, this equality is equiva-

lent to

$$\frac{1}{\varepsilon}\sum_{\lambda\in\Lambda}\theta_\lambda^{(1)}\phi_\lambda(x) = \log\beta_y + \frac{1}{\varepsilon}\sum_{\lambda\in\Lambda}\theta_\lambda^{(2)}\phi_\lambda(x).$$

By integrating the above equality with respect to μ, and from our normalization condition (2.3), the equality case in the Hölder inequality (A.11) implies that $\log\beta_y = 0$ and thus $\beta_y = 1$. But then one has that $e_1(x,y) = e_2(x,y)$, implying that $\sum_{\lambda\in\Lambda}(\theta_\lambda^{(1)} - \theta_\lambda^{(2)})\phi_\lambda(x) = 0$ for all $x \in \mathcal{X}$. Hence, we necessarily have that $\theta^{(1)} = \theta^{(2)}$ which yields a contradiction. Therefore, the function $\theta \mapsto h_\varepsilon(\theta, y)$ is strictly convex. Since $H_\varepsilon(\theta) = \mathbb{E}\left[h_\varepsilon(\theta, Y)\right]$ this also implies the strict convexity of H_ε, which achieves the proof of Proposition 2.7. □

A.2 Optimal transport with periodicity constraints

In this section, we focus our attention on conditions such that the dual potential (D_ε) is periodic at the boundary of $\mathcal{X}$.

A.2.1 The case of the standard quadratic cost

Using classical results in the analysis of multiple Fourier series (see e.g. [164, Corollary 1.8]), assuming that the Fourier coefficients $\theta^0 = (\theta^0_\lambda)_{\lambda\in\Lambda}$ of u_0 form an absolutely convergent series implies that u_0 can be extended as a continuous and $\mathbb{Z}^d$-periodic function on $\mathbb{R}^d$. Hence, under this assumption, u_0 has to be a continuous function that is constant at the boundary of $\mathcal{X} = [0,1]^d$. However, for the quadratic cost $c(x,y) = \frac{1}{2}\|x-y\|^2$, we are not aware of standard results on the regularity of optimal transport (through smoothness assumptions on ν) that would imply periodic properties of u_0 and its derivatives at the boundary of $\mathcal{X}$.

A.2.2 The quadratic cost on the torus

Nevertheless, guaranteeing the periodicity of u_0 and the summability of its Fourier coefficients is feasible by considering the setting $\mathcal{X} = \mathcal{Y} = \mathbb{T}^d$, where $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ is the d-dimensional torus, that is endowed with the usual distance

$$d_{\mathbb{T}^d}(x,y) = \min_{\lambda\in\mathbb{Z}^d}\|x-y+\lambda\|.$$

Hereafter, we identify the torus as the set of equivalence classes $\{x+\lambda \,:\, \lambda\in\mathbb{Z}^d\}$ for $x \in [0,1)^d$, and we use the notation $[x] = x + \lambda_0$ where $\lambda_0 \in \mathbb{Z}^d$ is such that $\|x+\lambda\|$ is minimal for $\lambda \in \mathbb{Z}^d$. We also recall that a function $u : \mathbb{T}^d \to \mathbb{R}$ can be identified as a $\mathbb{Z}^d$-periodic function on $\mathbb{R}^d$. Finally, one can observe that for a given $y \in \mathbb{T}^d$, the cost function $c(x,y) = \frac{1}{2}d^2_{\mathbb{T}^d}(x,y)$ is almost everywhere differentiable,

and its gradient is (see e.g. [151, Section 1.3.2])

$$\nabla_x c(x,y) = [x-y],$$

at every $x \notin y + \{\partial\Omega + \mathbb{Z}^d\}$ where $\partial\Omega$ denotes the boundary of $\Omega = [-\frac{1}{2}, \frac{1}{2}]^d$.

Assuming that the probability measure ν is also supported on the d-dimensional torus $\mathbb{T}^d$ allows to use existing results for optimal transport on the torus (see e.g. [38], [116, Section 2.2] and [151, Section 1.3.2]). Formally, taking $\mathcal{X} = \mathcal{Y} = \mathbb{T}^d$ implies that ν is considered as a periodic positive Radon measure on $\mathbb{R}^d$ with $\nu(\mathbb{T}^d) = 1$, and that μ is understood as the Lebesgue measure on $\mathbb{R}^d$. Note that this setting is not restrictive, as it allows to treat the example of an absolutely continuous measure ν with support on $[0,1]^d$ whose density f_ν takes a constant value on the boundary of $[0,1]^d$, implying that f_ν can be extended over $\mathbb{R}^d$ as a $\mathbb{Z}^d$-periodic function.

Then, thanks to the identification of $u : \mathbb{T}^d \to \mathbb{R}$ as a $\mathbb{Z}^d$-periodic function on $\mathbb{R}^d$, it follows that

$$\inf_{x\in\mathbb{T}^d} \left\{ \frac{1}{2} d^2_{\mathbb{T}^d}(x,y) - u(x) \right\} = \inf_{x\in\mathbb{R}^d} \left\{ \frac{1}{2}\|x-y\|^2 - u(x) \right\}.$$

Therefore, it is equivalent to define the conjugate of a function $u : \mathbb{T}^d \to \mathbb{R}$ with respect to the cost $c(x,y) = \frac{1}{2} d^2_{\mathbb{T}^d}(x,y)$ or to the quadratic cost $c(x,y) = \frac{1}{2}\|x-y\|^2$ using the periodization of u over $\mathbb{R}^d$. Now, using results on optimal transport on $\mathbb{T}^d$, previously established in [38] or [116, Proposition 4], it follows that

(i) there exists a unique optimal transport map $Q : \mathbb{T}^d \to \mathbb{T}^d$ from μ to ν such that

$$Q = \operatorname*{argmax}_{T\,:\,T\#\mu=\nu} \mathbb{E}\left(d^2_{\mathbb{T}^d}(X, T(X)) \right),$$

(ii) $Q(x) = x - \nabla u_0(x)$ where u_0 is a $\mathbb{Z}^d$-periodic function on $\mathbb{R}^d$ that is a solution of the dual problem (D_0) with $\mathcal{X} = \mathcal{Y} = \mathbb{T}^d$ and $c(x,y) = \frac{1}{2} d^2_{\mathbb{T}^d}(x,y)$,

(iii) $\|Q(x) - x\|^2 = d^2_{\mathbb{T}^d}(x, Q(x))$ for almost every $x \in \mathbb{R}^d$.

Entropically regularized optimal transport on the torus has also been recently considered in [23] and [36, Section E]. One can thus also consider the dual formulation of entropic OT as in (D_ε) with the cost $c(x,y) = \frac{1}{2} d^2_{\mathbb{T}^d}(x,y)$.

We conclude this section on optimal transport on the torus by a discussion on the regularity of the optimal dual functions in the un-regularized case $\varepsilon = 0$. For $s \in \mathbb{N}$, we denote by $\mathcal{C}^s(\mathbb{T}^d)$, the set of $\mathbb{Z}^d$-periodic functions f on $\mathbb{R}^d$ having everywhere defined continuous partial derivatives. Then, the following regularity result holds as an immediate application of results from [38] and [116, Theorem 5].

Lemma A.2. *Let u_0 be a a solution of the dual problem (D_0) with $\mathcal{X} = \mathcal{Y} = \mathbb{T}^d$ and $c(x,y) = \frac{1}{2} d^2_{\mathbb{T}^d}(x,y)$. Suppose that the probability distribution ν is absolutely*

continuous with a density f_ν that is lower and upper bounded by positive constants. Assume further that $f_\nu \in \mathcal{C}^{s-1}(\mathbb{T}^d)$ for some $s > 1$. Then, u_0 belongs to $\mathcal{C}^{s+1}(\mathbb{T}^d)$.

Consequently, under the assumptions of Lemma A.2, one has that if $f_\nu \in \mathcal{C}^{s-1}(\mathbb{T}^d)$ for some $s > d/2 - 1$, then u_0 belongs to $\mathcal{C}^k(\mathbb{T}^d)$ with $k > d/2$. Therefore, by standard results for multiple Fourier series (see e.g. [164, Corollary 1.9]), one has that $\sum_{\lambda\in\Lambda} |\theta^0_\lambda| < +\infty$. Hence we can conclude that if the density of ν is sufficiently smooth (and is upper and lower bounded by positive constants), then the Fourier serie of u_0 actually belongs to $\ell_1(\Lambda)$.

B

Annex for Chapter 3

B.1 Invariance properties

Invariance properties of empirical versions of $\mathbf{F}$ and $\mathbf{Q}$ were shown in [84]. The same holds in fact for the population counterparts, with the same argument : the transport problem inherits invariance from the Riemannian distance.

Proposition B.1. *In dimension d, let $\mathbf{O}$ be a $d \times d$ orthogonal matrix and let $\nu \in \mathbf{B}_2$. Denote by $\mathbf{O}_{\#}\nu$ the distribution of $\mathbf{O}Z$ if $Z \sim \nu$, and by $\mathbf{F}_Z, \mathbf{Q}_Z$, (resp. $\mathbf{F}_{\mathbf{O}Z}, \mathbf{Q}_{\mathbf{O}Z}$), the distribution and quantile functions of ν, (resp. $\mathbf{O}_{\#}\nu$). Then,*

$$\mathbf{F}_{\mathbf{O}Z}(\mathbf{O}z) = \mathbf{O}\mathbf{F}(z).$$

and

$$\mathbf{Q}_{\mathbf{O}Z}(\mathbf{O}z) = \mathbf{O}\mathbf{Q}(z).$$

Proof. Note that the Kantorovich problem, equivalent to (P_0) when $\nu \in \mathbf{B}_2$, minimizes

$$\int_{\mathbb{S}^2}\int_{\mathbb{S}^2} c(x,y)d\pi(x,y),$$

over the set of joint probabilities π supported on $\mathbb{S}^2 \times \mathbb{S}^2$ with marginals $\mu_{\mathbb{S}^2}, \nu$, see *e.g.* [84]. Because $c(\mathbf{O}x, \mathbf{O}y) = c(x,y)$, the transport problem between $\mu_{\mathbb{S}^2}$ and ν is equivalent to the one between $\mathbf{O}_{\#}\mu_{\mathbb{S}^2}$ and $\mathbf{O}_{\#}\nu$. Going back to Monge's problem (P_0), it immediately follows that the Monge map $T_{\#}\mu_{\mathbb{S}^2} = \nu$ verifies

$$\mathbf{O}T(z) = T_{\mathbf{O}Z}(\mathbf{O}z).$$

Up to interverting the reference and the target measures, the result follows. □

The following corollary is straightforward.

Corollary B.2. *For any* $\tau \in [0,1]$, $\mathbf{O}\mathcal{C}_\tau = \mathbf{Q}_{\mathbf{O}Z}(\mathbf{O}\mathcal{C}_\tau^U)$

B.2 Explicit forms

Closed-form expressions of $\mathbf{F}$ for rotationally invariant distributions were given in [84] and simplify in dimension $d = 3$, allowing to deduce the inverse map $\mathbf{Q}$. Let $Z \sim \nu$ be such a random vector with axis $\pm\theta_M$. Then, assume that ν has density

$$z \in \mathbb{S}^2 \mapsto c_f f(z^T\theta_M),$$

for f some positive *angular function* and c_f a normalizing constant. For $r \in [-1,1]$, denote by

$$F_f(r) = \int_{-1}^{r} f(s)ds \Big/ \int_{-1}^{1} f(s)ds$$

the distribution function of $Z^T\theta_M$ and by $Q_f = F_f^{-1}$ its quantile function. Then, letting $F_f^*(r) = 2F_f(r) - 1$, the directional distribution function of Z writes

$$\mathbf{F}(z) = F_f^*(z^T\theta_M)\theta_M + \sqrt{1 - F_f^*(z^T\theta_M)^2}S_{\theta_M}(z). \tag{B.1}$$

For instance, taking $f(s) = \exp(\kappa z^T\theta_M)$ corresponds to the von Mises-Fisher distribution with location parameter θ_M and concentration parameter $\kappa \in \mathbb{R}_+$. Crucially, the transport (B.1) reduces to univariate transport along the axis $\theta_M = \mathbf{F}(\theta_M)$. If $\theta_M = (0,0,1)^T$, that is to say *up to some rotation* thanks to Proposition B.1 and Corollary B.2, this corresponds to changing the latitude *w.r.t.* the usual coordinate system (3.1). Indeed, as soon as $\theta_M = (0,0,1)^T$,

$$\mathbf{F}(z) = F_f^*(z_3)\theta_M + \sqrt{1 - F_f^*(z_3)^2}\frac{(z_1, z_2, 0)^T}{\|(z_1, z_2, 0)^T\|}.$$

The third coordinate is changed to $F_f^*(z_3)$, and the other coordinates are adapted to the constraint $\mathbf{F}(z) \in \mathbb{S}^2$. This rewrites, in accordance with (3.1),

$$\mathbf{F}(z) = \mathbf{F}(\Phi(\theta,\phi)) = \Phi(\overline{\theta},\phi) \qquad \text{for} \qquad \overline{\theta} = \arccos\Big((F_f^*)(z_3)\Big). \tag{B.2}$$

Consequently, to get the inverse map $\mathbf{Q} = \mathbf{F}^{-1}$, it suffices to change the pseudo latitude of $\mathbf{F}(z) \in \mathcal{C}_\tau^U$ *w.r.t.* the axis $\pm\theta_M$. If $x = \mathbf{F}(z)$, $x_3 = F_f^*(z_3)$ and $z_3 = (F_f^*)^{-1}(x_3)$, that is

$$\mathbf{Q}(x) = \mathbf{Q}(\Phi(\theta,\phi)) = \Phi(\tilde{\theta},\phi) \qquad \text{for} \qquad \tilde{\theta} = \arccos\Big((F_f^*)^{-1}(x_3)\Big). \tag{B.3}$$

As highlighted in [84], this shows that MK quantile contours coincide with Mahalanobis ones from [108] under the rotationally symmetric model.

Bibliographie

[1] Carlo Acerbi, Claudio Nordio et Carlo Sirtori. *Expected Shortfall as a Tool for Financial Risk Management*. arXiv :cond-mat/0102304. 2001.

[2] Carlo Acerbi et Dirk Tasche. « On the coherence of Expected Shortfall ». In : *Journal of Banking & Finance* 26.7 (2002), p. 1487-1503.

[3] C Agostinelli et M Romanazzi. « Nonparametric analysis of directional data based on data depth ». In : *Environmental and ecological statistics* 20 (2013), p. 253-270.

[4] Jason M Altschuler, Jonathan Niles-Weed et Austin J Stromme. « Asymptotics for semidiscrete entropic optimal transport ». In : *SIAM Journal on Mathematical Analysis* 54.2 (2022), p. 1718-1741.

[5] Jose Ameijeiras-Alonso et Rosa M Crujeiras. « Directional statistics for wildfires ». In : *Applied directional statistics*. Chapman et Hall/CRC, 2018, p. 203-226.

[6] Sara Armaut, Roland Diel et Thomas Laloë. *On some depth-based risk measurement for high losses*. hal-03933578. 2023.

[7] Francis Bach. « Self-concordant analysis for logistic regression ». In : *Electronic Journal of Statistics* 4.none (2010), p. 384 -414.

[8] Francis R. Bach. « Adaptivity of averaged stochastic gradient descent to local strong convexity for logistic regression ». In : *Journal of Machine Learning Research* 15.1 (2014), p. 595-627.

[9] François Bachoc et al. « Gaussian processes on distributions based on regularized optimal transport ». In : *International Conference on Artificial Intelligence and Statistics*. PMLR. 2023, p. 4986-5010.

[10] Vic Barnett. « The ordering of multivariate data ». In : *Journal of the Royal Statistical Society : Series A (General)* 139.3 (1976), p. 318-344.

[11] Eustasio del Barrio, Alberto González Sanz et Marc Hallin. « Nonparametric multiple-output center-outward quantile regression ». In : *Journal of the American Statistical Association* just-accepted (2024), p. 1-43.

[12] Eustasio del Barrio et Jean-Michel Loubes. « Central limit theorems for empirical transportation cost in general dimension ». In : *Ann. Probab.* 47.2 (2019), p. 926-951.

[13] Eustasio del Barrio et al. « An improved central limit theorem and fast convergence rates for entropic transportation costs ». In : *SIAM Journal on Mathematics of Data Science* 3.5 (2023), p. 639-669.

[14] Franck Barthe et al. « A probabilistic approach to the geometry of the ℓ_p^n-ball ». In : *The Annals of Probability* 33.2 (2005), p. 480 -513.

[15] J. Beirlant et al. « Center-outward quantiles and the measurement of multivariate risk ». In : *Insurance : Mathematics and Economics* 95.C (2020), p. 79-100.

[16] B Beranger, S Padoan et G Marcon. *ExtremalDep : extremal dependence models*. R Package. 2021.

[17] Bernard Bercu et Jérémie Bigot. « Asymptotic distribution and convergence rates of stochastic algorithms for entropic optimal transportation between probability measures ». In : *The Annals of Statistics* 49.2 (2021), p. 968 -987.

[18] Bernard Bercu, Jeremie Bigot et Gauthier Thurin. *Monge-Kantorovich superquantiles and expected shortfalls with applications to multivariate risk measurements*. 2023.

[19] Bernard Bercu, Jérémie Bigot et Gauthier Thurin. *Stochastic optimal transport in Banach Spaces for regularized estimation of multivariate quantiles*. 2023.

[20] Bernard Bercu, Jérémie Bigot et Gauthier Thurin. *Regularized estimation of Monge-Kantorovich quantiles for spherical data*. 2024.

[21] Bernard Bercu, Manon Costa et Sébastien Gadat. « Stochastic approximation algorithms for superquantiles estimation ». In : *Electronic Journal of Probability* 26 (2021), p. 1-29.

[22] Bernard Bercu et al. « A stochastic Gauss-Newton algorithm for regularized semi-discrete optimal transport ». In : *Information and Inference : A Journal of the IMA* (mai 2022). issn : 2049-8772.

[23] Robert J. Berman. « The Sinkhorn Algorithm, Parabolic Optimal Transport and Geometric Monge-Ampére Equations ». In : *Numer. Math.* 145.4 (2020), p. 771-836.

[24] Espen Bernton, Promit Ghosal et Marcel Nutz. *Entropic Optimal Transport : Geometry and Large Deviations*. arXiv. 2021. eprint : 2102.04397 (math.OC).

[25] Filippo Bigi et al. « Fast evaluation of spherical harmonics with sphericart ». In : *The Journal of Chemical Physics* 159.6 (2023).

[26] Vladimir Igorevich Bogachev et Maria Aparecida Soares Ruas. *Measure theory*. T. 1. Springer, 2007.

[27] Y. Brenier. « Polar factorization and monotone rearrangement of vector-valued functions ». In : *Comm. Pure Appl. Math.* 44.4 (1991), p. 375-417. issn : 0010-3640.

[28] Sébastien Bubeck. « Convex Optimization : Algorithms and Complexity ». In : *Found. Trends Mach. Learn.* 8.3-4 (2015). issn : 1935-8237.

[29] Jun Cai, Huameng Jia et Tiantian Mao. « A multivariate CVaR risk measure from the perspective of portfolio risk management ». In : *Scandinavian Actuarial Journal* 2022.3 (2022), p. 189-215.

[30] Guillaume Carlier et al. « Convergence of entropic schemes for optimal transport and gradient flows ». In : *SIAM Journal on Mathematical Analysis* 49.2 (2017), p. 1385-1418.

[31] Guillaume Carlier et al. « Vector quantile regression and optimal transport, from theory to numerics ». In : *Empirical Economics* 62.1 (2022), p. 35-62.

[32] Probal Chaudhuri. « On a geometric notion of quantiles for multivariate data ». In : *Journal of the American statistical association* 91.434 (1996), p. 862-872.

[33] Victor Chernozhukov et al. « Monge–Kantorovich depth, quantiles, ranks and signs ». In : *The Annals of Statistics* 45.1 (2017), p. 223 -256.

[34] Sinho Chewi et Aram-Alexandre Pooladian. « An entropic generalization of Caffarelli's contraction theorem via covariance inequalities ». In : *Comptes Rendus. Mathématique* 361 (nov. 2023), p. 1471-1482.

[35] Gregory S Chirikjian et Alexander B Kyatkin. *Engineering applications of noncommutative harmonic analysis : with emphasis on rotation and motion groups*. CRC press, 2000.

[36] L. Chizat et al. « Faster Wasserstein Distance Estimation with the Sinkhorn Divergence ». In : *Proc. NeurIPS'20*. 2020.

[37] Samuel Cohen, Brandon Amos et Yaron Lipman. « Riemannian convex potential maps ». In : *International Conference on Machine Learning*. PMLR. 2021, p. 2028-2038.

[38] Dario Cordero-Erausquin. In : 329.3 (1999), p. 199-202. issn : 0764-4442.

[39] Dario Cordero-Erausquin, Robert J McCann et Michael Schmuckenschläger. « A Riemannian interpolation inequality à la Borell, Brascamp and Lieb ». In : *Inventiones mathematicae* 146.2 (2001), p. 219-257.

[40] Areski Cousin et Elena Di Bernardino. « On multivariate extensions of Value-at-Risk ». In : *Journal of Multivariate Analysis* 119 (2013), p. 32-46. issn : 0047-259X.

[41] Areski Cousin et Elena Di Bernardino. « On Multivariate Extensions of Conditional-Tail-Expectation ». In : *Insurance : Mathematics and Economics* 55 (2014), p. 272-282.

[42] Dennis D Cox. *The Theory of Statistics and Its Applications*. 2004.

[43] Harald Cramér. *Mathematical methods of statistics*. T. 26. Princeton university press, 1999.

[44] J. A. Cuesta et C. Matrán. « Notes on the Wasserstein Metric in Hilbert Spaces ». In : *Annals of Probability* 17 (1989), p. 1264-1276.

[45] J.A. Cuesta-Albertos et A. Nieto-Reyes. « The Tukey and the random Tukey depths characterize discrete distributions ». In : *Journal of Multivariate Analysis* 99.10 (2008), p. 2304-2311. issn : 0047-259X.

[46] M. Cuturi et G. Peyré. *Computational Optimal Transport*. Book available at https ://optimaltransport.github.io/book/, 2017.

[47] Marco Cuturi. « Sinkhorn Distances : Lightspeed Computation of Optimal Transport ». In : *Advances in Neural Information Processing Systems* 26 (2013).

[48] Marco Cuturi, Michal Klein et Pierre Ablin. « Monge, Bregman and Occam : Interpretable Optimal Transport in High-Dimensions with Feature-Sparse Maps ». In : t. 202. Proceedings of Machine Learning Research. 2023, p. 6671-6682.

[49] Nabarun Deb, Bhaswar B Bhattacharya et Bodhisattva Sen. *Pitman Efficiency Lower Bounds for Multivariate Distribution-Free Tests Based on Optimal Transport*. arXiv. 2021.

[50] Nabarun Deb et Bodhisattva Sen. « Multivariate rank-based distribution-free nonparametric testing using measure transportation ». In : *Journal of the American Statistical Association* 118.541 (2023), p. 192-207.

[51] Eustasio Del Barrio, Alberto González-Sanz et Marc Hallin. « A note on the regularity of optimal-transport-based center-outward distribution and quantile functions ». In : *J. Multivar. Anal.* 180 (2020).

[52] Alex Delalande. « Nearly tight convergence bounds for semi-discrete entropic optimal transport ». In : *International Conference on Artificial Intelligence and Statistics*. PMLR. 2022, p. 1619-1642.

[53] Philippe Delanoë et Grégoire Loeper. « Gradient estimates for potentials of invertible gradient–mappings on the sphere ». In : *Calculus of Variations and Partial Differential Equations* 26.3 (2006), p. 297-311.

[54] Houyem Demni, Amor Messaoud et Giovanni C Porzio. « The cosine depth distribution classifier for directional data ». In : *Applications in Statistical Computing : From Music Data Analysis to Industrial Quality Improvement* (2019), p. 49-60.

[55] Houyem Demni et Giovanni Camillo Porzio. « Directional DD-classifiers under non-rotational symmetry ». In : *2021 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. 2021, p. 1-6.

[56] Elena Di Bernardino et Clémentine Prieur. « Estimation of the multivariate conditional tail expectation for extreme risk levels : Illustration on environmental data sets ». In : *Environmetrics* 29.7 (2018).

[57] Elena Di Bernardino et al. « Plug-in estimation of level sets in a non-compact setting with applications in multivariate risk theory ». In : *ESAIM : Probability and Statistics* 17 (2013), p. 236-256.

[58] Jean-Luc Dortet-Bernadet et Nicolas Wicker. « Model-based clustering on the unit sphere with an illustration using gene expression profiles ». In : *Biostatistics* 9.1 (2008), p. 66-80.

[59] Ivar Ekeland, Alfred Galichon et Marc Henry. « Comonotonic measures of multivariate risks ». In : *Mathematical Finance* 22.1 (2012), p. 109-132.

[60] Yanqin Fan et al. *Lorenz map, inequality ordering and curves based on multidimensional rearrangements*. arXiv. 2022.

[61] OP Ferreira, AN Iusem et SZ Németh. « Concepts and techniques of optimization on the sphere ». In : *Top* 22 (2014), p. 1148-1170.

[62] Jean Feydy et al. « Interpolating between optimal transport and mmd using sinkhorn divergences ». In : *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR. 2019, p. 2681-2690.

[63] Alessio Figalli. « On the continuity of center-outward distribution and quantile functions ». In : *Nonlinear Analysis* 117 (2018), p. 413-421.

[64] Rémi Flamary et al. « POT : Python Optimal Transport ». In : *Journal of Machine Learning Research* 22.78 (2021), p. 1-8.

[65] João Pedro M Franco. *Multivariate Risk Analyzes in Cryptocurrency Market : An Optimal Transport Theory Approach*. 2023.

[66] Maximiliano Frungillo. *Discrete Approximation of Optimal Transport on Compact Spaces*. 2024.

[67] Alfred Galichon et Marc Henry. « Dual theory of choice with multivariate risks ». In : *Journal of Economic Theory* 147.4 (2012), p. 1501-1516.

[68] Eduardo García-Portugués, Davy Paindaveine et Thomas Verdebout. « On optimal tests for rotational symmetry against new classes of hyperspherical distributions ». In : *Journal of the American Statistical Association* 115.532 (2020), p. 1873-1887.

[69] Ferdinand Genans-Boiteux et al. *Semi-Discrete Optimal Transport : Nearly Minimax Estimation With Stochastic Gradient Descent and Adaptive Entropic Regularization*. 2024.

[70] Aude Genevay. « Entropy-regularized Optimal Transport for Machine Learning ». Thèse de doct. Université Paris sciences et lettres, 2019.

[71] Aude Genevay et al. « Stochastic Optimization for Large-scale Optimal Transport ». In : *Advances in neural information processing systems* 29 (2016).

[72] Aude Genevay et al. « Sample complexity of sinkhorn divergences ». In : *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR. 2019, p. 1574-1583.

[73] Promit Ghosal, Marcel Nutz et Espen Bernton. « Stability of entropic optimal transport and Schrödinger bridges ». In : *Journal of Functional Analysis* 283.9 (2022), p. 109622.

[74] Promit Ghosal et Bodhisattva Sen. « Multivariate Ranks and Quantiles using Optimal Transport : Consistency, Rates, and Nonparametric Testing ». In : *The Annals of Statistics* 50.2 (2022), p. 1012-1037.

[75] Anil K Ghosh et Probal Chaudhuri. « On maximum depth and related classifiers ». In : *Scandinavian Journal of Statistics* 32.2 (2005), p. 327-350.

[76] Yuri Goegebeur, Armelle Guillou et Jing Qin. « Dependent conditional tail expectation for extreme levels ». In : *Stochastic Processes and their Applications* 171 (2024), p. 104330. issn : 0304-4149.

[77] Ziv Goldfeld et al. « Limit theorems for entropic optimal transport maps and Sinkhorn divergence ». In : *Electronic Journal of Statistics* 18.1 (2024), p. 980-1041.

[78] Alberto Gonzalez-Sanz, Jean-Michel Loubes et Jonathan Niles-Weed. *Weak limits of entropy regularized Optimal Transport : potentials, plans and divergences.* arXiv. 2022.

[79] Alexander A Gushchin et Dmitriy A Borzykh. « Integrated quantile functions : properties and applications ». In : *Modern Stochastics : Theory and Applications* 4.4 (2017), p. 285-314.

[80] Marc Hallin. « From Mahalanobis to Bregman via Monge and Kantorovich : Towards a "General Generalized Distance" ». In : *Sankhya B* 80.Suppl 1 (2018), p. 135-146.

[81] Marc Hallin. « Measure Transportation and Statistical Decision Theory ». In : *Annual Review of Statistics and Its Application* 9 (2022), p. 401-424.

[82] Marc Hallin, Daniel Hlubinka et Šárka Hudecová. « Efficient fully distribution-free center-outward rank tests for multiple-output regression and MANOVA ». In : *Journal of the American Statistical Association* 118.543 (2023), p. 1923-1939.

[83] Marc Hallin, Davide La Vecchia et Hang Liu. « Center-outward R-estimation for semiparametric VARMA models ». In : *Journal of the American Statistical Association* 117.538 (2022), p. 925-938.

[84] Marc Hallin, Hang Liu et Thomas Verdebout. « Nonparametric measure-transportation-based methods for directional data ». In : *Journal of the Royal Statistical Society Series B : Statistical Methodology* (2024).

[85] Marc Hallin et Gilles Mordant. *Center-Outward Multiple-Output Lorenz Curves and Gini Indices a measure transportation approach.* Working Papers ECARES. ULB – Universite Libre de Bruxelles, 2022.

[86] Marc Hallin, D. Vecchia et Hang Liu. « Rank-based testing for semiparametric VAR models : A measure transportation approach ». In : *Bernoulli* 29.1 (fév. 2023), p. 229-273.

[87] Marc Hallin et al. « Distribution and quantile functions, ranks and signs in dimension d : A measure transportation approach ». In : *The Annals of Statistics* 49.2 (2021), p. 1139 -1165.

[88] Brittany Froese Hamfeldt et Axel GR Turnquist. « A convergence framework for optimal transport on the sphere ». In : *Numerische Mathematik* 151.3 (2022), p. 627-657.

[89] Trevor Hastie et al. *The elements of statistical learning : data mining, inference, and prediction.* T. 2. Springer, 2009.

[90] Konstantin Hauch et Claudia Redenbach. *Quantiles and depth for directional data from elliptically symmetric distributions.* 2022.

[91] Janet E Heffernan et Jonathan A Tawn. « A conditional approach for multivariate extreme values (with discussion) ». In : *Journal of the Royal Statistical Society : Series B (Statistical Methodology)* 66.3 (2004), p. 497-546.

[92] Zhen Huang et Bodhisattva Sen. *Multivariate symmetry : Distribution-free testing via optimal transport*. arXiv. 2023.

[93] Peter J Huber et Elvezio M Ronchetti. *Robust statistics*. John Wiley & Sons, 2011.

[94] Rebecka Jörnsten. « Clustering and classification based on the L1 data depth ». In : *Journal of Multivariate Analysis* 90.1 (2004), p. 67-89.

[95] Jost Jurgen. *Postmodern analysis*. eng. 2nd edition. Universitext. Berlin : Springer, 2003. isbn : 3-540-43873-4.

[96] Hubert Kalf. « On the expansion of a function in terms of spherical harmonics in arbitrary dimensions ». In : *Bulletin of the Belgian Mathematical Society-Simon Stevin* 2.4 (1995), p. 361-380.

[97] Ohad Kammar. « A note on Fréchet differentiation under Lebesgue integrals ». http ://denotational.co.uk/notes/kammar-a-note-on-frechet-differentiation-under-lebesgue-integrals.pdf. 2016.

[98] Oliver Dimon Kellogg. *Foundations of potential theory*. T. 31. Springer Science & Business Media, 2012.

[99] J. Kitagawa, Q. Mérigot et Thibert B. « Convergence of a Newton algorithm for semi-discrete optimal transport ». In : *Journal of the European Math Society* 21.9 (2019), p. 2603-2651.

[100] Roger Koenker et Gilbert Bassett Jr. « Regression quantiles ». In : *Econometrica : journal of the Econometric Society* (1978), p. 33-50.

[101] Andreĭ Nikolaevich Kolmogorov et Albert T Bharucha-Reid. *Foundations of the theory of probability : Second English Edition*. Courier Dover Publications, 2018.

[102] Dimitri Konen et Davy Paindaveine. « Spatial quantiles on the hypersphere ». In : *The Annals of Statistics* 51.5 (2023), p. 2221-2245.

[103] Linglong Kong et Yijun Zuo. « Smooth depth contours characterize the underlying distribution ». In : *Journal of Multivariate Analysis* 101.9 (2010), p. 2222-2226. issn : 0047-259X.

[104] Stefan Kunis et Daniel Potts. « Fast spherical Fourier algorithms ». In : *Journal of Computational and Applied Mathematics* 161.1 (2003), p. 75-98. issn : 0377-0427.

[105] Yassine Laguel et al. « Superquantiles at work : Machine learning applications and efficient subgradient computation ». In : *Set-Valued and Variational Analysis* (2021).

[106] Rajmadan Lakshmanan, Alois Pichler et Daniel Potts. « Nonequispaced Fast Fourier Transform Boost for the Sinkhorn Algorithm ». In : *ETNA - Electronic Transactions on Numerical Analysis* 58 (jan. 2023), p. 289-315.

[107] Christian Léonard. « From the Schrödinger problem to the Monge–Kantorovich problem ». In : *Journal of Functional Analysis* 262.4 (2012), p. 1879-1920.

[108] Christophe Ley, Camille Sabbah et Thomas Verdebout. « A new concept of quantiles for directional data and the angular Mahalanobis depth ». In : *Electronic Journal of Statistics [electronic only]* 8 (jan. 2014).

[109] Christophe Ley et Thomas Verdebout. *Modern directional statistics*. CRC Press, 2017.

[110] Christophe Ley et Thomas Verdebout. *Applied directional statistics : modern methods and case studies*. CRC Press, 2018.

[111] Regina Y Liu. « On a notion of data depth based on random simplices ». In : *The Annals of Statistics* (1990), p. 405-414.

[112] Regina Y Liu, Jesse M Parelius et Kesar Singh. « Multivariate analysis by data depth : descriptive statistics, graphics and inference,(with discussion and a rejoinder by liu and singh) ». In : *The annals of statistics* 27.3 (1999), p. 783-858.

[113] Regina Y Liu et Kesar Singh. « Ordering directional data : concepts of data depth on circles and spheres ». In : *The Annals of Statistics* 20.3 (1992), p. 1468-1484.

[114] Grégoire Loeper. « Regularity of optimal maps on the sphere : The quadratic cost and the reflector antenna ». In : *Archive for rational mechanics and analysis* 199 (2011), p. 269-289.

[115] Grégoire Loeper et Cédric Villani. « Regularity of optimal transport in curved geometry : The nonfocal case ». In : *Duke Mathematical Journal* 151.3 (2010), p. 431 -485.

[116] Tudor Manole et al. *Plugin Estimation of Smooth Optimal Transport Maps*. 2021.

[117] Giulia Marcon, Philippe Naveau et Simone Padoan. « A semi-parametric stochastic generator for bivariate extreme events ». In : *Stat* 6.1 (2017), p. 184-201.

[118] Kanti V Mardia, Peter E Jupp et KV Mardia. *Directional statistics*. T. 2. Wiley Online Library, 2000.

[119] D Marinucci et al. « Spherical needlets for cosmic microwave background data analysis ». In : *Monthly Notices of the Royal Astronomical Society* 383.2 (2008), p. 539-545.

[120] Shoaib Bin Masud et al. « Multivariate soft rank via entropic optimal transport : sample efficiency and generative modeling ». In : *Journal of Machine Learning Research* 24.160 (2023), p. 1-65.

[121] Robert J. McCann. « Existence and uniqueness of monotone measure-preserving maps ». In : *Duke Mathematical Journal* 80.2 (1995), p. 309 -323.

[122] Robert J McCann. « Polar factorization of maps on Riemannian manifolds ». In : *Geometric & Functional Analysis GAFA* 11.3 (2001), p. 589-608.

[123] Gonzalo Mena et Jonathan Niles-Weed. « Statistical bounds for entropic optimal transport : sample complexity and the central limit theorem ». In : *Advances in Neural Information Processing Systems*. Sous la dir. de H. Wallach et al. T. 32. Curran Associates, Inc., 2019.

[124] Arthur Mensch et Gabriel Peyré. « Online sinkhorn : Optimal transport distances from sample streams ». In : *Advances in Neural Information Processing Systems* 33 (2020), p. 1657-1667.

[125] Volker Michel. *Lectures on constructive approximation : Fourier, spline, and wavelet methods on the real line, the sphere, and the ball*. Springer Science & Business Media, 2012.

[126] Nina Miolane et al. « Geomstats : A Python Package for Riemannian Geometry in Machine Learning ». In : *Journal of Machine Learning Research* 21.223 (2020), p. 1-9.

[127] Karl Mosler. « Depth statistics ». In : *Robustness and complex data structures : Festschrift in Honour of Ursula Gather* (2013), p. 17-34.

[128] Stanislav Nagy. « Halfspace depth does not characterize probability distributions ». In : *Statistical Papers* 62.3 (2021), p. 1135-1139.

[129] Stanislav Nagy et Petra Laketa. *Theoretical properties of angular halfspace depth*. 2024.

[130] Stanislav Nagy et al. « Theory of angular depth for classification of directional data ». In : *Advances in Data Analysis and Classification* (2023).

[131] Greg T von Nessi. « On the regularity of optimal transportation potentials on round spheres ». In : *Acta applicandae mathematicae* 123 (2013), p. 239-259.

[132] Ziang Niu et Bhaswar B. Bhattacharya. *Distribution-free joint independence testing and robust independent component analysis using optimal transport*. arXiv. 2022.

[133] Marcel Nutz et Johannes Wiesel. « Entropic optimal transport : convergence of potentials ». In : *Probability Theory and Related Fields* 184.1 (2022), p. 401-424.

[134] Giuseppe Pandolfo et Antonio D'ambrosio. « Clustering directional data through depth functions ». In : *Computational Statistics* 38.3 (2023), p. 1487-1506.

[135] Giuseppe Pandolfo, Davy Paindaveine et Giovanni C Porzio. « Distance-based depths for directional data ». In : *Canadian Journal of Statistics* 46.4 (2018), p. 593-609.

[136] Marco Pegoraro et al. « Vector Quantile Regression on Manifolds ». In : *ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems*. 2023.

[137] Arthur Pewsey et Eduardo García-Portugués. « Recent advances in directional statistics ». In : *Test* 30.1 (2021), p. 1-58.

[138] Gabriel Peyré, Marco Cuturi et al. « Computational optimal transport : With applications to data science ». In : *Foundations and Trends® in Machine Learning* 11.5-6 (2019), p. 355-607.

[139] Aram-Alexandre Pooladian, Vincent Divol et Jonathan Niles-Weed. « Minimax estimation of discontinuous optimal transport maps : The semi-discrete case ». In : *International Conference on Machine Learning*. PMLR. 2023, p. 28128-28150.

[140] Aram-Alexandre Pooladian et Jonathan Niles-Weed. *Entropic estimation of optimal transport maps*. arXiv :2109.12004. 2021. eprint : 2109.12004 (math.ST).

[141] Daniel Potts et Michael Schmischke. « Approximation of high-dimensional periodic functions with Fourier-based methods ». In : *SIAM Journal on Numerical Analysis* 59.5 (2021), p. 2393-2429.

[142] A Prékopa. « Multivariate value at risk and related topics ». In : *Annals of Operations Research* 193(1) (2012), p. 49-69.

[143] Philippe Rigollet et Austin J. Stromme. *On the sample complexity of entropic optimal transport*. arXiv. 2022.

[144] Herbert Robbins et Sutton Monro. « A Stochastic Approximation Method ». In : *The Annals of Mathematical Statistics* 22.3 (1951), p. 400-407.

[145] R. T. Rockafellar et J. O. Royset. « Superquantiles and their applications to risk, random variables and regression ». In : *Theory Driven by Influential Applications*. Informs, 2013, p. 151-167.

[146] R. Tyrrell Rockafellar. *Convex analysis*. Princeton University Press, 1970.

[147] R Tyrrell Rockafellar. « Favorable classes of Lipschitz continuous functions in subgradient optimization ». In : *Progress in nondifferentiable optimization*. Sous la dir. d'E. Nurminski. IIASA, 1981.

[148] R Tyrrell Rockafellar et Johannes O Royset. « Random variables, monotone relations, and convex analysis ». In : *Mathematical Programming* 148 (2014), p. 297-331.

[149] Peter J Rousseeuw et Anja Struyf. « Characterizing angular symmetry and regression symmetry ». In : *Journal of Statistical Planning and Inference* 122.1-2 (2004), p. 161-173.

[150] Ritwik Sadhu, Ziv Goldfeld et Kengo Kato. *Stability and statistical inference for semidiscrete optimal transport maps*. 2023.

[151] F. Santambrogio. *Optimal Transport for Applied Mathematicians : Calculus of Variations, PDEs, and Modeling*. Progress in Nonlinear Differential Equations and Their Applications. Springer International Publishing, 2015. isbn : 9783319208282.

[152] Gideon Schechtman et Joel Zinn. « On the volume of the intersection of two L_p^n balls ». In : *Proceedings of the American Mathematical Society* 110.1 (1990), p. 217-224.

[153] Vivien Seguy et al. « Large-Scale Optimal Transport and Mapping Estimation ». In : *ICLR 2018 - International Conference on Learning Representations*. 2018, p. 1-15.

[154] Tomonari Sei. « Gradient modeling for multivariate quantitative data ». In : *Annals of the Institute of Statistical Mathematics* 63 (2011), p. 675-688.

[155] Tomonari Sei. « A Jacobian inequality for gradient maps on the sphere and its application to directional statistics ». In : *Communications in Statistics-Theory and Methods* 42.14 (2013), p. 2525-2542.

[156] Robert Serfling. « Quantile functions for multivariate analysis : approaches and applications ». In : *Statistica Neerlandica* 56.2 (2002), p. 214-232.

[157] Robert Serfling. « Depth functions in nonparametric multivariate inference ». In : *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* 72 (2006), p. 1.

[158] Robert Serfling. *Approximation theorems of mathematical statistics*. John Wiley & Sons, 2009.

[159] Hongjian Shi, Mathias Drton et Fang Han. « Distribution-free consistent independence tests via center-outward ranks and signs ». In : *Journal of the American Statistical Association* 117.537 (2022), p. 395-410.

[160] Hongjian Shi et al. *Distribution-free tests of multivariate independence based on center-outward quadrant, Spearman, Kendall, and van der Waerden statistics.* arXiv :2111.15567v4. 2021.

[161] Hongjian Shi et al. « On universally consistent and fully distribution-free rank tests of vector independence ». In : *The Annals of Statistics* 50.4 (2022), p. 1933-1959.

[162] Christopher G Small. « Measures of centrality for multivariate and directional distributions ». In : *Canadian Journal of Statistics* 15.1 (1987), p. 31-39.

[163] Stefan Sommer, Tom Fletcher et Xavier Pennec. « Introduction to differential and Riemannian geometry ». In : *Riemannian Geometric Statistics in Medical Image Analysis*. Elsevier, 2020, p. 3-37.

[164] Elias M. Stein et Guido Weiss. « VII. Multiple Fourier Series ». In : *Introduction to Fourier Analysis on Euclidean Spaces (PMS-32), Volume 32*. Princeton University Press, 2016, p. 245-286.

[165] Austin Stromme. « Sampling from a Schrödinger bridge ». In : *International Conference on Artificial Intelligence and Statistics*. PMLR. 2023, p. 4058-4067.

[166] « Subdifferential Calculus Rules for Possibly Nonconvex Integral Functions ». In : *SIAM Journal on Control and Optimization* 58.1 (2020), p. 462-484.

[167] R Torres, R. E. Lillo et H. Laniado. « A directional multivariate value at risk ». In : *Insurance : Mathematics and Economics* 65 (2015), p. 111-123.

[168] Ivana Tošić et Pascal Frossard. « Dictionary learning ». In : *IEEE Signal Processing Magazine* 28.2 (2011), p. 27-38.

[169] J. W. Tukey. « Mathematics and the Picturing of Data ». In : *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974)* 2 (1975), p. 523-531.

[170] Kerem Uğurlu. « A new coherent multivariate average-value-at-risk ». In : *Optimization* 72.2 (2023), p. 493-519.

[171] Cees Fouad de Valk et Johan Segers. *Stability and tail limits of transport-based quantile contours*. LIDAM Discussion Papers ISBA 2018031. Université catholique de Louvain, Institute of Statistics, Biostatistics et Actuarial Sciences (ISBA), 2018.

[172] C. Villani. *Topics in optimal transportation*. T. 58. Graduate Studies in Mathematics. American Mathematical Society, 2003.

[173] Cédric Villani. *Optimal transport : old and new*. eng. T. 338. Berlin : Springer, 2009. isbn : 978-3-540-71049-3.

[174] Wei Wang et al. « A Linear Optimal Transportation Framework for Quantifying and Visualizing Variations in Sets of Images ». In : *International Journal of Computer Vision* 101.2 (2013), p. 254-269.

[175] Mark A Wieczorek et Matthias Meschede. « SHTools : Tools for working with spherical harmonics ». In : *Geochemistry, Geophysics, Geosystems* 19.8 (2018), p. 2574-2592.

[176] Yijun Zuo et Robert Serfling. « General notions of statistical depth function ». In : *The Annals of Statistics* (2000), p. 461-482.

[177] Yijun Zuo et Robert Serfling. « Structural properties and convergence results for contours of sample statistical depth functions ». In : *Annals of Statistics* (2000), p. 483-499.