Universidad de Valladolid

Escuela de Doctorado Universidad de Valladolid

PROGRAMA DE DOCTORADO EN MATEMÁTICAS

TESIS DOCTORAL:

STATISTICAL ANALYSIS OF THE OPTIMAL TRANSPORT PROBLEM

Presentada por Alberto González Sanz para optar al grado de Doctor/a por la Universidad de Valladolid

Dirigida por:
Dr. Eustasio del Barrio
Dr. Jean-Michel Loubes

Universidad de Valladolid

esdUVa
Escuela de Doctorado Universidad de Valladolid

Abstract

Optimal transportation is a resource allocation problem present in fields such as economics, finance, physics or artificial intelligence. From a probabilistic point of view, the optimal transport cost endows the space of probability measures with a metric topology. In particular, this topology is equivalent to the weak topology of probability measures together with the convergence of moments. This makes the transport cost an appropriate tool for measuring discrepancies between distributions. On the other hand, the solution of the transport problem is known as optimal plan. That is, an unambiguous way to relate two distributions following an optimality criterion. This optimal plan, when deterministic, is called a transport map.

However, in many cases the probability distribution is a theoretical, unattainable entity. It is only visible to the practitioner through its empirical version, i.e. a finite data set of size n. This work examines the asymptotic behaviour of the transport cost in its empirical version. In other words, we study the limits of the empirical cost and plans when the data grows to infinity. It is well-known that the empirical transport cost converges to the population one. Moreover, for continuous measures it does so at a rate that decreases with dimension. In this thesis we prove the consistency of the transport map using topology of set-valued maps. This leads, indirectly, to being able to state that the rate at which the fluctuations–difference between the expected empirical cost and the empirical cost itself–approximate zero is the parametric $n^{-\frac{1}{2}}$, irrespective of the dimension. Moreover, these fluctuations multiplied by $n^{\frac{1}{2}}$ tend toward a Gaussian random variable. In economics the transportation problem appears in numerous occasions in its semi-discrete version, i.e. one of the probability distributions is discrete. In this case, we show that the rate at which the empirical transport cost converges to the population one does not depend on the dimension.

We also show that the well-known entropy regularization (or Sinkhorn regularization), apart from simplifying the computation of the transport problem by giving it a differentiable structure, has highly satisfactory statistical properties. In particular, its bias and the divergence–that the regularization defines–converge with speed greater than the parametric one; the empirical regularized plans converge to the population ones with rate $n^{-\frac{1}{2}}$ and, moreover, tending to a Gaussian process.

The transport map endows a probability measure P with an order with respect to a given reference. This property leads to the successful definition of M.Hallin's multivariate distribution function by choosing as a reference measure the spherical uniform. This thesis provides sufficient conditions under which this function defines a homeomorphism between the support of the probability measure P and the unitary ball–i.e. to support of the spherical uniform. Finally, we provide a conditional version of the multivariate distribution function, with applications to quantile regression.

Résumé

Le transport optimal est un problème d'allocation de ressources que l'on retrouve dans des domaines tels que l'économie, la finance, la physique et l'intelligence artificielle. D'un point de vue probabiliste, le coût de transport optimal dote l'espace des mesures de probabilité d'une topologie métrique. Cela fait du coût de transport un outil approprié pour mesurer les écarts entre les distributions. D'autre part, la solution du problème de transport est connue comme le plan optimal. C'est-à-dire une manière non ambiguë de mettre en relation deux distributions suivant un critère d'optimalité. Ce plan optimal, lorsqu'il est déterministe, s'appelle une application de transport.

Cependant, la distribution de probabilité est souvent une entité théorique, irréalisable. Elle n'est visible pour le praticien qu'à travers sa version empirique, c'est-à-dire un ensemble de données fini de taille n. Ce document examine le comportement asymptotique du coût de transport dans sa version empirique. En d'autres termes, nous étudions les limites du coût empirique et de le plan lorsque les données croissent à l'infini. Les travaux précédents ont montré que le coût de transport empirique converge vers le coût théorique. De plus, pour les mesures continues, elle le fait à un taux qui diminue avec la dimension. Dans cette thèse, nous démontrons la cohérence de l'application de transport en utilisant la topologie des applications qui prennent des valeurs dans un espace d'ensembles. Cela conduit, indirectement, à pouvoir affirmer que le taux auquel les fluctuations–différence entre l'espérance du coût empirique et le coût empirique lui-même–se rapprochent de zéro est le paramètre $n^{-\frac{1}{2}}$. De plus, ces fluctuations multipliées par $n^{\frac{1}{2}}$ tendent vers une variable gaussienne. Dans les applications économiques, le problème du transport apparaît à de nombreuses reprises dans sa version semi-discrète, c'est-à-dire qu'une des distributions est discrète. Dans ce cas, nous montrons que la vitesse à laquelle le coût de transport empirique converge vers le coût de population ne dépend pas de la dimension.

Nous montrons également que la régularisation entropique (ou régularisation de Sinkhorn), outre qu'elle simplifie le calcul du problème de transport en lui donnant une structure différentiable, possède des propriétés statistiques très satisfaisantes. En particulier, leur biais et la divergence que la régularisation définit convergent avec une vitesse supérieure à celle du paramétrique ; les plans régularisés empiriques convergent vers ceux de la population, avec une erreur gaussienne et décroissante à la vitesse $n^{-\frac{1}{2}}$.

L'application du transport confère à une mesure de probabilité P un ordre par rapport à une référence donnée. Cette propriété permet de définir avec succès la fonction de répartition multivariée de M.Hallin en choisissant comme mesure de référence l'uniforme sphérique. Cette thèse fournit des conditions suffisantes pour lesquelles cette fonction définit un homéomorphisme entre le support de la mesure de probabilité P et la balle unitaire, c'est-à-dire le support de l'uniforme sphérique. Enfin, nous fournissons une version conditionnelle de la fonction de répartition multivariée, avec des applications à la régression quantile.

Resumen

El transporte óptimo es un problema de asignación de recursos presente en ámbitos como economía, finanzas, física o inteligencia artificial. Desde un punto de vista probabilístico, el coste de transporte óptimo dota al espacio de medidas de probabilidad de una topología métrica. En particular, esta topología es equivalente a la topología débil de medidas junto con la convergencia de los momentos. Esto hace el coste del transporte una herramienta apropiada para la medición de discrepancias entre distribuciones. Por otro lado, la solución del problema de transporte es conocido como plan óptimo. Es decir, una manera inequívoca de relacionar dos distribuciones siguiendo un criterio de optimalidad. Este plan óptimo, cuando es determinista, es llamado aplicación de transporte.

Sin embargo, en muchas ocasiones la distribution de probabilidad es un ente teórico, inalcanzable. Solo es visible para el practicante a través de su versión empírica, es decir, de un conjunto de datos de tamaño finito n. Este trabajo examina el comportamiento asintótico del coste de transporte en su versión empírica. En otras palabras, se estudian los límites del coste y planes de transporte empíricos cuando los datos tienden a infinito. Es conocido, en varios trabajos precedentes, que el coste de transporte empírico converge hacia el poblacional. Es más, para medidas continuas lo hace a una velocidad que decrece con la dimensión. En esta tesis se demuestra la consistencia de la aplicación de transporte utilizando topología de aplicaciones que toman valores en un espacio de conjuntos. Esto lleva, de manera indirecta, a poder afirmar que la velocidad a las fluctuaciones –diferencia entre esperanza empírica del coste y el propio coste empírico– se aproxima a cero es la paramétrica $n^{-\frac{1}{2}}$. Además, estas fluctuaciones multiplicadas por $n^{\frac{1}{2}}$ tienden hacia una variable gausiana. En aplicaciones económicas el problema de transporte aparece en numerosas ocasiones en su versión semidiscreta, i.e. una de las distribuciones de probabilidad es discreta. En este caso, mostramos que la velocidad a la que el coste de transporte empírico converge hacia el poblacional no depende de la dimensión.

Demostramos también que la conocida regularización por la entropía (o regularización de Sinkhorn), aparte de simplificar la computación del problema de transporte dotándole de una estructura diferenciable, tiene propiedades estadísticas altamente satisfactorias. En particular, su sesgo y la divergencia que la regularización define convergen con velocidad mayor a la paramétrica; los planes regularizados empíricos convergen hacia los poblacionales, con un error gausiano y decreciente a velocidad $n^{-\frac{1}{2}}$.

La transformación de transporte otorga a una medida de probabilidad P un orden con respecto a una referencia dada. Esta propiedad permite la exitosa definición de la función de distribución multivariada de M.Hallin eligiendo como medida de referencia la uniforme esférica. Esta tesis proporciona condiciones suficientes bajo las cuales esta función define un homeomorfismo entre el soporte de la medida de probabilidad P y la bola unitaria, es decir, el soporte de la uniforme esférica. Finalmente, proporcionamos una versión condicional de la función de distribución multivariada, con aplicaciones a la regresión cuantílica.

Contents

CHAPTER 1

Introduction (English version)

Optimal transport is a resource allocation problem present in multiple areas of mathematics, and thus in its applications. This versatility is also manifested in its own theoretical framework: the study of its regularity relies on advanced techniques of differential equations (Caffarelli, 1990, 1991, 1992); the development of efficient computational methods, numerical analysis and combinatorics (Peyré and Cuturi, 2019, Chapter 3); its asymptotic behavior, convex analysis (del Barrio and Loubes, 2019), and empirical processes (del Barrio et al., 2005).

Optimal transportation consists of finding, among all probability measures with the same fixed marginals, the one that minimizes the average transportation cost. This minimum average value is known as the optimal transportation cost. In cases where one of the probabilities has density, the solution is deterministic and is given by a map (Gangbo and McCann, 1996), usually known as the transport map. The optimal transportation cost provides a metric structure, by means of the so-called Wasserstein distance, to the space of probability measures. Therefore, from a statistical point of view, optimal transportation offers a tool for data comparison that takes into account the geometry of the latent space, which has proven to be effective in solving problems such as bias correction in machine learning (Risser et al., 2021; Gordaliza et al., 2019a; Black et al., 2020), in modeling contrafactual reasoning (de Lara et al., 2021; Black et al., 2020) or in diffeomorphic registration (Feydy et al., 2017; De Lara et al., 2022+).

A natural application of any significant distance between distributions is the goodness-of-fit problem, that is, the problem of testing the null hypothesis that a sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ comes from a population with a completely specified distribution P (Hallin et al., 2021b). Effectively, in moderate dimensions, the Wasserstein distance can provide a consistent statistical test against any fixed alternative, eg. González-Delgado et al. (2021) proposes a consistent test based on the Wasserstein distance in the 2-dimensional torus. These applications, where the Wasserstein distance quantifies the similarity between different data samples, require a rigorous mathematical justification.

The first part of this thesis focuses on the asymptotic-statistical study of the transportation problem. We will see that the dimension of the latent space affects the error of the empirical approximation of the optimal transportation cost but not its variance, in fact, the weak limit of the fluctuations (difference between the empirical optimal transport cost distance and its mean) multiplied by $n^{\frac{1}{2}}$ follows a Gaussian distribution. The influence of dimension on the convergence rate of the empirical to population version is known as the "curse of dimensionality". The first rigorous proof of this fact date back to 1969 with the seminal paper of Dudley (1969). Recently, the works of Fournier and Guillin (2013) and Weed and Bach (2019) confirm the fact that the empirical Wasserstein distance between two continuous distributions converges to the population with rate $n^{-\frac{1}{d}}$, except for possibly logarithmic factors. Despite the fact that this bound can be improved for sufficiently separated probability distributions (see Manole and Niles-Weed (2021)), the convergence rate continues to depend exponentially on the dimension. Therefore, the bounds of the p-values of any statistical test based on the Wasserstein distance depend exponentially on the dimension, requiring a massive amount of data to obtain significant results.

The fluctuations have a different asymptotic behavior. Furthermore, they can be bounded with rate $n^{-\frac{1}{2}}$, irrespective of the dimension (Weed and Bach, 2019, proposition 20). The arguments presented in Weed and Bach (2019), which are based on the McDiarmid inequality, are not applicable to probability measures with unbounded supports. A more precise study of the fluctuations is the one performed by del Barrio and Loubes (2019) through the Efron-Stein inequality. However, most of the arguments are specific to the quadratic cost –in which the transport maps are gradients of convex functions– making their generalization to other costs not trivial.

The aforementioned curse of dimensionality appears for probabilities with density. When the two probabilities are discrete, the optimization problem becomes parametric and satisfies a central limit theorem with a rate $n^{-\frac{1}{2}}$, see Sommerfeld and Munk (2018). If one of the probabilities is discrete –the so-called semi-discrete problem–, del Barrio and Loubes (2019) proved the central limit theorem centered on the population value as an application of the result obtained for the fluctuations. This strategy, as we will see in this work, is not the most appropriate. The functional version of the delta-method provides a methodology that requires fewer hypotheses. This has been observed in parallel by Hundrieser et al. (2022).

The control of fluctuations is more useful for analyzing the entropy-regularized transport problem. Proposed by Cuturi (2013), it is undoubtedly the most widely used method for regularizing the transport problem. Mena and Niles-Weed (2019) proved, using the arguments of del Barrio and Loubes (2019), that the fluctuations of the regularized problem are asymptotically Gaussian. In this work we will see that in addition the bias converges faster than the variance, resulting in the central limit theorem for the entropy-regularized transport cost. The regularized transport cost cannot be used to perform a goodness-of-fit test. The added term as a penalty causes a phenomenon known as entropic bias, which substantially reduces the usefulness of the regularized transport cost for statistical inference. Feydy et al. (2019) proposes a modification of the regularized transport cost –the Sinkhorn divergence– that repairs this problem. This thesis provides a second-order development of the Sinkhorn divergence with respect to the empirical process. As a result, a precise characterization of its asymptotic behavior is obtained. This can be potentially used to derive a goodness-of fit-test statistic based on the Sinkhorn divergence.

It is known that ranks, based on the univariate order notion, provide a general methodology for addressing the goodness-of-fit testing problem. However, the multivariate generalization of the concept of ranks has not been possible due to the absence of a notion of multivariate distribution function. Recently, the works of Marc Hallin and co-authors (eg. Hallin et al. (2021a)) propose the application of transport between the sample of data and the spherical uniform distribution as a candidate for a multivariate distribution function. It turns out that –always according to Hallin et al. (2021a)– this proposal satisfies the main properties that make the univariate distribution function a useful tool for statistical inference. This has provided, as mentioned for the univariate case, a general methodology for creating goodness-of-fit tests (Deb and Sen, 2019; Deb et al., 2021) or independence tests (Shi et al., 2022; Hallin and Mordant, 2021). The second part of this thesis proves the continuity of the multivariate distribution function for probabilities supported on a convex set, thus extending the result of Figalli (2018). The singularity at the origin of coordinates of the uniform spherical probability distribution means that the results of Caffarelli (see eg. Caffarelli (1990, 1991, 1992)) do not apply in this case. To conclude, we will provide, using the new concept of multivariate distribution function, an innovative methodology to solve the multivariate output non-parametric regression problem. We will see that this proposal is, to date, the only one that maintains the fundamental property of quantile regression, the probabilistic control of quantile regions. This property, in the univariate case, dictates that the quantile region of order $0 < r < 1$ contains a proportion of r points of the sample.

The two parts that compose this work are clearly delimited. In my opinion, the link is to be found in the mathematical aspect, in the reasoning and the tools used and not in the object of study itself. Before presenting in detail the results obtained, with the technical definitions and the tools used, it is worth underlining the fact that each chapter is sefl-contained and adapted from its on-line (or published) version. In other words, each of them can be read and analyzed separately. Those who are not interested in the links between the different chapters can skip the rest of the introduction.

1.0.1 Optimal transport in a nutshell

Since the end of the last century, the Monge assignment problem has become an important research topic in statistics and probability, with applications to machine learning, economics, physics and astronomy, to name only a few. Let $c : \mathbb{R}^d \times \mathbb{R}^d \to [0, +\infty)$ be a continuous function, the *optimal transport cost* between two probability measures $P, Q \in \mathcal{P}(\mathbb{R}^d)$ for the cost c is defined as the solution of the Monge's problem

$$\inf_{T:\, T_{\#}P=Q} \int_{\mathbb{R}^d} c(\mathbf{x}, T(\mathbf{x}))dP(\mathbf{x}), \tag{1.1}$$

where the notation $T_{\#}P$ represents the *push-forward* measure, i.e., the measure such that $T_{\#}P(A) := P(T^{-1}(A))$, for each measurable set A. It took until the 1990s, with the parallel work of Brenier (1991) and Cuesta and Matrán (1989), to prove the existence of the solution for the quadratic cost (($c(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2$), see also Gangbo and McCann (1996) for more general costs or Villani (2003) as a general resource for further understanding of the optimal transport theory.

Meanwhile, Kantorovich in 1942 (see Kantorovich (2006) for an English translation of the original article) formulated the famous relaxation of Monge's problem;

$$\mathcal{T}_c(P,Q) = \inf_{\pi \in \Pi(P,Q)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(\mathbf{x},\mathbf{y}) d\pi(\mathbf{x},\mathbf{y}), \tag{1.2}$$

where $\Pi(P,Q)$ is the set of probability measures $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ such that $\pi(A \times \mathbb{R}^d) = P(A)$ and $\pi(\mathbb{R}^d \times B) = Q(B)$, for all measurable sets A, B. A probability measure $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is said to be a *optimal transportation plan for cost* c between P and Q if it is a minimizer in (1.2).

The main advantage of the Kantorovich relaxation is the existence of optimal solutions: for (1.2) it only requires integrability of the cost (Theorem 4.1 in Villani (2008)) while the Monge problem needs some hypotheses on one of both probabilities–absolutely continuity with respect to the Lebesgue measure–and on the cost–the well-known Gangbo and McCann (1996)'s conditions for the existence of optimal transport maps, i.e. $c(\mathbf{x},\mathbf{y}) = h(\mathbf{x}-\mathbf{y})$, where $h : \mathbb{R}^d \to [0,\infty)$ is a non negative function satisfying

(A1) h is strictly convex on $\mathbb{R}^d$,

(A2) given a height $r \in \mathbb{R}^+$ and an angle $\theta \in (0,\pi)$, there exists some $M := M(r,\theta) > 0$ such that for all $|\mathbf{p}| > M$, one can find a cone

$$K(r,\theta,\mathbf{z},\mathbf{p}) := \left\{ \mathbf{x} \in \mathbb{R}^d : |\mathbf{x}-\mathbf{p}||\mathbf{z}| \cos(\theta/2) \leq \langle \mathbf{z}, \mathbf{x}-\mathbf{p} \rangle \leq r|\mathbf{z}| \right\},$$

with vertex at $\mathbf{p}$ (and $\mathbf{z} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$) on which h attains its maximum at $\mathbf{p}$,

(A3) $\lim_{|\mathbf{x}|\to\infty} \frac{h(\mathbf{x})}{|\mathbf{x}|} = \infty$.

The transportation problem admits a dual formulation:

$$\mathcal{T}_c(P,Q) = \sup_{(f,g)\in\Phi_c(P,Q)} \int f(\mathbf{x}) dP(\mathbf{x}) + \int g(\mathbf{y}) dQ(\mathbf{y}), \tag{1.3}$$

where $\Phi_c(P,Q) = \{(f,g) \in L_1(P) \times L_1(Q) : \ f(\mathbf{x}) + g(\mathbf{y}) \leq c(\mathbf{x},\mathbf{y})\}$, see Theorem 5.10 inVillani (2008). Call $\psi \in L_1(P)$ *an optimal transport potential from* P *to* Q *for cost* c if there exists $\varphi \in L_1(Q)$ such that the couple (ψ,φ) solves (1.3). Surprisingly, the equivalence between (1.1) and (1.2) passes through the regularity of the potentials, which in turn follows from that of the cost (cf. Gangbo and McCann (1996) eg.). As far as the content treated in this thesis is concerned, optimal transport potentials will describe: in Chapters 2 and 3, the variance of the bounds of the fluctuations; and in Chapter 3 the prices given by a company to certain product for the *"Hotelling's location model"* (Galichon, 2016, Chapter 5.1).

Each formulation has a different interest, and its use depends on the application. On the one hand, when it exists, the solution of (1.1) defines a transport map between probability measures. This allows to infer the properties of one probability measure through another already known (or some reference measure). This is the idea behind the successful M. Hallin's multivariate quantile function (Hallin et al. (2021a)); of multivariate quantile regression (Chapter 7); of mass-transportation based counterfactual explanations (de Lara et al., 2021) or of bias repair (Gordaliza et al. (2019b); Black et al. (2020)).

Let $\mathcal{P}_p(\mathbb{R}^d)$ be the space of probabilities in $\mathbb{R}^d$ with finite moments of order $p \geq 1$. On the other hand, when a potential cost ($c_p(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^p$, for $p \geq 1$) is considered, the Kantorovich formulation between probabilities with finite moments of order $p \geq 1$ admits always a solution and the function

$$\mathcal{P}_p(\mathbb{R}^d) \times \mathcal{P}_p(\mathbb{R}^d) \ni (P, Q) \to \mathcal{W}_p(P, Q) = (\mathcal{T}_p(P, Q))^{\frac{1}{p}} = \left(\mathcal{T}_{c_p}(P, Q)\right)^{\frac{1}{p}}$$

defines a distance $\mathcal{P}_p(\mathbb{R}^d)$ characterized by:

$$\mathcal{W}_p(\mu_n, \mu) \to 0 \quad \iff \quad \mu_n \xrightarrow{w} \mu \quad \text{and} \quad \int \|\mathbf{x}\|^p d\mu_n(\mathbf{x}) \to \int \|\mathbf{x}\|^p d\mu(\mathbf{x}),$$

see (Villani, 2003, Chapter 7). This is the so-called *"Wasserstein distance"*. It is used –to name only a few–, as a discriminator in Generatives adversarial networks Arjovsky et al. (2017), in dipheomorphic registration or as a penalty term in algorithmic bias reparation (Risser et al., 2021).

1.1 Asymptotic behavior

The population probability P is not usually available to practitioners; what they observe is an i.i.d. sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ of size n of P, defining the empirical measure P_n. Let us suppose for this introduction that Q is known. However, the results here exposed still hold in the two-sample case. The value $\mathcal{T}_c(P_n, Q)$ is thus the empirical counterpart of the population $\mathcal{T}_c(P, Q)$. Of course, $\mathcal{T}_c(P_n, Q)$ tends to $\mathcal{T}_c(P, Q)$, but, at what rate? That is, if the number of data we have, n, tends towards infinity, how does the difference $\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q)$ vary? To give a quick answer we can make a bias-variance division of the error:

$$\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q) = (\mathbb{E}(\mathcal{T}_c(P_n, Q)) - \mathcal{T}_c(P, Q)) + (\mathcal{T}_c(P_n, Q) - \mathbb{E}(\mathcal{T}_c(P_n, Q))) . \tag{1.4}$$

Before formally analyzing the asymptotic behaviour of bias $(\mathbb{E}(\mathcal{T}_c(P_n, Q_n)) - \mathcal{T}_c(P, Q))$ and variance $(\mathcal{T}_c(P_n, Q) - \mathbb{E}(\mathcal{T}_c(P_n, Q)))$, a simulation may be useful. To do this we set a dimension "relatively high" (as is $d = 10$) and simulate $\sqrt{n}(\mathcal{T}_2(P_n, Q_n))$, for $P, Q \sim \mathcal{U}_{[0,1]^{10}}$. Figure 1.1 shows, with different colors and repeating 10 independent times the procedure, the different values of $\sqrt{n}(\mathcal{T}_2(P_n, Q_n))$ on the ordinate axis and the sample size n on the abscissa axis.

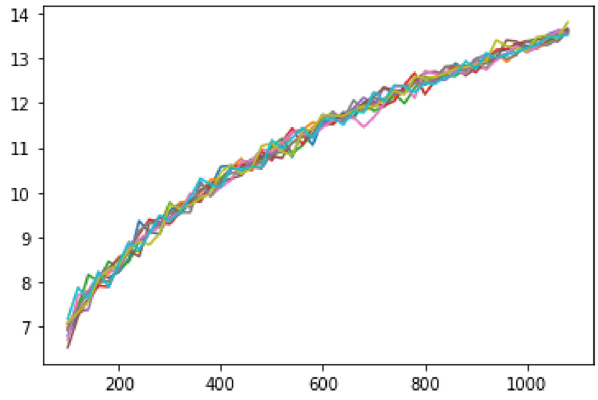

Figure 1.1: x's-axis, sample size n; y's-axis, difference $\sqrt{n}(\mathcal{T}_2(P_n, Q_n))$, for $P, Q \sim \mathcal{U}_{[0,1]^{10}}$. The experiment is repeated 10 independent times and the interpolated results plotted in different colours.

The heuristic observation turns out to be quite enlightening: the convergence rate of the bias cannot be the parametric rate $\sqrt{n}$, but $\sqrt{n}(\mathcal{T}_2(P_n, Q_n) - \mathbb{E}\mathcal{T}_2(P_n, Q_n))$ seems bounded. That is, apparently the fluctuations are not influenced by the fact of having considerably high dimension ($d = 10$). Indeed, for absolutely continuous measures with respect to the d-dimensional Lebesgue measure ℓ_d, the correct convergence rate of the bias depends on the dimension of the space (Weed and Bach, 2019).

1.1.1 Fluctuation analysis

For the fluctuation analysis (variance) we assume that $c(\mathbf{x}-\mathbf{y}) = h(\mathbf{x}-\mathbf{y})$ satisfies Gangbo and McCann (1996)'s conditions, P and Q are absolutely continuous wit respect to the Lebesgue measure and

$$\int h(2\mathbf{x})^2 dP(\mathbf{x}) < \infty \text{ and } \int h(-2\mathbf{y})^2 dQ(\mathbf{y}) < \infty.$$

Chapter 2 shows

$$\sqrt{n}\left(\mathcal{T}_c(P_n, Q) - \mathbb{E}\mathcal{T}_c(P_n, Q)\right) \xrightarrow{w} N(0, \sigma_c^2(P, Q)), \tag{1.5}$$

where

$$\sigma_c^2(P, Q) := \int \varphi(\mathbf{x})^2 dP(\mathbf{x}) - \left(\int \varphi(\mathbf{x}) dP(\mathbf{x})\right)^2, \tag{1.6}$$

being φ an optimal transport potential from P to Q for the cost c. In the case $P = Q$ the optimal transport potential is constant, so that its variance is null and the limit (1.5) is degenerate. Returning to the simulated example above (Figure 1.1), this implies that the fluctuations are not only controlled, but tend in probability to 0. The correct rate-of convergence –the one giving non-degenerated limits– of the fluctuations under $P = Q$ is still an open problem.

The proof is mainly based on Efron-Stein inequality (Boucheron et al., 2013, Chapter 3.1), i.e. if $(\mathbf{X}_1', \ldots, \mathbf{X}_n')$ and $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ are i.i.d, set $Z := f(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ and for each $i \in \{1, \ldots, n\}$ denote $Z_i' := f(\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}, \mathbf{X}_i', \mathbf{X}_{i+1}, \ldots, \mathbf{X}_n)$, then

$$\mathit{Var}(Z) \leq \frac{n}{2} E(Z - Z_i')^2 = nE(Z - Z_i')_+^2.$$

where $(\cdot)_+$ denotes the positive part. Efron-Stein inequality gives thus

$$n\mathit{Var}(\mathcal{T}_c(P_n, Q) - \int \varphi(\mathbf{x}) dP_n(\mathbf{x})) \leq \mathbb{E}\big(\varphi_n(\mathbf{X}_1) - \varphi(\mathbf{X}_1) - \varphi_n(\mathbf{X}_1') + \varphi(\mathbf{X}_1')\big)^2, \tag{1.7}$$

which will mean the tightness of $\sqrt{n}\,(\mathcal{T}_c(P_n, Q) - \mathbb{E}\mathcal{T}_c(P_n, Q))$. The central limit theorem will follow from the stability of the potentials –Corollary 2.2.7 gives the uniqueness (up to additive constant) of the potentials for probabilities with connected support and Theorem 2.3.4 the stability of the transport map. We extract here the content of Theorem 2.3.4 due to its importance–it is the first result showing the stability of the optimal transport map and potential in unbounded domains and for general costs.

Theorem 1.1.1. *Let $Q \in \mathcal{P}(\mathbb{R}^d)$ be such that $Q \ll \ell_d$ and has a connected support with negligible boundary. Assume $Q_n, P_n, P \in \mathcal{P}(\mathbb{R}^d)$ are such that $P_n \stackrel{w}{\to} P$, $Q_n \stackrel{w}{\to} Q$ and*

$$\mathcal{T}_c(P_n, Q_n) < \infty \ \text{ and } \ \mathcal{T}_c(P, Q) < \infty$$

for a cost $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$, with h differentiable and satisfying (A1)-(A3). *If ψ_n (resp. ψ) are optimal transport potentials from Q_n to P_n (resp. from Q to P) for the cost c. Then:*

(i) There exist constants $a_n \in \mathbb{R}$ such that $\tilde{\psi}_n := \psi_n - a_n \to \psi$ in the sense of uniform convergence on the compacts sets of $\mathrm{Supp}(Q)$.

(ii) For each compact $K \subset \mathrm{Supp}(Q) \cap \mathrm{dom}(\nabla\psi)$

$$\sup_{\mathbf{x} \in K} \sup_{\mathbf{y}_n \in \partial^c \psi_n(\mathbf{x})} |\mathbf{y}_n - \nabla^c \psi(\mathbf{x})| \longrightarrow 0.$$

However, this result only yields the a.s. convergence of the empirical potentials towards their population counterpart, but one cannot conclude from this that (1.7) tends to 0. To avoid further assumptions (as in del Barrio and Loubes (2019)), one can show that the sequence $\mathbb{E}\big|\varphi_n(\mathbf{X}_1) - \varphi(\mathbf{X}_1) - \varphi_n(\mathbf{X}_1') + \varphi(\mathbf{X}_1')\big|$ is bounded and then the Banach-Alaoglu theorem yields weak convergence in $L^2(\mathbb{P})$ of $\big|\varphi_n(\mathbf{X}_1) - \varphi(\mathbf{X}_1) - \varphi_n(\mathbf{X}_1') + \varphi(\mathbf{X}_1')\big|$ along sub-sequences. By taking Cesàro means we can go from weak to strong convergence.

1.2 Semi-discrete case

The behavior of the bias is very different in the the semi-discrete case–when one of the two probabilities is supported on a finite set of points $\mathbb{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$. Here, in Chapter 3 we obtain

$$\sqrt{n}\left(\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q)\right) \xrightarrow{w} \sup_{\mathbf{z}\in \mathrm{Opt}_c(P,Q)} \mathbb{G}(\mathbf{z}),$$

where $\mathrm{Opt}_c(P, Q)$ is the set of optimal transport potentials and $\mathbb{G}$ a Gaussian process acting on them. It is then evident that the uniqueness of the potentials (except for additive constants) implies Gaussian weak limits. Moreover, under this hypothesis and thanks to the work of Cárcamo et al. (2020), we know that the bootstrap is consistent as an approximation of the limit process. The fact that the curse of dimensionality seems to not affect the semi discrete case for both probabilities is quite astonishing. But it is partially hidden in the assumption that the support has a fixed finite size. For a better understanding we provide the upper bound

$$E\left|\mathcal{W}_1(P, Q_m) - \mathcal{W}_1(P, Q)\right| \leq \frac{8\sqrt{2N}}{\sqrt{m}} K(\mathrm{diam}(\mathbb{X}), Q),$$

where

$$K(\mathrm{diam}(\mathbb{X}), Q) \\ = \left(4\,\mathrm{diam}(\mathbb{X}) + 2\sqrt{\int |\mathbf{y}|^2 dQ(\mathbf{y})} + 2\,\mathrm{diam}(\mathbb{X})\right)\left(\log(2) + \sqrt{2\,\mathrm{diam}(\mathbb{X}) + 1}\right)$$

and $\mathcal{W}_1$ is the 1-Wasserstein distance for the Euclidean distance.

The proofs of Chapters 2 and 3 are based on completely different arguments; the limit of fluctuations uses the procedure introduced by del Barrio and Loubes (2019) based on Efron-Stein inequality, while the semi-discrete problem requires the functional derivation in the Hadamard sense of the transport cost. Lemma 3.1.1 shows that $\mathcal{T}_c(P, Q)$ is equivalent to the maximization (over a parametric class) of

$$g_c(P, Q, \mathbf{z}) = \sum_{i=1}^{N} z_i p_i + \int \inf_{i=1,\dots,N}\{c(\mathbf{x}_i, \mathbf{y}) - z_i\} dQ(\mathbf{y}). \tag{1.8}$$

This justifies the fact that the convergence rate is just the parametric one. The solution $\mathbf{z}^*$ (unique up to additive constant under some assumptions) of (1.8) defines the Laguerre cells

$$\mathrm{Lag}_k(\mathbf{z}^*) := \{\mathbf{y} \in \mathbb{R}^d : \ c(\mathbf{x}_k, \mathbf{y}) - z_k < c(\mathbf{x}_i, \mathbf{y}) - z_i^* \ \text{ for all } i \neq k\}, \quad k = 1, \dots, N,$$

which are generalizations of the Voronoi cells–equivalent to $\mathrm{Lag}_k(\mathbf{0})$ for the quadratic cost.

The semi-discrete optimal transport has applications to the *"Hotelling's location model"* (eg. (Galichon, 2016, Chapter 5.1)). This application is a typical example of socio-economic problem where the location of certain population is represented by a continuous probability Q and the fountains –businesses trying to sell a product– as a discrete probability P. Here the location of the fountain i is $\mathbf{x}_i$ and the capacity is p_i. Each inhabitant would choose the fountain at the same time closer and offering a better price, i.e. the strategy $\arg\inf_{i=1,\dots,N}\{c(\mathbf{x}_i,\mathbf{y}) - z_i\}$, where z_i represents the price of the fountain i. The set of population that prefers to consume from the fountain i is actually $\mathrm{Lag}_k(\mathbf{z})$. Under market clearing–supply equals to demand–each fountain is used to its full capacity and the problem of determining the prices reduces to a solution of a semi-discrete optimal transport cost.

Chapter 3 also gives, up to our knowledge, the first Central Limit Theorem for the solutions of the dual problem (1.8). We underline this result can not be generalized for continuous distributions. Indeed, if both probabilities are continuous and the space is not one dimensional, we cannot expect such type of central limit for the potentials, since, the expected value of the estimation of the transport cost converges with rate $O(n^{-\frac{1}{d}})$ and no longer $O(n^{-\frac{1}{2}})$. When the two samples are discrete, even if such a rate is $O(n^{-\frac{1}{2}})$, the lack of uniqueness of the dual problem does not allow to prove such type of problems. In consequence, the semi-discrete is one of the few cases where such results, for the optimal transport potentials in general dimension, can be expected. Chapter 3 provides weak limits for the Laguerre cells in terms of the L^p metric (cf. Vitale (1985) eg.) of between the empirical and population cells. Moreover, we give also asymptotic confidence neighbourhoods in terms of the Hausdorff distance (L^∞ metric). In all cases, the parametric rate is achieved.

1.3 Entropy Regularized Optimal Transport

Therefore, the problem of finding the weak limits of the transport cost is complicated even in moderate dimension, and, even if they were found, the rate would be slow. That is, too much data would be needed to approximate the limiting distribution. One needs, therefore, another notion of discrepancy between distributions that, while capturing the geometry of the space, has an empirical approximation that converges to its population counterpart with the parametric rate $\sqrt{n}$.

With a computational motivation, the influential article Cuturi (2013) proposes a regularization of (1.2) by an entropy penalty:

$$S_\epsilon(P,Q) = \min_{\pi\in\Pi(P,Q)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \tfrac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 d\pi(\mathbf{x},\mathbf{y}) + \epsilon H(\pi|P\times Q), \tag{1.9}$$

where H denotes the relative entropy, defined, for two probability measures α and β, as o $H(\alpha|\beta) = \int \log(\frac{d\alpha}{d\beta}(x))d\alpha(x)$ if α is absolutely continuous with respect to β and $+\infty$ otherwise. Let $\pi^\epsilon_{P,Q}$be the solution of (1.9), which is absolutely continuous with respect to $P \times Q$ with density $\xi^{\epsilon}_{P,Q}$. This problem, as well as its unregularized counterpart (1.2), admits a dual formulation.

$$S_\epsilon(P,Q) = \sup_{f\in L_1(P), g\in L_1(Q)} \Big\{ \int_{\mathbb{R}^d} f(\mathbf{x})dP(\mathbf{x}) + \int_{\mathbb{R}^d} g(\mathbf{y})dQ(\mathbf{y}) \\ - \epsilon \int_{\mathbb{R}^d\times\mathbb{R}^d} e^{\frac{f(\mathbf{x})+g(\mathbf{y})-\frac{1}{2}\|x-y\|^2}{\epsilon}} dP(\mathbf{x})dQ(\mathbf{y}) + \epsilon \Big\} \quad (1.10)$$

The pair of solutions of (1.10), call it $(f^{\epsilon}_{P,Q}, g^{\epsilon}_{P,Q})$, satisfies the following *optimality conditions*:

$$\int e^{\frac{f_{P,Q}(x)+g_{P,Q}(y)-\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2}{\epsilon}} dQ(\mathbf{y}) = 1, \quad \text{for all } \mathbf{x} \in \mathbb{R}^d,$$
$$\int e^{\frac{f_{P,Q}(\mathbf{x})+g_{P,Q}(\mathbf{y})-\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2}{\epsilon}} dP(\mathbf{x}) = 1, \quad \text{for all } \mathbf{y} \in \mathbb{R}^d. \quad (1.11)$$

This makes it possible to find solutions for (1.10) on the class $\mathcal{C}^s$, reduce the complexity of the optimization (cf. Genevay et al. (2019)) and *a fortiori* $\sqrt{n}\,\mathbb{E}|S_\epsilon(P_n,Q)-S_\epsilon(P,Q)| \le C$, valid for sub-Gaussian probabilities Mena and Niles-Weed (2019).

Moreover, del Barrio and Loubes (2019)'s concentration-argument holds also in this case (see Mena and Niles-Weed (2019)), giving rise to

$$\sqrt{n}(S_\epsilon(P_n,Q) - \mathbb{E}S_\epsilon(P_n,Q)) \xrightarrow{w} N(0, \mathrm{Var}_{\mathbf{X}\sim P}(f^{\epsilon}_{P,Q}(\mathbf{X}))). \quad (1.12)$$

Both (1.12) and the previously mentioned bound did not provide asymptotically valid confidence intervals one the population value $S_\epsilon(P,Q)$). For this purpose, it is necessary to know the exact limit of

$$\sqrt{n}\,(\mathbb{E}S_\epsilon(P_n,Q) - S_\epsilon(P,Q)). \quad (1.13)$$

In Chapter 4 it is proved that (1.13) tends to 0, so the expected value can be exchanged for the population value in (1.12), resulting in

$$\sqrt{n}(S_\epsilon(P_n,Q) - S_\epsilon(P,Q)) \xrightarrow{w} N(0, \mathrm{Var}_{\mathbf{X}\sim P}(f^{\epsilon}_{P,Q}(\mathbf{X}))). \quad (1.14)$$

Let z_α be the α-quantile for the standard normal distribution and

$$\hat{\sigma}^2_n := \mathrm{Var}_{P_n}(f^{\epsilon}_{P_n,Q}) = \frac{1}{n}\sum_{i=1}^{n}(f^{\epsilon}_{P_n,Q}(\mathbf{X}_i))^2 - \Big(\frac{1}{n}\sum_{i=1}^{n} f^{\epsilon}_{P_n,Q}(\mathbf{X}_i)\Big)^2 \quad (1.15)$$

be a consistent approximation of the limit variance $\mathrm{Var}_{\mathbf{X}\sim P}(f^{\epsilon}_{P,Q}(\mathbf{X}))$, the set

$$\left[S_\epsilon(P_n,Q) \pm \frac{\hat{\sigma}_n}{\sqrt{n}} z_{1-\alpha/2}\right], \quad (1.16)$$

is thus an asymptotic confidence intervals of level alpha for the value $S_\epsilon(P_n,Q)$.

Nevertheless, (1.16) does not provide any statistically significant consequences due to the entropic bias, the fact that $S_\epsilon(P,Q)=0$ does not mean that both probabilities are equal! The hypothesis $H_0:\ P=Q$ cannot be accepted or rejected via the value of $S_\epsilon(P,Q)$.

The most successful solution to curtail the influence of the entropy regularization is *Sinkhorn's divergence*, proposed by Feydy et al. (2019) and stated as

$$D_\epsilon(P,Q)=S_\epsilon(P,Q)-\tfrac{1}{2}\left(S_\epsilon(P,P)+S_\epsilon(Q,Q)\right).$$

Clearly $D_\epsilon(P,Q)$ is symmetric in P,Q and $SD_\epsilon(P,P)=0$. Moreover, (Feydy et al., 2019, Theorem 1) proves $D_\epsilon(P,Q)\geq 0$, with $D_\epsilon(P,Q)=0$ if and only if $P=Q$.

Asymptotic confidence intervals for $SD_\epsilon(P,Q)$ can be inferred from their empirical counterpoint by knowing the non-trivial limits of

$$a_n(D_\epsilon(P_n,Q)-D_\epsilon(P,Q)),$$

for certain sequence $\{a_n\}_{n\in\mathbb{N}}\subset\mathbb{R}$. Chapter 5 proves that the sequence in question depends on the hypothesis. In particular, we obtain the following limits:

- Under $H_0:\ P=Q$:

$$n\,D_\epsilon(P_n,P)\xrightarrow{w}\frac{1}{4}\sum_{j=1}^{\infty}\lambda_j\left(\sum_{i=1}^{\infty}x_{i,j}N_i\right)^2,$$

  where $\{N_i\}_{i\in\mathbb{N}}$ is a sequence of mutually independent random variables with $N_i\sim N(0,1)$ and $\{x_{i,j}\}_{i,j\in\mathbb{N}}\subset\mathbb{R}$, $\{\lambda_j\}_{j\in\mathbb{N}}\subset[0,\infty)$ real (and deterministic) sequences depending on P and ϵ.

- Under $H_1:\ P\neq Q$:

$$\sqrt{n}(D_\epsilon(P_n,Q)-D_\epsilon(P,Q))\xrightarrow{w}N(0,\mathrm{Var}_P(\psi^\epsilon_{P,Q})),$$

  where $\psi^\epsilon_{P,Q}=f^\epsilon_{P,Q}-\frac{1}{2}(f^\epsilon_{P,P}+g^\epsilon_{P,P})$.

The technique is different for each hypothesis. Under H_1 the Efron-Stein inequality provides the first-order development. However, since $\psi^\epsilon_{P,Q}$ is zero when $P=Q$, this is not sufficient to obtain nontrivial limits under H_0. So a second-order development is necessary, which is a consequence of the first-order development of the potentials. The proof is long and tedious; divided in two different works. On the one hand, Chapter 4 shows the tightness of the potentials and the divergence. Meaning that there exists a constant c_d, depending only on d, such that

$$\mathbb{E}\|g_{P_n,Q}-g_{P,Q}\|^2_{\mathcal{C}^s(\Omega)},\ \mathbb{E}\|f_{P_n,Q}-f_{P,Q}\|^2_{\mathcal{C}^s(\Omega)}\leq\tfrac{c_d}{n}D_\Omega^{5(d+1)}e^{15D_\Omega^2},$$

and

$$\mathbb{E}D_1(P_n,P) \leq \frac{c_d}{n} D_\Omega^{\frac{3d}{2}+1} \frac{32}{(d+1)^2} e^{\frac{19}{2}D_\Omega^2}.$$

On the other hand, Chapter 5 shows the first-order development of $\sqrt{n}\begin{pmatrix} f^\epsilon_{P_n,Q} - f^\epsilon_{P,Q} \\ g^\epsilon_{P_n,Q} - g^\epsilon_{P,Q} \end{pmatrix}$, in $\mathcal{C}^\alpha(\Omega) \times \mathcal{C}^\alpha(\Omega)$, with respect to the empirical process $\sqrt{n}(P_n - P)$, which follows

$$\begin{pmatrix} f^\epsilon_{P_n,Q} - f^\epsilon_{P,Q} \\ g^\epsilon_{P_n,Q} - g^\epsilon_{P,Q} \end{pmatrix} = \begin{pmatrix} (1 - \mathcal{A}^\epsilon_Q \mathcal{A}^\epsilon_P)^{-1} \mathcal{A}^\epsilon_Q \mathbb{G}^n_{P,s} \\ -(1 - \mathcal{A}^\epsilon_P \mathcal{A}^\epsilon_Q)^{-1} \mathbb{G}^n_{P,s} \end{pmatrix} + o_P\left(\frac{1}{\sqrt{n}}\right).$$

Here $\mathbb{G}^n_{P,s}$ denotes $\frac{1}{n}\sum_{k=1}^n \xi^\epsilon_{P,Q}(\mathbf{X}_k,\cdot) - \mathbb{E}\left(\xi^\epsilon_{P,Q}(\mathbf{X},\cdot)\right)$, and

$$\begin{aligned} \mathcal{A}^\epsilon_P &: L^2(P) \ni f \mapsto \int \xi^\epsilon_{P,Q}(\mathbf{x},\cdot) f(\mathbf{x}) dP(\mathbf{x}) \in \mathcal{C}^\alpha(\Omega), \\ \mathcal{A}^\epsilon_Q &: L^2_0(Q) \ni g \mapsto \int \xi^\epsilon_{P,Q}(\cdot,\mathbf{y}) g(\mathbf{y}) dQ(\mathbf{y}) \in \mathcal{C}^\alpha(\Omega). \end{aligned} \tag{1.17}$$

This first-order development of the potentials allows, through the formula

$$d\pi^\epsilon_{P,Q} = e^{\frac{f_{P,Q}+g_{P,Q}-\frac{1}{2}\|\cdot-\cdot\|^2}{\epsilon}} dP dQ,$$

to obtain the weak limits of the solutions of the primal problem, i.e.

$$\sqrt{n}\int \eta\, (d\pi^\epsilon_{P_n,Q} - d\pi^\epsilon_{P,Q}), \quad \text{with } \eta \in L^2(P \times Q).$$

In particular, we obtain

$$\sqrt{n}\left(\int \eta d\pi^\epsilon_{P_n,Q} - \int \eta d\pi^\epsilon_{P,Q}\right) \xrightarrow{w} N(0, \sigma^2_{\lambda,\epsilon}(\eta)), \quad \eta \in L^2(P \times Q), \tag{1.18}$$

where $\sigma^2_\lambda(\eta) = \mathrm{Var}_{\mathbf{X}\sim P}\left((1 - \mathcal{A}^\epsilon_Q \mathcal{A}^\epsilon_P)^{-1}(\eta^\epsilon_{\mathbf{x}} - \mathcal{A}^\epsilon_Q \eta^\epsilon_{\mathbf{y}})(\mathbf{X})\right)$, thus proving a conjecture of Harchaoui et al. (2020). (1.18) provides consistent confidence intervals for $\pi^\epsilon_{P_n,Q}$, which allows, among other things, the possibility to perform inference on the Sinkhorn distance (apply (1.18) to the function $(\mathbf{x},\mathbf{y}) \to \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2$) defined in Cuturi (2013) o the regularized colocalization measure RCol (apply (1.18) to the function $(\mathbf{x},\mathbf{y}) \to \mathbb{1}_{\|\mathbf{x}-\mathbf{y}\|^2 \leq t}$) defined in Klatt et al. (2020).

1.4 Multivariate center-outward distribution function; regularity and quantile regression.

The lack of canonical order of $\mathbb{R}^d$, for $d \geq 2$, prevents the generalization of statistical tools based on the univariate order relation. One of the most important is the distribution function of a univariate random variable $X \sim P$, defined as $F : x \to \mathbb{P}(X \leq x)$.

Many attempts have been made to define a multivariate distribution function. These include those based on depths, on copulas, on component-wise ranks, on spatial ranks or on Mahalanobis ranks; we refer to Hallin et al. (2021a) and references therein. All of them have to deal with the mentioned absence of canonical order, but none of them is able to mimic the properties that make the univariate distribution function useful for statistical inference. That is; distribution freeness (the distribution of $F(X)$ is uniform on $[0, 1]$ while its empirical counterpart $\{F^n(X_i)\}_{i=1}^n$ is uniform on the set $\{\frac{i}{n}\}_{i=1}^n$, irrespective of the distribution of the X_i's); satisfies the Glivenko-Cantelli property (the empirical approximation converges a.s. uniformly in $\mathbb{R}$) and maintains order, i.e. it is monotone.

Monotonicity and the fact that $F(X)$ is uniform on $[0, 1]$ uniquely define F, for any continuous random variable $X \sim P$. Monotonicity is equivalent to being the gradient of a convex function, distribution freeness to F pushing forward P towards $U_{(0,1)}$ (the uniform distribution in $(0, 1)$). Of course, the quantile function Q is the gradient of a convex function pushing forward $U_{(0,1)}$ toward P. These properties led Chernozhukov et al. (2017) to define a candidate multivariate quantile function, the so-called quantile distribution function with respect to a reference measure $\mathcal{U}$, defined, for a random variable $\mathbf{X} \sim P \in \mathcal{P}(\mathbb{R}^d)$, as the unique gradient of a convex function (defined at almost every point) that pushes forward $\mathcal{U}$ toward P. Its existence holds via the celebrated McCann's theorem McCann (1995).

The current debate is about the choice of an *idoneus* reference measure $\mathcal{U}$. Deb and Sen (2019) advocates for a uniform measure on the hypercube (successful for testing independence), Hallin et al. (2021a) for a spherical uniform (successful for giving a notion of center-outward ordering) and Deb et al. (2021) for a standard Gaussian (successful for Hotelling's T^2 test). The most correct answer is that it depends on the purpose (Hallin and Mordant, 2021). Here we will deal with the case of the spherical distribution on the unit ball $\mathbf{U}_d$, i.e. the one obtained by independently taking a radius $r \sim U_{(0,1)}$ and a direction θ uniform on the sphere $\mathcal{S}^{d-1}$. The advantage of using the spherical uniform as a reference measure lies in its invariance to rotations (changes of basis) and its uniform control of the balls centered at the origin and of radius $\tau \in (0, 1)$, i.e.

$$\mathbf{U}_d(\tau \mathbb{S}^d) = \tau, \quad \text{for } \tau \in (0, 1). \tag{1.19}$$

No other probability measure in $\mathbb{R}^d$ satisfies these properties. (1.19) provides a natural radial ordering and a clear notion of center –thus a center-outward ordering. The quantile function $\mathbf{Q}_\pm$ is defined in this case as the unique gradient of a convex function ψ (defined at almost every point) that pushes forward $\mathbf{U}_d$ toward P. The distribution function is defined as

$$\mathbf{F}_\pm(\mathbf{x}) := \arg \sup_{\|\mathbf{u}\| \leq 1} \{\langle \mathbf{x}, \mathbf{u} \rangle - \psi(\mathbf{u})\},$$

which coincides with the inverse of $\mathbf{Q}_\pm$ in the support of P. The importance given by $\mathbf{U}_d$ to the origin and that notion of center-outward ordering are inherited by P via $\mathbf{Q}_\pm$. This gives rise to the name *center-outward distribution and quantile map* coined in Hallin et al. (2021a) for $\mathbf{F}_\pm$ and $\mathbf{Q}_\pm$. This terminology will be used from now on.

Assume the observation of just an i.i.d. sample $\mathbf{X}^{[n]} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ of P, its empirical center-outward distribution $\mathbf{F}_\pm^n$ is defined, for each $\mathbf{X_i}$ as $\mathbf{F}_\pm^n(\mathbf{X_i}) = \mathfrak{G}_{\sigma_n(i)}^{(n)}$. Here σ_n is the one-to-one correspondence in $\{1, \ldots, n\}$ where

$$\sum_{i=1}^{n} \|\mathbf{X}_i - \mathfrak{G}_{\sigma(i)}^{(n)}\|^2 \tag{1.20}$$

achieves its minimum, and $\mathfrak{G}_1^{(n)}, \ldots, \mathfrak{G}_n^{(n)}$ is a regular grid of the uniform spherical of $\mathbb{U}_d$. It consists of factoring $n = n_s n_d + n_0$,with $n < \min(n_s, n_d)$ and computing n points $\mathfrak{G}_1^{(n)}, \ldots, \mathfrak{G}_n^{(n)}$ created by the intersection between

- the rays generated by an n_S-tuple $\mathbf{u}_1, \ldots, \mathbf{u}_{n_S} \in \mathcal{S}_{d-1}$ of unit vectors such that ${n_S}^{-1} \sum_{j=1}^{n_S} \delta_{\mathbf{u}_j}$ converges weakly to the uniform over $\mathcal{S}_{d-1}$ as $n_S \to \infty$, and
- the n_R hyperspheres with center $\mathbf{0}$ and radii $j/(n_R + 1)$, $j = 1, \ldots, n_R$,

the points corresponding to n_0 are identified with $\mathbf{0}$. The Glivenko-Cantelli property

$$\max_{1 \le k \le n} \|\mathbf{F}_\pm^n(\mathbf{X}_k) - \mathbf{F}_\pm(\mathbf{X}_k)\| \xrightarrow{a.s.} 0 \tag{1.21}$$

is valid if $\mathbf{F}_\pm$ is well defined on the whole space (recall that McCann's theorem yields a Lebesgue a.s. definition). This is quite common when it comes to proving the convergence of gradients of convex functions. In particular, three steps are usually followed; prove the weak convergence of the transport plans; prove the convergence of the sub-differentials in the sense of sets; and finally, prove the convergence of the potentials. On the compact sets where the sub-differential is single-valued (i.e. where the function is smooth) the limit is uniform, the particular (strongly convex) form of the support of $\mathbb{U}_d$, allows us to pass from uniform convergence in the compact sets to (global) uniform convergence.

If $\mathbf{F}_\pm$ is well-defined over the entire space, as a consequence of (1.19), the center-outward quantile map defines nested regions $\mathbb{C}_P(\tau) = \mathbf{Q}_\pm(r\,\mathbb{S}^d)$ and continuous contours $\mathcal{C}_P(\tau) = \mathbf{Q}_\pm(r\,\mathcal{S}^{d-1})$ indexed by $\tau \in ([0,1)$, such that, for all P absolutely continuous with respect to the Lebesgue measure, $P\big[\mathbb{C}_P(\tau)\big] = \tau$, irrespective of P. No other choice of reference measure defines a transport-based multivariate distribution function sharing the last property for all dimension, which is the key of the multivariate quantile regression (see Chapter 7).

1.4.1 Regularity of the the center-outward distribution

The center-outward distribution function $\mathbf{F}_\pm$ is well defined if and only if it is continuous (cf. Rockafellar (1970)). Therefore, the study of its regularity is important. It is known that in one dimension the distribution function is continuous for all absolutely continuous (with respect to the Lebesgue measure) probabilities. Similar results for its multivariate counterpart should be desirable. The regularity of convex functions satisfying a transport condition is usually dealt by means of the Monge-Ampère equation (see Figalli (2017)). We will see that in this case the continuity of $\mathbf{F}_{\pm}$ is attained by assuming that P has density p satisfying:

Assumption A. For any $R > 0$, there exists $0 < \lambda_R \leq \Lambda_R$ such that

$$\lambda_R \leq p(\mathbf{x}) \leq \Lambda_R \quad \text{for all } \mathbf{x} \in \mathcal{X} \cap R\,\mathbb{S}^d. \tag{1.22}$$

In particular, $\mathbf{F}_{\pm}$ satisfies the condition of pushing forward P to Q and it is the gradient of a convex function φ, so that its sub-differential satisfies

$$\int_{\partial\varphi(A)} u_d(\mathbf{y})d\mathbf{y} = \int_A p(\mathbf{x})d\mathbf{x}. \tag{1.23}$$

The main result of the Chapter 6 is as follows.

Theorem 1.4.1. *Let* P *be a probability measure with density* p *supported on the open convex set* $\mathcal{X} \subseteq \mathbb{R}^d$.

(i) *If* p *satisfies* (6.11)*, then* $K := \partial\psi(\mathbf{0})$ *is a compact, convex set with Lebesgue measure* 0 *such that the center-outward quantile function* $\mathbf{Q}_{\pm} := \nabla\psi$ *and the center-outward distribution function* $\mathbf{F}_{\pm} := \nabla\psi^*$ *are homeomorphisms between* $\mathbb{S}^d \setminus \{\mathbf{0}\}$ *and* $\mathcal{X} \setminus K$*, inverses of each other.*

(ii) *If, moreover,* $p \in \mathcal{C}^{k,\alpha}_{\text{loc}}(\mathcal{X})$ *for some* $k \in \mathbb{N}$ *and* $\alpha \in (0,1)$*, then* $\mathbf{Q}_{\pm}$ *and* $\mathbf{F}_{\pm}$ *are diffeomorphisms of class* $\mathcal{C}^{k+1,\alpha}_{\text{loc}}$ *between* $\mathbb{S}^d \setminus \{\mathbf{0}\}$ *and* $\mathcal{X} \setminus K$.

The proof is based on Caffarelli's theory (see eg. Caffarelli (1990, 1991, 1992)), which has been extensively studied by Figalli in Figalli and Kim (2010); Philippis and Figalli (2012); Cordero-Erausquin and Figalli (2019). I recommend the book Figalli (2017) as an introductory guide to this topic and its implications for the regularity of the transport map. Up to Figalli (2018), the regularity of $\mathbf{F}$ was not covered by any of these works, see Cordero-Erausquin and Figalli (2019) for a study of the most general case. However the assumptions of Figalli (2018) (it is assumed **A.** and that P is supported in the whole space) are too strong, as we will see in Chapter 6, they can be be relaxed .

As a consequence, if P is a probability with density p supported on the convex set $\mathcal{X} \subseteq \mathbb{R}^d$ such that p satisfies (1.22), the following properties are satisfied.

- The Glivenko-Cantelli property (1.21) (Hallin et al., 2021a, Proposition 2.4) holds. Which implies that the population center-outward distribution function can be uniformly estimated from the sample.
- For all $\mathbf{u}$ on the unit sphere $\mathcal{S}^{d-1}$, all sequence $(t_n)_{n\in\mathbb{N}}$ of real numbers such that $t_n \to \infty$, the limit

$$\lim_{n\to\infty} \mathbf{F}_{\pm}(t_n\mathbf{u}) = \mathbf{u} \tag{1.24}$$

holds. This property is the analogous to $\lim_{n\to\infty} F(t_n) = 1$ and $\lim_{n\to\infty} F(-t_n) = 0$ in the univariate case, that is, from the center-outward perspective.

- For all $r \in (0,1)$ and all $\mathbf{y}$ belonging to the boundary of $\mathbf{Q}_{\pm}(r\,\mathbb{S}^d)$, there exists a ray T emanating from $\mathbf{y}$ for which $\overline{\mathbf{Q}_{\pm}(r\,\mathbb{S}^d)} \cap T = \{\mathbf{y}\}$. In colloquial words, for every point of a quantile contour there exists a ray that does not touch the set again. Moreover $\mathbb{C}_P(\tau)$ are nested and connected regions.

- If, in addition, $\mathcal{X}$ is compact, for all $r \to 1$, $\mathbf{Q}_{\pm}(r\,\mathbb{S}^d)$ tends to $\mathcal{X}$ in Hausdoff distance.

$$\lim_{r\to 1} \mathrm{d}_H\big(\mathbf{Q}_{\pm}(r\,\mathbb{S}^d), \mathcal{X}\big) = 0. \tag{1.25}$$

The support of P can be approximated by the limits of the quantile regions.

1.4.2 Nonparametric quantile regression with multivariate output

Chapter 7 proposes a novel and significant nonparametric and multivariate extension, based on the previously introduced multivariate quantiles, of Koenker and Bassett's famous concept of quantile regression Koenker and Bassett (1978), a powerful tool in the statistical study of the dependence of a variable of interest Y with respect to the covariates $\mathbf{X} = (X_1, \ldots, X_m)$. Unlike classical regression, which, in a sense, focuses on the conditional means $\mathrm{E}[Y|\mathbf{X}]$, quantile regression requires full knowledge of the conditional distributions $\mathrm{P}_{\mathrm{Y}|\mathbf{X}=\mathbf{x}}$ of Y given $\mathbf{X} = \mathbf{x}$. Its non-parametric formulation, whose study begins with the pioneering work of Stone (1977), has become part of everyday statistical practice, with countless applications in all areas of scientific research in which a finite number of parameters provide a model too rigid to explain certain observed behavior.

The primary motivation for the use of quantile regression is the observation that $\mathrm{P}_{\mathrm{Y}|\mathbf{X}=\mathbf{x}}$ is much more informative than simple knowledge of a parameter of interest, such as conditional mean or median. Figure 1.2 shows an example of univariate quantile regression, with X =age and Y =triceps skinfold. Looking only at the median (red line) we would only be able to obtain a description of the trend of the model, however, taking into account the information given by the quantile tubes, we can observe the heteroscedasticity of the model.

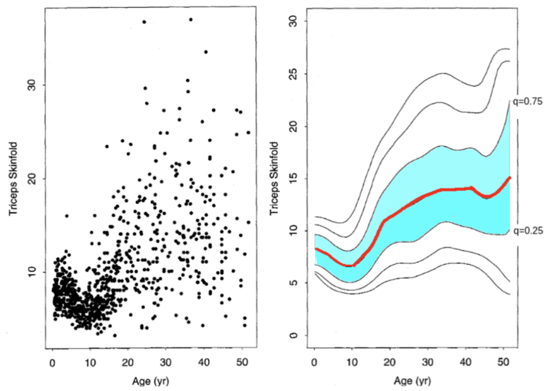

Figure 1.2: Example of univariate quantile regression X =age and Y =triceps skinfold. Figure extracted from Yu and Jones (1998). Red line represents the median and the blue region the center-outward quantile region of order 0.5.

The trend function in the quantile regression is the conditional median, which corresponds to $Q(\frac{1}{2}|\,\mathbf{x})$. However, the fundamental property of quantile regression, which differentiates it from all other regression models, *"probabilistic control"* of the tubes and quantile regions, i.e.

$$\mathbb{P}\left[Y \in \left[Q(\frac{1}{2} - \frac{\tau}{2}|\,\mathbf{x}), Q(\frac{1}{2} + \frac{\tau}{2}|\,\mathbf{x})\right] \Big|\, \mathbf{X} = \mathbf{x}\right] = \tau,$$

irrespective of $\mathbb{P}$. Thanks to this property we know, for example, that the blue region the figure 1.2 has a probability of 0.5. This is the basic principle of quantile regression. Any regression model that does not satisfy it cannot be considered as quantile regression, and if it was intended to be quantile regression, it has certainly missed its target.

Of course, we cannot be satisfied with a definition of quantile regression that is valid only for univariate outputs; consider for example the following model:

$$\mathbf{Y} = \begin{pmatrix} Y \\ Z \end{pmatrix} = \begin{pmatrix} V \\ \frac{U\sqrt{|X-\frac{1}{2}|}+V\,(1-\sqrt{|X-\frac{1}{2}|})}{\sqrt{|X-\frac{1}{2}|+(1-\sqrt{|X-\frac{1}{2}|})^2}} \end{pmatrix}, \quad X \sim \mathcal{U}_{(0,1)},\ U, V \sim \mathcal{N}(0,1), \quad (1.26)$$

being X, U and V mutually independent. The univariate regression on the marginals does not give us any useful information, both are independent of X. All the information is in the joint distribution. The reader may wonder if projecting across all directions would give us more information. Indeed more projections will give more information, since these characterize the probability of the vector $\mathbf{Y}$. This is along the lines of integrated depths (see Cuevas and Fraiman (2009)), which, like all depth-based regression (for a fuller thorough treatment, see (Serfling and Zuo, 2000) or (Serfling, 2002, 2019) and the references given there), loses probabilistic control of the regions.

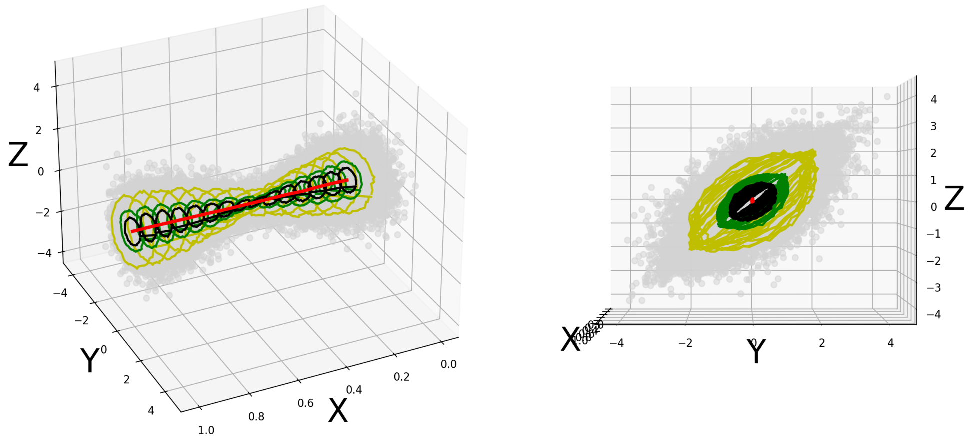

Figure 1.3: Example of univariate quantile regression; with X and $\mathbf{Y} = (Y, Z)$ as in (1.26). The number of observed points is $n = 7,000$. The quantile contours for $\tau = 0.2, 0.4, 0.8$ are represented in black ($\tau = 0.2$), green ($\tau = 0.4$) and yellow ($\tau = 0.8$). The center is represented in red.

The methodology proposed in Chapter 7 defines regions and quantile tubes, capable of analyzing nonparametric models while maintaining a probabilistically asymptotic control of the regions–the empirical probability control converges to the population one, which is known. Figure 1.3 analyzes the model (1.26), giving at the same time a very visual solution, the trend of the model is –as expected– constant however the covariate X reaches its peak of influence on the vector of interest in the direction $(1, 1)$.

The center-outward quantile function $\mathbf{Q}_{\pm}$ *of* $\mathbf{Y}$ *given* $\mathbf{X} = \mathbf{x}$ is defined as

$$\mathbf{u} \in \mathbb{S}_d \mapsto \mathbf{Q}_{\pm}(\mathbf{u}|\,\mathbf{x}) \in \mathbb{R}^d.$$

It naturally defines the conditional quantile regions as

$$\mathbb{C}_{\pm}(\tau|\mathbf{x}) := \mathbf{Q}_{\pm}(\tau\,\overline{\mathbb{S}}_d|\mathbf{x}) \quad \tau \in (0,1), \;\; \mathbf{x} \in \mathbb{R}^m,$$

which satisfy the fundamental property

$$\mathbb{P}\left[\mathbf{Y} \in \mathbb{C}_{\pm}(\tau|\,\mathbf{x})\,\middle|\, \mathbf{X} = \mathbf{x}\right] = \tau \quad \text{ for all } \mathbf{x} \in \mathbb{R}^m, \tau \in (0,1), \text{ and } \mathbb{P}.$$

For $\tau = 0$, the conditional median is

$$\mathbb{C}_{\pm}(0\,|\,\mathbf{x}) := \bigcap_{\tau \in (0,1)} \mathbb{C}_{\pm}(\tau|\,\mathbf{x})$$

It also characterizes the nested (non *"quantile-crossing"* phenomenon) *"regression quantile tubes of order* $\tau \in (0,1)$*"* (in $\mathbb{R}^{m+d}$)

$$\mathbb{T}_{\pm}(\tau) := \left\{ \left(\mathbf{x}, \mathbf{Q}_{\pm}(\tau\, \overline{\mathbb{S}}_d \middle|\, \mathbf{x})\right) \middle| \mathbf{x} \in \mathbb{R}^m \right\}, \quad \tau \in (0,1)$$

which satisfy the fundamental property

$$\mathbb{P}\left[(\mathbf{X}, \mathbf{Y}) \in \mathbb{T}_{\pm}(\tau)\right] = \tau \quad \text{irrespective of } \mathbb{P}, \tau \in (0,1).$$

For $\tau = 0$, we define

$$\mathbb{T}_{\pm}(0) := \left\{ (\mathbf{x}, \mathbf{y})) \middle| \mathbf{x} \in \mathbb{R}^m, \, \mathbf{y} \in \mathbb{C}_{\pm}(0 \,|\, \mathbf{x}) \right\} = \bigcap_{\tau \in (0,1)} \mathbb{T}_{\pm}(\tau)$$

(as the *graph* of $\mathbf{x} \mapsto \mathbb{C}_{\pm}(\tau|\, \mathbf{x})$); with a slight abuse of language, we also call it *regression median* of $\mathbf{Y}$ with respect to $\mathbf{X}$.

When what is observed is a sample $(\mathbf{X}, \mathbf{Y})^{(n)} := \left((\mathbf{X}_1, \mathbf{Y}_1), \ldots, (\mathbf{X}_n, \mathbf{Y}_n)\right)$ of n i.i.d. copies of $(\mathbf{X}, \mathbf{Y}) \sim \mathrm{P}_{\mathbf{XY}}$, Chapter 7 proposes an estimator of $\mathbf{u} \mapsto \mathbf{Q}_{\pm}(\mathbf{u}|\mathbf{X} = \mathbf{x})$, $\mathbf{x} \in \mathbb{R}^m$. Our estimator is obtained in two steps: in step 1, we construct an empirical distribution of $\mathbf{Y}$ conditional to $\mathbf{X} = \mathbf{x}$ and, in step 2, we compute the corresponding center-outward quantile.

The empirical distribution of $\mathbf{Y}$ conditional to $\mathbf{X} = \mathbf{x}$ is approximated by a weight function, so (if we do not want to restrict ourselves to piecewise constant weight functions) this will be an atomic probability non-equal weights. The solution we obtain requires the solution of the following optimal transport problem:

$$\begin{aligned}
&\min_{\pi := \{\pi_{i,j}\}} \sum_{i=1}^{N} \sum_{j=1}^{n} \frac{1}{2} \left|\mathbf{Y}_j - \mathfrak{G}_i\right|^2 \pi_{i,j}\,, \\
&\text{s.t.} \ \sum_{j=1}^{n} \pi_{i,j} = \frac{1}{N}, \quad i \in \{1, 2, \ldots, N\}, \\
&\qquad \sum_{i=1}^{N} \pi_{i,j} = w_j^{(n)}(\mathbf{x}; \mathbf{X}^{(n)}), \quad j \in \{1, 2, \ldots, n\}, \\
&\qquad \pi_{i,j} \geq 0, \quad i \in \{1, 2, \ldots, N\}, \ j \in \{1, 2, \ldots, n\},
\end{aligned} \tag{1.27}$$

where $\mathfrak{G}_1^{(n)}, \ldots, \mathfrak{G}_N^{(n)}$ is a regular grid of $\mathbb{U}_d$. In this case N is arbitrary, we will only ask it to tend to infinity. The problem is that the solution of (1.27) is a transport plan and not a map as such. To create a map we apply the criterion

$$\mathbf{T}^*(\mathfrak{G}_i|\, \mathbf{x}) := \arg\inf \left\{ \|\mathbf{y}\| : \ \mathbf{y} \in \mathrm{conv}\left(\{\mathbf{Y}_J : \ J \in \arg\max_j \pi^*_{i,j}(\mathbf{x})\} \right) \right\}, \tag{1.28}$$

where $\mathrm{conv}(A)$ denotes the convex envelope of A.We proceed as in Hallin et al. (2021a) by choosing the cyclically monotonic smooth interpolation with the largest Lipschitz constant.

Let us denote it as $\mathbf{u} \mapsto \mathbf{Q}^{(n)}_{w,\pm}(\mathbf{u}|\mathbf{x})$ and call it the *empirical center-outward conditional quantile map*.

We will show that if the weight function is consistent (in the sense of Stone (1977)), for all $\mathbf{u} \in \mathbb{S}_d$ and $\epsilon > 0$,

$$\mathbb{P}\left(\mathbf{Q}^{(n)}_{w,\pm}(\mathbf{u}\,|\mathbf{X}) \notin \mathbf{Q}_{\pm}(\mathbf{u}\,|\mathbf{X}) + \epsilon\mathbb{S}_d \right) \longrightarrow 0 \quad \text{as } n \text{ and } N \to \infty,$$

and for all $\tau \in (0,1)$,

$$\mathbb{P}\left(\mathbb{C}^{(n)}_{\pm}(\tau\,|\mathbf{X}) \not\subset \mathbb{C}_{\pm}(\tau\,|\mathbf{X}) + \epsilon\mathbb{S}_d \right) \to 0 \text{ and } \mathbb{P}\left(\mathcal{C}^{(n)}_{\pm}(\tau\,|\mathbf{X}) \not\subset \mathcal{C}^{(n)}_{\pm}(\tau\,|\mathbf{X}) + \epsilon\mathbb{S}_d \right) \to 0$$

as n and $N \to \infty$.

Remark 1.4.2. *The proof of this result follows the line of convergence of empirical multi-valued mappings, initiated by del Barrio and Loubes (2019) for proving convergence of optimal transport potentials in the quadratic case, continued by Hallin et al. (2021a) for the proof of the multivariate Glivenko-Cantelli theorem and finally formalized in Segers (2022). See also del Barrio et al. (2021) for the general cost case. In this case we have an added difficulty; the convergence is in probability, and, therefore, we need to pass numerous times through sub-sequences. It is a tedious procedure.*

Chapter 6 has its application in the 7-th; it is able to give the conditions that $(\mathbf{X}, \mathbf{Y})$ must satisfy in order to obtain stronger notions of convergence. In this case, assuming that $\mathbf{A}$ holds a.s. for the conditional probability, then, for all compact $K \subset \mathbb{S}_d \setminus \{\mathbf{0}\}$, when n and $N \to \infty$,

$$\sup_{\mathbf{u}\in K} |\mathbf{Q}^{(n)}_{w,\pm}(\mathbf{u}\,|\mathbf{X}) - \mathbf{Q}_{\pm}(\mathbf{u}\,|\mathbf{X})| \xrightarrow{\mathbb{P}} 0$$

and, for all $\tau \in (0,1)$ and $\epsilon > 0$,

$$\mathbb{P}\left(d_H\left(\mathcal{C}^{(n)}_{\pm}(\tau\,|\mathbf{X}), \mathcal{C}_{\pm}(\tau\,|\mathbf{X}) \right) > \epsilon \right) \to 0.$$

Moreover, under these conditions, we obtain the asymptotic probability control

$$\mathbb{P}\left(\mathbf{Y} \in \mathbb{C}^{(n)}_{\pm}(\tau\,|\mathbf{X}) | \mathbf{X} \right) \xrightarrow{\mathbb{P}} \tau \quad \text{for all } \tau \in (0,1),$$

that justifies the proposed methodology.

Chapter 7 ends with a series of experiments where good behaviour in models including heteroscedasticity and nonlinear trends is shown; its power as a data analysis tool is also illustrated on some real data sets.

Bibliography

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 214–223. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/arjovsky17a.html.

Emily Black, Samuel Yeom, and Matt Fredrikson. Fliptest: Fairness testing via optimal transport. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, page 111–121, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450369367. doi: 10.1145/3351095.3372845. URL https://doi.org/10.1145/3351095.3372845.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 02 2013. ISBN 9780199535255. doi: 10.1093/acprof:oso/9780199535255.001.0001. URL https://doi.org/10.1093/acprof:oso/9780199535255.001.0001.

Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991. doi: https://doi.org/10.1002/cpa.3160440402. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160440402.

Luis Caffarelli. Some regularity properties of solutions of monge ampère equation. *Communications on Pure and Applied Mathematics*, 44:965–969, 1991.

Luis A. Caffarelli. A localization property of viscosity solutions to the monge-ampere equation and their strict convexity. *Annals of Mathematics*, 131(1):129–134, 1990. ISSN 0003486X. URL http://www.jstor.org/stable/1971509.

Luis A. Caffarelli. The regularity of mappings with a convex potential. *Journal of the American Mathematical Society*, 5(1):99–104, 1992. ISSN 08940347, 10886834. URL http://www.jstor.org/stable/2152752.

Javier Cárcamo, Antonio Cuevas, and Luis-Alberto Rodríguez. Directional differentiability for supremum-type functionals: Statistical applications. *Bernoulli*, 26(3):2143 – 2175, 2020. doi: 10.3150/19-BEJ1188. URL https://doi.org/10.3150/19-BEJ1188.

Victor Chernozhukov, Alfred Galichon, Marc Hallin, and Marc Henry. Monge-Kantorovich depth, quantiles, ranks and signs. *Ann. Statist.*, 45(1):223–256, 2017. ISSN 0090-5364. doi: 10.1214/16-AOS1450. URL https://doi.org/10.1214/16-AOS1450.

Dario Cordero-Erausquin and Alessio Figalli. Regularity of monotone transport maps between unbounded domains. *Discrete and Continuous Dynamical Systems*, 39(12): 7101–7112, 2019.

Juan A. Cuesta and Carlos Matrán. Notes on the wasserstein metric in hilbert spaces. *Annals of Probability*, 17:1264–1276, 1989.

Antonio Cuevas and Ricardo Fraiman. On depth measures and dual statistics. a methodology for dealing with general data. *Journal of Multivariate Analysis*, 100(4):753–766, 2009. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2008.08.002. URL https://www.sciencedirect.com/science/article/pii/S0047259X08001796.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pages 2292–2300, 2013. URL http://papers.nips.cc/paper/4927-sinkhorn-distances-lightspeed-computation-of-optimal-transport.

Lucas de Lara, Alberto González-Sanz, Nicholas Asher, and Jean-Michel Loubes. Transport-based counterfactual models. *Pre-Print*, 2021.

Lucas De Lara, Alberto González-Sanz, and Jean-Michel Loubes. Diffeomorphic registration using sinkhorn divergences. *To appear in SIAM Journal on Imaging Sciences*, 2022+.

Nabarun Deb and Bodhisattva Sen. Multivariate rank-based distribution-free nonparametric testing using measure transportation. *J. Amer. Statist. Assoc.*, 2019. To appear.

Nabarun Deb, Bhaswar B. Bhattacharya, and Bodhisattva Sen. Efficiency lower bounds for distribution-free hotelling-type two-sample tests based on optimal transport, 2021. URL https://arxiv.org/abs/2104.01986.

Eustasio del Barrio and Jean-Michel Loubes. Central limit theorems for empirical transportation cost in general dimension. *The Annals of Probability*, 47(2):926 – 951, 2019. doi: 10.1214/18-AOP1275. URL https://doi.org/10.1214/18-AOP1275.

Eustasio del Barrio, Evarist Giné, and Frederic Utzet. Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli*, 11(1):131–189, 2005. ISSN 1350-7265. doi: 10.3150/bj/1110228245. URL https://doi.org/10.3150/bj/1110228245.

Eustasio del Barrio, Alberto González-Sanz, and Jean-Michel Loubes. Central limit theorems for general transportation costs. *Pre-Print*, 2021.

R. M. Dudley. The speed of mean glivenko-cantelli convergence. *Ann. Math. Statist.*, 40 (1):40–50, 02 1969. doi: 10.1214/aoms/1177697802. URL https://doi.org/10.1214/aoms/1177697802.

Jean Feydy, Benjamin Charlier, François-Xavier Vialard, and Gabriel Peyré. Optimal transport for diffeomorphic registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 291–299. Springer, 2017.

Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trouvé, and Gabriel Peyré. Interpolating between optimal transport and mmd using sinkhorn divergences. In *AISTATS*, 2019.

Alessio Figalli. *The Monge-Ampère Equation and Its Applications*. Zurich Lectures in Advanced Mathematics, European Mathematical Society (EMS), Zurich, 2017. ISBN 978-3-03719-670-0.

Alessio Figalli. On the continuity of center-outward distribution and quantile functions. *arXiv: Analysis of PDEs*, 2018.

Alessio Figalli and Young-Heon Kim. Partial regularity of brenier solutions of the monge-ampère equation. *Discrete and Continuous Dynamical Systems*, 28:559–565, 2010.

N. Fournier and A. Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162:707–738, 2013.

Alfred Galichon. *Optimal Transport Methods in Economics*. Princeton University Press, 09 2016. ISBN 9780691172767. doi: 10.23943/princeton/9780691172767.001.0001. URL https://doi.org/10.23943/princeton/9780691172767.001.0001.

Wilfrid Gangbo and Robert J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996. doi: 10.1007/BF02392620. URL https://doi.org/10.1007/BF02392620.

Aude Genevay, Lénaïc Chizat, Francis Bach, Marco Cuturi, and Gabriel Peyré. Sample complexity of sinkhorn divergences. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pages 1574–1583. PMLR, 16–18 Apr 2019. URL http://proceedings.mlr.press/v89/genevay19a.html.

Javier González-Delgado, Alberto González-Sanz, Juan Cortés, and Pierre Neuvial. Two-sample goodness-of-fit tests on the flat torus based on wasserstein distance and their relevance to structural biology. *arXiv: Probability*, 2021.

Paula Gordaliza, Eustasio Del Barrio, Gamboa Fabrice, and Jean-Michel Loubes. Obtaining fairness using optimal transport theory. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 2357–2365. PMLR, 09–15 Jun 2019a. URL http://proceedings.mlr.press/v97/gordaliza19a.html.

Paula Gordaliza, Eustasio Del Barrio, Gamboa Fabrice, and Jean-Michel Loubes. Obtaining fairness using optimal transport theory. In *International Conference on Machine Learning*, pages 2357–2365. PMLR, 2019b.

Marc Hallin and Gilles Mordant. On the finite-sample performance of measure transportation-based multivariate rank tests, 2021.

Marc Hallin, Eustasio del Barrio, Juan Cuesta-Albertos, and Carlos Matrán. Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *The Annals of Statistics*, 49(2):1139 – 1165, 2021a. doi: 10.1214/20-AOS1996.

Marc Hallin, Gilles Mordant, and Johan Segers. Multivariate goodness-of-fit tests based on Wasserstein distance. *Electron. J. Stat.*, 15(1):1328–1371, 2021b. doi: 10.1214/21-ejs1816. URL https://doi.org/10.1214/21-ejs1816.

Zaïd Harchaoui, Lang Liu, and Soumik Pal. Asymptotics of entropy-regularized optimal transport via chaos decomposition. *arXiv: Probability*, 2020.

Shayan Hundrieser, Marcel Klatt, Thomas Staudt, and Axel Munk. A unifying approach to distributional limits for empirical optimal transport. *Pre-Print*, 02 2022. URL https://arxiv.org/pdf/2202.12790.pdf.

Lev Kantorovich. On the translocation of masses. *Journal of Mathematical Sciences*, 133: 1381–1382, 2006.

Marcel Klatt, Carla Tameling, and Axel Munk. Empirical regularized optimal transport: statistical theory and applications. *SIAM J. Math. Data Sci.*, 2(2):419–443, 2020. doi: 10.1137/19M1278788. URL https://doi.org/10.1137/19M1278788.

Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1913643.

Tudor Manole and Jonathan Niles-Weed. Sharp convergence rates for empirical optimal transport with smooth costs. *Pre-Print*, 2021.

Robert J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80(2):309 – 323, 1995. doi: 10.1215/S0012-7094-95-08013-2. URL https://doi.org/10.1215/S0012-7094-95-08013-2.

Gonzalo Mena and Jonathan Niles-Weed. Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/5acdc9ca5d99ae66afdfe1eea0e3b26b-Paper.pdf.

Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019. ISSN 1935-8237. doi: 10.1561/2200000073. URL http://dx.doi.org/10.1561/2200000073.

Guido Philippis and Alessio Figalli. Partial regularity for optimal transport maps. *Publications mathématiques de l'IHÉS*, 121, 09 2012. doi: 10.1007/s10240-014-0064-7.

Laurent Risser, Alberto González Sanz, Quentin Vincenot, and Jean-Michel Loubes. Tackling algorithmic bias in neural-network classifiers using wasserstein-2 regularization. *J. Math Imaging Vis*, 2021.

R. Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series. Princeton University Press, Princeton, N. J., 1970.

Johan Segers. Graphical and uniform consistency of estimated optimal transport plans. *arXiv*, 2022.

Robert Serfling. Quantile functions for multivariate analysis: approaches and applications. *Statistica Neerlandica*, 56(2):214–232, 2002. doi: https://doi.org/10.1111/1467-9574.00195. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9574.00195.

Robert Serfling. Depth functions on general data spaces, i. perspectives, with consideration of "density" and "local" depths". Available at https://www.utdallas.edu~serfling, 2019.

Robert Serfling and Yijun Zuo. General notions of statistical depth function. *The Annals of Statistics*, 28(2):461–482, 2000. doi: 10.1214/aos/1016218226. URL https://doi.org/10.1214/aos/1016218226.

Hongjian Shi, Mathias Drton, and Fang Han. Distribution-free consistent independence tests via center-outward ranks and signs. *J. Amer. Statist. Assoc.*, 117(537):395–410, 2022. ISSN 0162-1459. doi: 10.1080/01621459.2020.1782223. URL https://doi.org/10.1080/01621459.2020.1782223.

Max Sommerfeld and Axel Munk. Inference for empirical wasserstein distances on finite spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):219–238, 2018.

Charles J. Stone. Consistent nonparametric regression. *Ann. Statist.*, 5(4):595–620, 07 1977. doi: 10.1214/aos/1176343886. URL https://doi.org/10.1214/aos/1176343886.

Cédric Villani. *Topics in optimal transportation*. Number 58 in Graduate studies in mathematics. American Mathematical Soc., 2003.

Cédric Villani. *Optimal Transport: Old and New*. Number 338 in Grundlehren der mathematischen Wissenschaften. Springer, Berlin, 2008. ISBN 978-3-540-71049-3. OCLC: ocn244421231.

Richard A Vitale. Lp metrics for compact, convex sets. *Journal of Approximation Theory*, 45(3):280–287, 1985. ISSN 0021-9045. doi: https://doi.org/10.1016/0021-9045(85)90051-6. URL https://www.sciencedirect.com/science/article/pii/0021904585900516.

Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A):2620 – 2648, 2019. doi: 10.3150/18-BEJ1065. URL https://doi.org/10.3150/18-BEJ1065.

Keming Yu and M. C. Jones. Local linear quantile regression. *Journal of the American Statistical Association*, 93(441):228–237, 1998. doi: 10.1080/01621459.1998.10474104. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1998.10474104.

CHAPTER 1

Introduction (version française)

Le transport optimal est un problème d'allocation de ressources présent dans de nombreux domaines mathématiques; et donc dans leurs applications. Cette polyvalence se manifeste également dans son propre cadre théorique : l'étude de sa régularité repose sur des techniques avancées d'équations différentielles (Caffarelli, 1990, 1991, 1992); le développement de méthodes de calcul efficaces sur l'analyse numérique et de combinatoire (Peyré and Cuturi, 2019, Chapitre 3); son comportement asymptotique sur l'analyse convexe (del Barrio and Loubes, 2019) et les processus empiriques (del Barrio et al., 2005).

Le transport optimal consiste à trouver, parmi toutes les lois de probabilité jointes ayant des marginales fixes, celle qui minimise le coût de transport moyen. Cette valeur minimale moyenne est connue sous le nom de coût de transport optimal. Dans les cas où l'une des probabilités a une densité, la solution est déterministe et est donnée par l'application du transport (Gangbo and McCann, 1996). Le coût de transport optimal fournit une structure métrique, par la distance de Wasserstein, à l'espace des mesures de probabilité. Par conséquent, du point de vue statistique, le transport optimal offre un outil pour la comparaison de données qui tient compte de la géométrie de l'espace latent. Ce qui s'est avéré efficace pour résoudre des problèmes tels que la correction de biais en apprentissage automatique (Risser et al., 2021; Gordaliza et al., 2019a; Black et al., 2020), dans la modélisation de la raisonnement contrefactuels (de Lara et al., 2021; Black et al., 2020) ou dans l'appariement difféomorphique (Feydy et al., 2017; De Lara et al., 2022+).

Une application naturelle de toute distance entre distributions de probabilités est le problème de l'ajustement, c'est-à-dire le problème de tester l'hypothèse nulle selon laquelle un échantillon $\mathbf{X}_1, \dots, \mathbf{X}_n$ provient d'une population avec une distribution complètement spécifiée P (Hallin et al., 2021b). En effet, dans des dimensions modérées, la distance de Wasserstein peut fournir un test statistique convergent contre toute alternative fixe, par exemple González-Delgado et al. (2021) propose un test convergent basé sur la distance de Wasserstein sur le tore 2-dimensionnel. Ces applications, où la distance de Wasserstein quantifie la similarité entre différents échantillons de données, nécessitent une justification mathématique rigoureuse.

La première partie de cette thèse se concentre sur l'étude statistique asymptotique du problème de transport. Nous verrons que la dimension de l'espace latent affecte l'erreur de l'approximation empirique du coût de transport optimal, mais pas sa variance. En fait, la limite des fluctuations (différence entre la distance de coût de transport optimal empirique et sa moyenne) est gaussienne. L'influence de la dimension sur le taux de convergence de la version empirique à la version de la population est connue sous le nom de *malédiction de la dimension*. Les premières démonstrations rigoureuses de ce fait remontent à 1969 avec le travail de Dudley (1969). Récemment, le travail de Fournier and Guillin (2013) et Weed and Bach (2019) confirme le fait que la distance de Wasserstein empirique entre deux distributions continues converge vers la population avec un taux $n^{-\frac{1}{d}}$, modulo d'éventuels facteurs logarithmiques. Cela prouve que la distance de Wasserstein ne peut pas être utilisée pour fournir un test de conformité consistant en dimension générale. Les fluctuations ont un comportement asymptotique différent. De plus, elles peuvent être bornées avec un taux $n^{-\frac{1}{2}}$, indépendamment de la dimension (Weed and Bach, 2019, proposition 20). Les arguments de Weed and Bach (2019), basés sur l'inégalité de McDiarmid, n'ont pas de sens pour les mesures de probabilité avec des supports non bornés. Une étude plus précise des fluctuations est celle effectuée par del Barrio and Loubes (2019) à travers l'inégalité Efron-Stein. Cependant, la plupart des arguments sont spécifiques au coût quadratique —dans lequel les applications de transport sont des gradients de fonctions convexes— ce qui rend leur généralisation à d'autres coûts non triviale.

Le fléau de la dimension mentionné précédemment apparaît pour les probabilités avec densité. Lorsque les deux probabilités sont discrètes, le problème d'optimisation devient paramétrique et satisfait un théorème limite central (TLC) avec un taux $n^{-\frac{1}{2}}$, comme démontré par Sommerfeld and Munk (2018). Dans le cas où l'une des probabilités est discrète (c'est-à-dire le cas dit semi-discret), del Barrio and Loubes (2019) a prouvé que le TLC est centré sur la valeur de la population en utilisant les résultats obtenus pour les fluctuations. Cependant, cette stratégie peut ne pas être la plus appropriée, comme nous le verrons dans cette étude. La version fonctionnelle de la méthode delta fournit une méthodologie qui nécessite moins d'hypothèses, comme observé parallèlement par Hundrieser et al. (2022).

La maîtrise des fluctuations est plus utile pour analyser le problème de transport régularisé par l'entropie, proposé par Cuturi (2013). Il s'agit indéniablement de la méthode la plus utilisée pour régulariser le problème de transport. Mena and Niles-Weed (2019) ont prouvé, en utilisant les arguments de del Barrio and Loubes (2019), que les fluctuations du problème régularisé sont asymptotiquement gaussiennes. Dans cette étude, nous verrons également que le biais converge plus rapidement que la variance, ce qui aboutit au TLC pour le coût de transport régularisé par l'entropie. Cependant, cela n'est pas suffisant pour donner une réponse positive au problème de l'ajustement basé sur les notions de transport. Le terme supplémentaire comme pénalité provoque un phénomène connu sous le nom de biais entropique, qui réduit significativement l'intérêt du coût de transport régularisé pour l'inférence statistique. Feydy et al. (2019) propose une modification du coût de transport régularisé, appelé divergence de Sinkhorn, qui répare ce problème. Cette thèse fournit un développement de second ordre de la divergence Sinkhorn par rapport au processus empirique, ce qui permet une caractérisation précise de son comportement asymptotique. Cela peut potentiellement être utilisé pour dériver une statistique d'ajustement basée sur la divergence de Sinkhorn.

Il est connu que les rangs, basés sur la notion d'ordre univarié, fournissent une méthodologie générale pour résoudre le problème de l'ajustement. La généralisation multivariée du concept de rang n'a pas été possible en raison de l'absence d'une notion de fonction de distribution multivariée. Récemment, les travaux de Marc Hallin et ses co-auteurs (par exemple Hallin et al. (2021a)) proposent l'application du transport entre l'échantillon de données et la distribution sphérique uniforme comme candidat pour une fonction de distribution multivariée. Il s'avère que, toujours selon Hallin et al. (2021a), cette proposition satisfait les propriétés principales qui font de la fonction de distribution univariée un outil utile pour l'inférence statistique. Cela a fourni, comme mentionné pour le cas univarié, une méthodologie générale pour créer des tests d'ajustement (Deb and Sen (2019); Deb et al. (2021)) ou des tests d'indépendance (Shi et al. (2022); Hallin and Mordant (2021)). La deuxième partie de cette thèse prouve la continuité de la fonction de distribution multivariée pour les probabilités supportées sur un ensemble convexe, étendant ainsi le résultat de Figalli (2018). La singularité à l'origine des coordonnées de la probabilité uniforme sphérique signifie que les résultats de Caffarelli (voir par exemple Caffarelli (1990, 1991, 1992)) ne s'appliquent pas dans ce cas. Pour conclure, nous fournirons, en utilisant le nouveau concept de fonction de distribution multivariée, une méthodologie innovante pour l'étude du problème de régression non paramétrique à sortie multivariée. Nous verrons que cette proposition est, à ce jour, la seule qui maintient la propriété fondamentale de la régression des quantiles, le contrôle probabiliste des régions des quantiles. Cette propriété, dans le cas univarié, stipule que la région de quantile d'ordre $0 < r < 1$ contient une proportion de r points de l'échantillon.

Les deux parties qui composent ce travail sont clairement délimitées. Selon moi, le lien se trouve dans l'aspect mathématique, dans le raisonnement et les outils utilisés, et non dans l'objet d'étude lui-même. Avant de présenter en détail les résultats obtenus, avec les définitions techniques et les outils utilisés, il convient de souligner le fait que chaque chapitre est autosuffisant et adapté de sa version en ligne (ou publiée). En d'autres termes, chacun d'entre eux peut être lu et analysé séparément. Ceux qui ne sont pas intéressés par les liens entre les différents chapitres peuvent sauter le reste de l'introduction.

1.0.1 Le transport optimal en quelques mots

Depuis la fin du siècle dernier, le problème de l'affectation de Monge est devenu un sujet de recherche important en statistique et en probabilité, avec des applications à l'apprentissage automatique, à l'économie, à la physique et à l'astronomie, pour n'en citer que quelques-unes. Soit $c : \mathbb{R}^d \times \mathbb{R}^d \to [0, +\infty)$ une fonction continue, le *coût de transport optimal* entre deux mesures de probabilité $P, Q \in \mathcal{P}(\mathbb{R}^d)$, pour le coût c, est défini comme la solution du problème de Monge

$$\inf_{T:T_{\#}P=Q} \int_{\mathbb{R}^d} c(\mathbf{x}, T(\mathbf{x}))dP(\mathbf{x}), \tag{1.1}$$

où la notation $T_{\#}P$ représente la mesure *push-forward*, c'est-à-dire la mesure telle que $T_{\#}P(A) := P(T^{-1}(A))$, pour chaque ensemble mesurable A. Il a fallu attendre les années 1990, avec les travaux parallèles de Brenier (1991) et Cuesta and Matrán (1989), pour prouver l'existence de la solution du coût quadratique $((c(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2)$, voir aussi Gangbo and McCann (1996) pour des coûts plus généraux et Villani (2003) pour une étude complète.

Entre-temps, Kantorovich a formulé en 1942 (voir Kantorovich (2006) pour une traduction anglaise de l'article original) la célèbre relaxation du problème de Monge;

$$\mathcal{T}_c(P, Q) = \inf_{\pi \in \Pi(P,Q)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(\mathbf{x}, \mathbf{y})d\pi(\mathbf{x}, \mathbf{y}), \tag{1.2}$$

où $\Pi(P, Q)$ est l'ensemble des mesures de probabilité $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ telles que $\pi(A \times \mathbb{R}^d) = P(A)$ y $\pi(\mathbb{R}^d \times B) = Q(B)$, pour tout ensemble mesurable A, B. Une mesure de probabilité $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ est dite être un *plan de transport optimal pour un coût* c entre P et Q si elle est un minimiseur dans (1.2).

Le principal avantage de la relaxation de Kantorovich est l'existence de solutions optimales : pour (1.2), elle ne requiert que l'intégrabilité du coût (Théorème 4.1 dans Villani (2008)) tandis que le problème de Monge nécessite certaines hypothèses sur l'une des deux probabilités—absolument la continuité par rapport à la mesure de Lebesgue–et sur le coût–les conditions bien connues de Gangbo and McCann (1996) pour l'existence de applications de transport optimales, à savoir $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$, où $h : \mathbb{R}^d \to [0, \infty)$ est une fonction non négative satisfaisant

(A1) h est strictement convexe sur $\mathbb{R}^d$,

(A2) étant donné une hauteur $r \in \mathbb{R}^+$ et un angle $\theta \in (0, \pi)$, il existe un certain $M := M(r, \theta) > 0$ tel que pour tous les $|\mathbf{p}| > M$, on puisse trouver un cône

$$K(r, \theta, \mathbf{z}, \mathbf{p}) := \left\{ \mathbf{x} \in \mathbb{R}^d : |\mathbf{x} - \mathbf{p}||\mathbf{z}| \cos(\theta/2) \le \langle \mathbf{z}, \mathbf{x} - \mathbf{p} \rangle \le r|\mathbf{z}| \right\},$$

dont le sommet est en $\mathbf{p}$ (et $\mathbf{z}$ $dans\mathbb{R}^d \setminus \{\mathbf{0}\}$) sur laquelle h atteint son maximum en $\mathbf{p}$,

(A3) $\lim_{|\mathbf{x}| \to \infty} \frac{h(\mathbf{x})}{|\mathbf{x}|} = \infty$.

Le problème du transport admet une formulation duale :

$$\mathcal{T}_c(P, Q) = \sup_{(f,g) \in \Phi_c(P,Q)} \int f(\mathbf{x})dP(\mathbf{x}) + \int g(\mathbf{y})dQ(\mathbf{y}), \tag{1.3}$$

où $\Phi_c(P, Q) = \{(f, g) \in L_1(P) \times L_1(Q) : f(\mathbf{x}) + g(\mathbf{y}) \leq c(\mathbf{x}, \mathbf{y})\}$, voir le théorème 5.10 inVillani (2008). On appelle ψ *dans*$L_1(P)$ *un potentiel de transport optimal de* P *à* Q *pour un coût* c s'il existe φ *dans*$L_1(Q)$ tel que le couple (ψ, φ) résout (1.3). De manière surprenante, l'équivalence entre (1.1) et (1.2) passe par la régularité des potentiels, qui elle-même découle de celle du coût (cf. Gangbo and McCann (1996) eg.). En ce qui concerne le contenu traité dans cette thèse, la formulation duale sera utilisée : dans les chapitres 2 et 3, la variance des bornes des fluctuations; et dans le Chapitre 3 les prix donnés par une entreprise à certains produits.

Chaque formulation a un intérêt différent, et son utilisation dépend de chaque application. D'une part, lorsqu'elle existe, la solution de (1.1) définit une application de transport entre mesures de probabilité. Cela permet d'inférer les propriétés d'une mesure de probabilité à travers une autre déjà connue (ou une mesure de référence). C'est l'idée derrière le succès de la fonction quantile multivariée de M. Hallin (Hallin et al. (2021a)); de la régression quantile multivariée (Chapitre 7); des explications contrefactuelles basées sur le transport de masse (de Lara et al., 2021) ou de la réparation des biais (Gordaliza et al., 2019b; Black et al., 2020).

Soit $\mathcal{P}_p(\mathbb{R}^d)$ l'espace des probabilités dans $\mathbb{R}^d$ avec des moments finis d'ordre $p \geq 1$. D'autre part, lorsqu'on considère un coût potentiel ($c_p(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^p$, pour $p \geq 1$), la formulation de Kantorovich entre probabilités à moments finis d'ordre $p \geq 1$ admet toujours une solution et la fonction

$$\mathcal{P}_p(\mathbb{R}^d) \times \mathcal{P}_p(\mathbb{R}^d) \ni (P, Q) \to \mathcal{W}_p(P, Q) = (\mathcal{T}_p(P, Q))^{\frac{1}{p}} = \left(\mathcal{T}_{c_p}(P, Q)\right)^{\frac{1}{p}}$$

définit une distance $\mathcal{P}_p(\mathbb{R}^d)$ caractérisée par :

$$\mathcal{W}_p(\mu_n, \mu) \to 0 \quad \iff \quad \mu_n \xrightarrow{w} \mu \quad \text{y} \quad \int \|\mathbf{x}\|^p d\mu_n(\mathbf{x}) \to \int \|\mathbf{x}\|^p d\mu(\mathbf{x}),$$

voir (Villani, 2003, Chapitre 7). Il s'agit de la distance dite *"Wasserstein"* ou de *"Monge-Kantorovich"*. Elle est largement utilisée comme discriminateur dans les réseaux adversaires génératifs Arjovsky et al. (2017), dans le recalage diphéomorphique ou comme terme de pénalité dans la réparation algorithmique des biais (Risser et al., 2021).

1.1 Comportement asymptotique

La probabilité de vraie population P n'est généralement pas connue pour les chercheurs; ce qu'ils observent est un échantillon i.i.d. $\mathbf{X}_1, \ldots, \mathbf{X}_n$ de taille n de P, définissant la mesure empirique P_n. Supposons pour cette introduction que Q soit connu. Cependant, les résultats exposés ici restent valables dans le cas à deux échantillons. La valeur $\mathcal{T}_c(P_n, Q)$ est donc la contrepartie empirique du vraie valeur $\mathcal{T}_c(P, Q)$. Bien sûr, $\mathcal{T}_c(P_n, Q)$ tend vers $\mathcal{T}_c(P, Q)$, mais à quel rythme ? Autrement dit, si le nombre de données dont nous disposons n tend vers l'infini, comment varie la différence $\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q)$? Pour donner une réponse rapide, nous pouvons faire une décomposition biais-variance de l'erreur :

$$\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q) = (\mathbb{E}(\mathcal{T}_c(P_n, Q)) - \mathcal{T}_c(P, Q)) + (\mathcal{T}_c(P_n, Q) - \mathbb{E}(\mathcal{T}_c(P_n, Q))) \,. \tag{1.4}$$

Avant d'analyser formellement le comportement asymptotique du biais $(\mathbb{E}(\mathcal{T}_c(P_n, Q_n)) - \mathcal{T}_c(P, Q))$ et de la variance $(\mathcal{T}_c(P_n, Q) - \mathbb{E}(\mathcal{T}_c(P_n, Q)))$, une simulation peut être utile. Pour ce faire, nous fixons une dimension 'relativement élevée" (par exemple $d = 10$) et simulons $\sqrt{n}(\mathcal{T}_2(P_n, Q_n))$, pour $P, Q \sim \mathcal{U}_{[0,1]^{10}}$. La figure 1.1 montre, avec différentes couleurs et en répétant 10 fois indépendamment la procédure, les différentes valeurs de $\sqrt{n}(\mathcal{T}_2(P_n, Q_n))$ sur l'axe des ordonnées et la taille de l'échantillon n sur l'axe des abscisses.

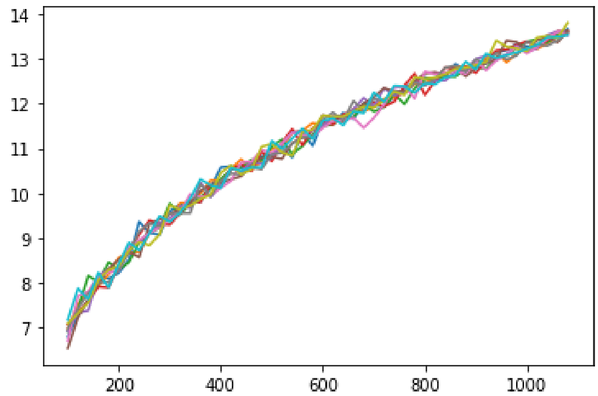

Figure 1.1: Axe des Y, taille de l'échantillon n ; axe des y, différence $\sqrt{n}(\mathcal{T}_2(P_n, Q_n))$, pour $P, Q \sim \mathcal{U}_{[0,1]^{10}}$. L'expérience est répétée 10 fois de manière indépendante et les résultats interpolés sont représentés par des couleurs différentes.

L'observation heuristique permet d'éclairer ce phénomène : le taux de convergence du biais ne peut pas être le taux paramétrique $\sqrt{n}$, mais $\sqrt{n}(\mathcal{T}_2(P_n, Q_n) - \mathbb{E}\mathcal{T}_2(P_n, Q_n))$ semble borné. Il semble donc que les fluctuations ne sont pas influencées par le fait d'avoir une dimension considérablement élevée ($d = 10$). En effet, pour des mesures absolument continues par rapport à la mesure de Lebesgue à d dimensions ℓ_d, le taux de convergence correct du biais dépend de la dimension de l'espace Weed and Bach (2019).

1.1.1 L'analyse des fluctuation

Pour l'analyse des fluctuations (variance), nous supposons que le coût $c(\mathbf{x} - \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$ satisfait aux conditions de Gangbo and McCann (1996), que P et Q sont absolument continus par rapport à la mesure de Lebesgue et que

$$\int h(2\mathbf{x})^2 dP(\mathbf{x}) < \infty \text{ and } \int h(-2\mathbf{y})^2 dQ(\mathbf{y}) < \infty.$$

Le chapitre 2 montre que

$$\sqrt{n}\,(\mathcal{T}_c(P_n, Q) - \mathbb{E}\mathcal{T}_c(P_n, Q)) \xrightarrow{w} N(0, \sigma_c^2(P, Q)), \tag{1.5}$$

où

$$\sigma_c^2(P,Q) := \int \varphi(\mathbf{x})^2 dP(\mathbf{x}) - \left(\int \varphi(\mathbf{x}) dP(\mathbf{x})\right)^2, \tag{1.6}$$

φ étant un potentiel de transport optimal de P à Q pour le coût c. Dans le cas où $P = Q$, le potentiel de transport optimal est constant, de sorte que sa variance est nulle et que la limite (1.5) est dégénérée. Pour en revenir à l'exemple simulé ci-dessus (Figure 1.1), cela implique que les fluctuations sont non seulement bornées, mais tendent en probabilité vers 0. Le taux de convergence de la fluctuation sous l'hypothèse $P = Q$ est encore un problème ouvert. González-Delgado et al. (2021) prouve que la limite (1.5) reste valide sur le tore plat de dimension d, $\mathbb{T}^d$.

La preuve est principalement basée sur l'inégalité d'Efron-Stein (Boucheron et al., 2013, Chapitre 3.1), c'est-à-dire que si $(\mathbf{X}_1', \ldots, \mathbf{X}_n')$ et $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ sont i.i.d, on définit $Z := f(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ et pour chaque $i \in \{1, \ldots, n\}$ on désigne

$$Z_i' := f(\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}, \mathbf{X}_i', \mathbf{X}_{i+1}, \ldots, \mathbf{X}_n),$$

alors

$$\mathit{Var}(Z) \leq \frac{n}{2} E(Z - Z_i')^2 = nE(Z - Z_i')_+^2.$$

où $(\cdot)_+$ désigne la partie positive. L'inégalité d'Efron-Stein donne donc

$$n\mathit{Var}(\mathcal{T}_c(P_n, Q) - \int \varphi(\mathbf{x}) dP_n(\mathbf{x})) \leq \mathbb{E}\big(\varphi_n(X_1) - \varphi(X_1) - \varphi_n(X_1') + \varphi(X_1')\big)^2 \tag{1.7}$$

ce qui signifie que $\sqrt{n}\,(\mathcal{T}_c(P_n, Q) - \mathbb{E}\mathcal{T}_c(P_n, Q))$ est une suite tendue. Le théorème central limite découlera de la stabilité des potentiels—Corollaire 2.2.7 donne l'unicité (jusqu'à la constante additive) des potentiels pour les probabilités à support connecté et le Théorème 2.3.4 la stabilité de la application de transport. Nous extrayons ici le contenu du Théorème 2.3.4 en raison de son importance— c'est le premier résultat montrant la stabilité de la application de transport et du potentiel optimal dans des domaines non bornés et pour des coûts généraux.

Theorem 1.1.1. *Soit $Q \in \mathcal{P}(\mathbb{R}^d)$ tel que $Q \ll \ell_d$ a un support connecté avec une frontière négligeable. Supposons que $Q_n, P_n, P \in \mathcal{P}(\mathbb{R}^d)$ sont tels que $P_n \xrightarrow{w} P$, $Q_n \xrightarrow{w} Q$ et*

$$\mathcal{T}_c(P_n, Q_n) < \infty \ \text{ and } \ \mathcal{T}_c(P, Q) < \infty$$

pour un coût $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$, avec h différentiable et satisfaisant (A1)-(A3). *Si ψ_n (resp. ψ) sont des potentiels de transport optimaux de Q_n à P_n (resp. de Q à P) pour le coût c. Alors :*

(i) Il existe des constantes $a_n \in \mathbb{R}$ telles que $\tilde{\psi}_n := \psi_n - a_n \to \psi$ au sens de la convergence uniforme sur les ensembles compacts de $\mathrm{Supp}(Q)$.

(ii) Pour chaque K sous-ensemble compact $\mathrm{Supp}(Q) \cap \mathrm{dom}(\nabla\psi)$

$$\sup_{\mathbf{x}\in K} \sup_{\mathbf{y}_n \in \partial^c \psi_n(\mathbf{x})} |\mathbf{y}_n - \nabla^c \psi(\mathbf{x})| \longrightarrow 0.$$

Cependant, ce résultat ne donne que la convergence p.s. des potentiels empiriques (1.7) vers leur contrepartie théorique, mais on ne peut pas en conclure que (1.7) tend vers 0. Pour éviter d'autres hypothèses (comme dans del Barrio and Loubes (2019)), on peut montrer que la séquence $\mathbb{E}\big|\varphi_n(X_1) - \varphi(X_1) - \varphi_n(X_1') + \varphi(X_1')\big|$ est bornée, puis le théorème de Banach-Alaoglu donne une convergence faible dans $L^2(\mathbb{P})$ de $\big|\varphi_n(X_1) - \varphi(X_1) - \varphi_n(X_1') + \varphi(X_1')\big|$ le long des sous-séquences. En prenant la moyenne de Cesàro signifie que nous pouvons passer d'une convergence faible à une convergence forte.

1.2 Cas semi-discret

Le comportement du biais est très différent dans le cas semi-discret — lorsque l'une des deux probabilités est supportée pour un ensemble fini de points $\mathbb{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$. Ici, dans le chapitre 3 nous obtenons

$$\sqrt{n}\,(\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q)) \xrightarrow{w} \sup_{\mathbf{z}\in \mathrm{Opt}_c(P,Q)} \mathbb{G}(\mathbf{z}),$$

où $\mathrm{Opt}_c(P, Q)$ est l'ensemble des potentiels de transport optimaux et $\mathbb{G}$ un processus gaussien agissant sur eux. Il est alors évident que l'unicité á des constantes additives près de ceux-ci implique que la limite est gaussienne. De plus, sous cette hypothèse et grâce aux travaux de Cárcamo et al. (2020), nous savons que le bootstrap est cohérent comme approximation du processus limite. Le fait que la fléau de la dimension ne semble pas affecter le cas semi-discret pour les deux probabilités est assez étonnant. Mais il est partiellement caché dans l'hypothèse que le support a une taille finie fixe. Pour une meilleure compréhension, nous fournissons la limite supérieure

$$E\,|\mathcal{W}_1(P, Q_m) - \mathcal{W}_1(P, Q)| \le \frac{8\sqrt{2N}}{\sqrt{m}} K(\mathrm{diam}(\mathbb{X}), Q)$$

où

$$K(\mathrm{diam}(\mathbb{X}), Q) = (4\,\mathrm{diam}(\mathbb{X}) + 2\sqrt{\int |\mathbf{y}|^2 dQ(\mathbf{y})} + 2\,\mathrm{diam}(\mathbb{X}))\left(\log(2) + \sqrt{2\,\mathrm{diam}(\mathbb{X}) + 1}\right)$$

et $\mathcal{W}_1$ est la distance de 1-Wasserstein pour la distance euclidienne.

Les preuves des chapitres 2 et 3 sont basées sur des arguments complètement différents; la limite des fluctuations utilise la procédure introduite par del Barrio and Loubes (2019) basée sur l'inégalité d'Efron-Stein, tandis que le problème semi-discret requiert la dérivation fonctionnelle au sens de Hadamard du coût de transport. Le lemme 3.1.1 montre que $\mathcal{T}_c(P,Q)$ est équivalent à la maximisation (sur une classe paramétrique) de

$$g_c(P,Q,\mathbf{z}) = \sum_{i=1}^{N} z_i p_i + \int \inf_{i=1,\dots,N} \{c(\mathbf{x}_i,\mathbf{y}) - z_i\} dQ(\mathbf{y}). \tag{1.8}$$

C'est donc la raison pour laquelle le taux de convergence devient paramétrique. La solution $\mathbf{z}^*$ (unique jusqu'à la constante additive sous certaines hypothèses) de (1.8) définit un ensemble de cellules de Laguerre

$$\mathrm{Lag}_k(\mathbf{z}^*) := \{\mathbf{y} \in \mathbb{R}^d : \ c(\mathbf{x}_k,\mathbf{y}) - z_k < c(\mathbf{x}_i,\mathbf{y}) - z_i^* \text{ for all } i \neq k\}, \quad k = 1,\dots,N,$$

qui sont des généralisations des cellules de Voronoï– équivalentes à $\mathrm{Lag}_k(\mathbf{0})$ pour le coût quadratique.

Le transport optimal semi-discret a des applications au *"modèle de localisation de Hotelling"* (Galichon, 2016, Chapitre 5.1). Cette application est un exemple typique de problème socio-économique où la localisation d'une certaine population est représentée par une probabilité continue Q et les fontaines—des entreprises essayant de vendre un produit– par une probabilité discrète P. Ici, l'emplacement de la fontaine i est $\mathbf{x}_i$ et sa capacité est p_i. Chaque habitant choisirait la fontaine à la fois plus proche et offrant un meilleur prix, c'est-à-dire la stratégie $\arg\inf_{i=1,\dots,N}\{c(\mathbf{x}_i,\mathbf{y}) - z_i\}$, où z_i représente le prix de la fontaine i. L'ensemble de la population qui préfère consommer à la fontaine i est en fait $\mathrm{Lag}_k(\mathbf{z})$. En cas de compensation du marché - l'offre est égale à la demande - chaque fontaine est utilisée à sa pleine capacité et le problème de la détermination des prix se réduit à la solution d'un coût de transport optimal semi-discret..

Le chapitre 3 donne également, à notre connaissance, le premier théorème de limite centrale pour les solutions du problème dual (1.8). Nous soulignons que ce résultat ne peut pas être généralisé pour des distributions continues. En effet, si les deux probabilités sont continues et que l'espace n'est pas unidimensionnel, on ne peut s'attendre à un tel type de limite centrale pour les potentiels, puisque la valeur attendue de l'estimation du coût de transport converge avec un taux $O(n^{-\frac{1}{d}})$ et non plus $O(n^{-\frac{1}{2}})$. Lorsque les deux échantillons sont discrets, même si un tel taux est de $O(n^{-\frac{1}{2}})$, le manque d'unicité du problème dual ne permet pas de prouver ce type de problèmes. En conséquence, le semi-discret est le seul cas où de tels résultats, pour les potentiels du problème O.T. en dimension générale, peuvent être attendus. Le chapitre 3 fournit des limites faibles pour les cellules de Laguerre en termes de métrique L^p (cf. Vitale (1985) eg.) entre les cellules empiriques et de vraie population. De plus, nous donnons également des voisinages de confiance asymptotiques en termes de distance de Hausdorff (métrique L^∞). Dans tous les cas, la vitesse paramétrique est atteint.

1.3 Transport optimal régularisé par l'entropie

Le problème de la recherche des limites faibles du coût de transport est compliqué, même en dimension modérée, et, même si elles étaient trouvées, le rythme serait lent. C'est-à-dire qu'il faudrait trop de données pour approcher la distribution limite. On a donc besoin d'une autre notion d'écart entre les distributions qui, tout en capturant la géométrie de l'espace, a une approximation empirique qui converge vers sa contrepartie de vraie population avec la vitesse paramétrique $\sqrt{n}$.

Avec une motivation computationnelle, l'article influent Cuturi (2013) propose une régularisation de (1.2) avec une pénalité entropique :

$$S_\epsilon(P,Q) = \min_{\pi \in \Pi(P,Q)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \tfrac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 d\pi(\mathbf{x},\mathbf{y}) + \epsilon H(\pi | P \times Q), \tag{1.9}$$

où H désigne l'entropie relative, définie, pour deux mesures de probabilité α et β, comme suit : $H(\alpha|\beta) = \int \log(\frac{d\alpha}{d\beta}(x))d\alpha(x)$ si α est absolument continue par rapport à β et $+\infty$ sinon. Soit $\pi^\epsilon_{P,Q}$ la solution de (1.9), qui est absolument continue par rapport à $P \times Q$ avec la densité $\xi^\epsilon_{P,Q}$ Ce problème, ainsi que son homologue non régularisé (1.2), admet une formulation double.

$$\begin{aligned} S_\epsilon(P,Q) = \sup_{f \in L_1(P), g \in L_1(Q)} \Big\{ & \int_{\mathbb{R}^d} f(\mathbf{x})dP(\mathbf{x}) + \int_{\mathbb{R}^d} g(\mathbf{y})dQ(\mathbf{y}) \\ & - \epsilon \int_{\mathbb{R}^d \times \mathbb{R}^d} e^{\frac{f(\mathbf{x})+g(\mathbf{y})-\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2}{\epsilon}} dP(\mathbf{x})dQ(\mathbf{y}) + \epsilon \Big\} \end{aligned} \tag{1.10}$$

La paire de solutions de (1.10), appelée $(f^\epsilon_{P,Q}, g^\epsilon_{P,Q})$, satisfait les *conditions d'optimalité* suivantes:

$$\begin{aligned} \int e^{f_{P,Q}(\mathbf{x})+g_{P,Q}(\mathbf{y})-\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2} dQ(\mathbf{y}) = 1, \quad \text{for all } x \in \mathbb{R}^d, \\ \int e^{f_{P,Q}(\mathbf{x})+g_{P,Q}(\mathbf{y})-\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2} dP(\mathbf{x}) = 1, \quad \text{for all } y \in \mathbb{R}^d. \end{aligned} \tag{1.11}$$

Cela permet de trouver des solutions pour (1.10) sur la classe $\mathcal{C}^s$, de réduire la complexité de l'optimisation (cf. Genevay et al. (2019)) et *a fortiori* $\sqrt{n}\,\mathbb{E}|S_\epsilon(P_n,Q) - S_\epsilon(P,Q)| \leq C$, valable pour des probabilités sub-gaussiennes Mena and Niles-Weed (2019).

De plus, l'argument de concentration de del Barrio and Loubes (2019) tient également dans ce cas (voir Mena and Niles-Weed (2019)), donnant lieu à

$$\sqrt{n}(S_\epsilon(P_n,Q) - \mathbb{E}S_\epsilon(P_n,Q)) \xrightarrow{w} N(0, \mathrm{Var}_{\mathbf{X}\sim P}(f^\epsilon_{P,Q}(\mathbf{X}))). \tag{1.12}$$

Ni (1.12) ni la limite mentionnée précédemment ne fournissent d'intervalles de confiance asymptotiquement valides pour la valeur de la vraie population $S_\epsilon(P,Q)$). Pour cela, il est nécessaire de connaître la limite exacte de

$$\sqrt{n}\,(\mathbb{E}S_\epsilon(P_n,Q) - S_\epsilon(P,Q)). \tag{1.13}$$

Dans le chapitre 4, il est prouvé que (1.13) tend vers 0 quand $n \to \infty$. Ainsi, la valeur attendue peut être remplacée par la valeur de la vraie population dans (1.12), résultant en

$$\sqrt{n}(S_\epsilon(P_n, Q) - S_\epsilon(P, Q)) \xrightarrow{w} N(0, \mathrm{Var}_{\mathbf{X}\sim P}(f^\epsilon_{P,Q}(\mathbf{X}))). \tag{1.14}$$

Soit z_α le quantile α de la distribution normale standard et

$$\hat{\sigma}_n^2 := \mathrm{Var}_{P_n}(f^\epsilon_{P_n,Q}) = \frac{1}{n}\sum_{i=1}^{n}(f^\epsilon_{P_n,Q}(\mathbf{X}_i))^2 - \left(\frac{1}{n}\sum_{i=1}^{n} f^\epsilon_{P_n,Q}(\mathbf{X}_i)\right)^2 \tag{1.15}$$

soit une approximation cohérente de la variance limite $\mathrm{Var}_{\mathbf{X}\sim P}(f^\epsilon_{P,Q}(\mathbf{X}))$, la limite

$$\left[S_\epsilon(P_n, Q) \pm \frac{\hat{\sigma}_n}{\sqrt{n}} z_{1-\alpha/2}\right], \tag{1.16}$$

est valide.

Néanmoins, (1.16) ne fournit aucune conséquence statistiquement significative en raison du biais entropique. Le fait que $S_\epsilon(P, Q) = 0$ ne signifie pas que les deux probabilités sont égales ! L'hypothèse $P = Q$ ne peut pas être acceptée ou rejetée via la valeur de $S_\epsilon(P, Q)$.

La solution la plus aboutie pour limiter l'influence de la régularisation entropique est la divergence de Sinkhorn, proposée par Feydy et al. (2019) et énoncée comme suit

$$D_\epsilon(P, Q) = S_\epsilon(P, Q) - \tfrac{1}{2}\left(S_\epsilon(P, P) + S_\epsilon(Q, Q)\right).$$

Clairement, $D_\epsilon(P, Q)$ est symétrique en P, Q et $D_\epsilon(P, P) = 0$. De plus, (Feydy et al., 2019, Théorème 1) prouve que $D_\epsilon(P, Q) \geq 0$, avec $D_\epsilon(P, Q) = 0$ si et seulement si $P = Q$.

Les intervalles de confiance asymptotiques pour $D_\epsilon(P, Q)$ peuvent être déduits de leur contre-partie empirique en connaissant les limites non triviales de

$$a_n(D_\epsilon(P_n, Q) - D_\epsilon(P, Q)),$$

pour une certaine suite $\{a_n\}_{n \ dans \mathbb{N}} \subset \mathbb{R}$. Le chapitre 5 prouve que la séquence en question dépend de l'hypothèse. En particulier, on obtient les limites suivantes :

- Sous $H_0 : \ P = Q$:

$$n\, D_\epsilon(P_n, P) \xrightarrow{w} \frac{1}{4}\sum_{j=1}^{\infty}\lambda_j\left(\sum_{i=1}^{\infty} x_{i,j} N_i\right)^2,$$

où $\{N_i\}_{i\in\mathbb{N}}$ est une séquence de variables aléatoires mutuellement indépendantes avec $N_i \sim N(0, 1)$ et $\{x_{i,j}\}_{i,j\in\mathbb{N}} \subset \mathbb{R}$, $\{\lambda_j\}_{j\in\mathbb{N}} \subset [0, \infty)$ séquences réelles (et déterministes) dépendant de P y ϵ.

- Sous H_1 : $P \neq Q$:

$$\sqrt{n}(D_\epsilon(P_n, Q) - D_\epsilon(P, Q)) \xrightarrow{w} N(0, \mathrm{Var}_P(\psi^\epsilon_{P,Q})),$$

  où $\psi^\epsilon_{P,Q} = f^\epsilon_{P,Q} - \frac{1}{2}(f^\epsilon_{P,P} + g^\epsilon_{P,P})$.

La technique est différente pour chaque hypothèse. Sous H_1, l'inégalité d'Efron-Stein fournit le développement au premier ordre. Cependant, puisque $\psi^\epsilon_{P,Q}$ est nul lorsque $P = Q$, cela n'est pas suffisant pour obtenir des limites non triviales sous H_0. Il faut donc un développement du second ordre, qui est une conséquence du développement du premier ordre des potentiels. La preuve est longue et fastidieuse; elle est divisée en deux ouvrages différents. D'une part, le chapitre 4 montre l'étanchéité des potentiels et la divergence. Ce qui signifie que il existe une constante c_d, dépendant uniquement de d, telle que

$$\mathbb{E}\|g_{P_n,Q} - g_{P,Q}\|^2_{\mathcal{C}^s(\Omega)},\ \mathbb{E}\|f_{P_n,Q} - f_{P,Q}\|^2_{\mathcal{C}^s(\Omega)} \leq \tfrac{c_d}{n} D_\Omega^{5(d+1)} e^{15 D_\Omega^2},$$

et

$$\mathbb{E} D_1(P_n, P) \leq \frac{c_d}{n} D_\Omega^{\frac{3d}{2}+1} \frac{32}{(d+1)^2} e^{\frac{19}{2} D_\Omega^2}.$$

D'une part, le chapitre 5 montre le développement de premier ordre de $\sqrt{n}\begin{pmatrix} f^\epsilon_{P_n,Q} - f^\epsilon_{P,Q} \\ g^\epsilon_{P_n,Q} - g^\epsilon_{P,Q} \end{pmatrix}$, in $\mathcal{C}^\alpha(\Omega) \times \mathcal{C}^\alpha(\Omega)$, par rapport au processus empirique $\sqrt{n}(P_n - P)$, qui suit

$$\begin{pmatrix} f^\epsilon_{P_n,Q} - f^\epsilon_{P,Q} \\ g^\epsilon_{P_n,Q} - g^\epsilon_{P,Q} \end{pmatrix} = \begin{pmatrix} (1 - \mathcal{A}^\epsilon_Q \mathcal{A}^\epsilon_P)^{-1} \mathcal{A}^\epsilon_Q \mathbb{G}^n_{P,s} \\ -(1 - \mathcal{A}^\epsilon_P \mathcal{A}^\epsilon_Q)^{-1} \mathbb{G}^n_{P,s} \end{pmatrix} + o_P\left(\frac{1}{\sqrt{n}}\right).$$

Ici $\mathbb{G}^n_{P,s}$ désignent $\frac{1}{n}\sum_{k=1}^n \xi^\epsilon_{P,Q}(\mathbf{X}_k, \cdot) - \mathbb{E}\left(\xi^\epsilon_{P,Q}(\mathbf{X}, \cdot)\right)$, et

$$\begin{aligned} \mathcal{A}^\epsilon_P &: L^2(P) \ni f \mapsto \int \xi^\epsilon_{P,Q}(\mathbf{x}, \cdot) f(\mathbf{x}) dP(\mathbf{x}) \in \mathcal{C}^\alpha(\Omega), \\ \mathcal{A}^\epsilon_Q &: L^2_0(Q) \ni g \mapsto \int \xi^\epsilon_{P,Q}(\cdot, \mathbf{y}) g(\mathbf{y}) dQ(\mathbf{y}) \in \mathcal{C}^\alpha(\Omega), \end{aligned} \tag{1.17}$$

Ce développement au premier ordre des potentiels permet, par la formule

$$d\pi^\epsilon_{P,Q} = e^{\frac{f_{P,Q} + g_{P,Q} - \frac{1}{2}\|\cdot - \cdot\|^2}{\epsilon}} dP dQ,$$

pour obtenir les limites faibles du problème primaire, c'est-à-dire

$$\sqrt{n} \int \eta\, (d\pi^\epsilon_{P_n,Q} - d\pi^\epsilon_{P,Q}), \quad \text{with } \eta \in L^2(P \times Q).$$

En particulier, nous obtenons

$$\sqrt{n}\left(\int \eta d\pi^\epsilon_{P_n,Q} - \int \eta d\pi^\epsilon_{P,Q}\right) \xrightarrow{w} N(0, \sigma^2_{\lambda,\epsilon}(\eta)), \quad \eta \in L^2(P \times Q), \tag{1.18}$$

où $\sigma^2_\lambda(\eta) = \mathrm{Var}_{\mathbf{X}\sim P}\Big((1 - \mathcal{A}^\epsilon_Q \mathcal{A}^\epsilon_P)^{-1}(\eta^\epsilon_{\mathbf{x}} - \mathcal{A}^\epsilon_Q \eta^\epsilon_{\mathbf{y}})(\mathbf{X})\Big)$, prouvant ainsi une conjecture de Harchaoui et al. (2020). (1.18) fournit des intervalles de confiance consistants pour $\pi^\epsilon_{P_n,Q}$, ce qui permet, entre autres, d'effectuer une inférence sur la distance de Sinkhorn (appliquer (1.18) à la fonction $(\mathbf{x}, \mathbf{y}) \to \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2$) définie dans Cuturi (2013) o the regularized colocalization measure RCol (appliquer (1.18) à la fonction $(\mathbf{x}, \mathbf{y}) \to \mathbb{1}_{\|\mathbf{x}-\mathbf{y}\|^2 \leq t}$) définie dans Klatt et al. (2020).

1.4 Fonction de distribution multivariée centre-extérieur; régularité et régression quantile.

L'absence d'ordre canonique de $\mathbb{R}^d$, pour $d \geq 2$, empêche la généralisation d'outils statistiques basés sur la relation d'ordre univarié. L'un des plus importants est la fonction de distribution d'une variable aléatoire univariée $X \sim P$, définie comme $F : x \to \mathbb{P}(X \leq x)$.

De nombreuses tentatives ont été faites pour définir une fonction de distribution multivariée. Parmi celles-ci, citons celles basées sur les profondeurs, les copules, les rangs par composante, les rangs spatiaux ou les rangs de Mahalanobis; nous nous référons à Hallin et al. (2021a) et à ses références. Toutes ces méthodes doivent tenir compte de l'absence d'ordre canonique, mais aucune d'entre elles n'est capable d'imiter les propriétés qui rendent la fonction de distribution univariée utile. Il s'agit de la liberté de distribution (la distribution de $F(X)$ est uniforme sur $[0, 1]$, tandis que sa contrepartie empirique $\{F^n(X_i)\}_{i=1}^n$ est uniforme sur l'ensemble $\{\frac{i}{n}\}_{i=1}^n$, quelle que soit la distribution des X_i); de la propriété de Glivenko-Cantelli (l'approximation empirique converge a. s. uniformément dans $\mathbb{R}$) et maintient l'ordre, c'est-à-dire qu'elle est monotone.

La monotonicité et le fait que $F(X)$ est uniforme sur $[0, 1]$ définissent de manière unique F, pour toute variable aléatoire continue $X \sim P$. La monotonicité équivaut à être le gradient d'une fonction convexe, la gratuité de la distribution à F faisant avancer P vers $U_{(0,1)}$ (la distribution uniforme sur $(0, 1)$). Bien sûr, la fonction quantile Q est le gradient d'une fonction convexe poussant en avant $U_{(0,1)}$ vers P. Ces propriétés ont conduit Chernozhukov et al. (2017) à définir une fonction quantile multivariée candidate, dite fonction de distribution quantile par rapport à une mesure de référence $\mathcal{U}$, définie, pour une variable aléatoire $\mathbf{X} \sim P \in \mathcal{P}(\mathbb{R}^d)$, comme l'unique gradient d'une fonction convexe (définie en presque tout point) qui fait avancer $\mathcal{U}$ vers P. Son existence est donnée en vue du célèbre théorème de McCann McCann (1995).

Le débat actuel porte sur le choix d'une *idoneus* mesure de référence $\mathcal{U}$. Deb and Sen (2019) plaide pour une mesure uniforme sur l'hypercube (réussissant à tester l'indépendance), Hallin et al. (2021a) pour une mesure uniforme sphérique (réussissant à donner une notion d'ordre centre-extérieur) et Deb et al. (2021) pour une gaussienne standard (réussissant le test T^2 de Hotelling). La réponse la plus correcte est que cela dépend du but recherché : (Hallin and Mordant, 2021). Nous traiterons ici le cas de la distribution sphérique sur la boule unité $\mathbf{U}_d$, c'est-à-dire celle obtenue en prenant indépendamment un rayon $r \sim U_{(0,1)}$

et une direction θ uniforme sur la sphère $\mathcal{S}^{d-1}$. L'avantage d'utiliser l'uniforme sphérique comme mesure de référence réside dans son invariance aux rotations (changements de base) et son contrôle uniforme des boules centrées sur l'origine et de rayon $\tau \in (0,1)$, soit

$$\mathbf{U}_d(\tau\mathbb{S}^d) = \tau \quad \tau \in (0,1). \tag{1.19}$$

Aucune autre mesure de probabilité dans $\mathbb{R}^d$ ne satisfait à ces propriétés. (1.19) fournit un ordonnancement radial naturel et une notion claire de centre—donc un ordonnancement centre-extérieur. La fonction quantile $\mathbf{Q}_\pm$ est définie dans ce cas comme l'unique gradient d'une fonction convexe ψ (définie en presque chaque point) qui fait avancer $\mathbf{U}_d$ vers P. La fonction de distribution est définie comme

$$\mathbf{F}_\pm(\mathbf{x}) := \arg \sup_{\|\mathbf{u}\|\leq 1} \{\langle \mathbf{x}, \mathbf{u}\rangle - \psi(\mathbf{u})\},$$

qui coïncide avec l'inverse de $\mathbf{Q}_\pm$ dans le support de P. L'importance accordée par $\mathbf{U}_d$ à l'origine et cette notion d'ordre centre-extérieur sont héritées par P via $\mathbf{Q}_\pm$. Cela donne lieu au nom de *distribution centre-extérieur et mappings de quantiles* inventé dans Hallin et al. (2021a) pour $\mathbf{F}_\pm$ et $\mathbf{Q}_\pm$. Cette terminologie sera utilisée à partir de maintenant.

Supposons l'observation d'un simple échantillon i.i.d. échantillon $\mathbf{X}^{[n]} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ de P, on définit sa distribution empirique centre-extérieur $\mathbf{F}^n_\pm$, pour chaque $\mathbf{X}_i$ comme $\mathbf{F}^n_\pm(\mathbf{X_i}) = \mathfrak{G}^{(n)}_{\sigma_n(i)}$. Ici, σ_n est la correspondance biunivoque dans $\{1, \ldots, n\}$ où

$$\sum_{i=1}^{n} \|\mathbf{X}_i - \mathfrak{G}^{(n)}_{\sigma(i)}\|^2 \tag{1.20}$$

atteint son minimum, et $\mathfrak{G}^{(n)}_1, \ldots, \mathfrak{G}^{(n)}_n$ est une grille régulière de la sphérique uniforme de $\mathbb{U}_d$. Elle consiste à factoriser $n = n_s n_d + n_0$, avec $n < \min(n_s, n_d)$ et à calculer n points $\mathfrak{G}^{(n)}_1, \ldots, \mathfrak{G}^{(n)}_n$ créés par l'intersection entre

- les rayons générés par un couple $\mathbf{u}_1, \ldots, \mathbf{u}_{n_S} \in \mathcal{S}_{d-1}$ de vecteurs unitaires tels que $n_S{}^{-1} \sum_{j=1}^{n_S} \delta_{\mathbf{u}_j}$ converge faiblement vers l'uniforme sur $\mathcal{S}_{d-1}$ comme $n_S \to \infty$, et
- les hypersphères n_R de centre $\mathbf{0}$ et de rayon $j/(n_R+1)$, $j = 1, \ldots, n_R$,

les points correspondant à n_0 sont identifiés à $\mathbf{0}$. La propriété de Glivenko-Cantelli

$$\max_{1\leq k\leq n} \|\mathbf{F}^n_\pm(\mathbf{X}_k) - \mathbf{F}_\pm(\mathbf{X}_k)\| \xrightarrow{p.s.} 0 \tag{1.21}$$

est valide si $\mathbf{F}_\pm$ est bien défini sur l'espace entier (rappelons que le théorème de McCann donne une définition de Lebesgue p.s.). Ceci est assez courant lorsqu'il s'agit de prouver la convergence des gradients de fonctions convexes. En particulier, on suit généralement trois étapes : prouver la convergence faible des plans de transport, prouver la convergence des sous-différentielles au sens des ensembles, et enfin, prouver la convergence des potentiels. Sur les ensembles compacts où le sous-différentiel est monovalent (i.e. où la fonction est lisse) la limite est uniforme, la forme particulière (fortement convexe) du support de $\mathbb{U}_d$, nous permet de passer de la convergence uniforme dans les ensembles compacts à la convergence uniforme (globale).

Si $\mathbf{F}_\pm$ est bien défini sur l'espace entier, en conséquence de (1.19), la application du quantile centre-extérieur définit des régions imbriquées $\mathbb{C}_P(\tau) = \mathbf{Q}_\pm(r\,\mathbb{S}^d)$ et des contours continus $\mathcal{C}_P(\tau) = \mathbf{Q}_\pm(r\,\mathcal{S}^{d-1})$ indexés par $\tau \in ([0,1)$, tels que, pour tout P absolument continu par rapport à la mesure de Lebesgue, $P\big[\mathbb{C}_P(\tau)\big] = \tau$, indépendamment de P. Aucun autre choix de mesure de référence ne définit une fonction de distribution multivariée basée sur le transport partageant la dernière propriété pour toute dimension, qui est la clé de la régression quantile multivariée (voir le chapitre 7).

1.4.1 Régularité de la distribution centre-extérieur

La fonction de distribution centre-extérieur $\mathbf{F}_\pm$ est bien définie si et seulement si elle est continue (cf. Rockafellar (1970)). Par conséquent, l'étude de sa régularité est importante. On sait qu'en une dimension la fonction de distribution est continue pour toutes les probabilités absolument continues (par rapport à la mesure de Lebesgue). Des résultats similaires pour sa contrepartie multivariée devraient être souhaitables. La régularité des fonctions convexes satisfaisant une condition de transport est généralement traitée au moyen de l'équation de Monge-Ampère (cf. Figalli (2017) eg.). Nous verrons que dans ce cas la continuité de $\mathbf{F}_\pm$ est atteinte en supposant que P a une densité p satisfaisant :

Hypothèse A. Pour tout $R > 0$, il existe $0 < \lambda_R \le \Lambda_R$ tel que

$$\lambda_R \le p(\mathbf{x}) \le \Lambda_R \quad \text{pour tous } \mathbf{x} \in \mathcal{X} \cap R\,\mathbb{S}^d. \tag{1.22}$$

En particulier, $\mathbf{F}_\pm$ satisfait la condition d'avancement de P vers Q et c'est le gradient d'une fonction convexe φ, de sorte que son sous-différentiel satisfait

$$\int_{\partial\varphi(A)} u_d(\mathbf{y})d\mathbf{y} = \int_A p(\mathbf{x})d\mathbf{x}. \tag{1.23}$$

Le résultat principal du chapitre 6 est le suivant.

Theorem 1.4.1. *Soit* P *une mesure de probabilité avec une densité* p *supportée sur l'ensemble convexe ouvert* $\mathcal{X} \subseteq \mathbb{R}^d$.

(i) *Si* p *satisfait à* (1.22), *alors* $K := \partial\psi(\mathbf{0})$ *est un ensemble compact et convexe de mesure de Lebesgue* 0 *tel que la fonction quantile centre-extérieur* $\mathbf{Q}_\pm := \nabla\psi$ *et la fonction de distribution centre-extérieur* $\mathbf{F}_\pm := \nabla\psi^*$ *sont des homéomorphismes entre* $\mathbb{S}^d \setminus \{\mathbf{0}\}$ *et* $\mathcal{X} \setminus K$, *inverses l'un de l'autre.*

(ii) *Si, en outre,* $p \in \mathcal{C}^{k,\alpha}_{\text{loc}}(\mathcal{X})$ *pour un certain* $k \in \mathbb{N}$ *and* $\alpha \in (0,1)$, *alors* $\mathbf{Q}_\pm$ *et* $\mathbf{F}_\pm$ *sont difféomorphismes de classe* $\mathcal{C}^{k+1,\alpha}_{\text{loc}}$ *entre* $\mathbb{S}^d \setminus \{\mathbf{0}\}$ *et* $\mathcal{X} \setminus K$.

La preuve repose sur la théorie de Caffarelli Caffarelli (1990, 1991, 1992), Caffarelli (1990) qui a été largement étudiée par Figalli dans Figalli and Kim (2010); Philippis and Figalli (2012); Cordero-Erausquin and Figalli (2019). Nous recommandons le livre Figalli (2017) comme guide d'introduction à ce sujet et ses implications pour la régularité de la application de transport.Jusqu'à Figalli (2018), la régularité de $\mathbf{F}$ n'était couverte par aucun de ces travaux, voir Cordero-Erausquin and Figalli (2019) pour une étude du cas le plus général. Cependant les hypothèses de Figalli (2018) (on suppose $\mathbf{A}$ et que P est supporté dans tout l'espace) sont trop fortes, comme nous le verrons dans le chapitre 6, elles peuvent être être assouplies .

En conséquence, si P est une probabilité avec une densité p supportée sur l'ensemble convexe $\mathcal{X} \subseteq \mathbb{R}^d$ tel que p satisfait à (1.22),, les propriétés suivantes sont satisfaites.

- La propriété de Glivenko-Cantelli (1.21) (Hallin et al., 2021a, Proposition 2.4) est vérifiée. Ce qui implique que la fonction de distribution du centre de la population vers l'extérieur peut être estimée uniformément à partir de l'échantillon.

- Pour toute $\mathbf{u}$ sur la sphère unité $\mathcal{S}^{d-1}$, toute séquence $(t_n)_{n\in\mathbb{N}}$ de nombres réels telle que $t_n \to \infty$, on a la limite
$$\lim_{n\to\infty} \mathbf{F}_{\pm}(t_n\mathbf{u}) = \mathbf{u} \tag{1.24}$$
Cette propriété est analogue à $\lim_{n\to\infty} F(t_n) = 1$ and $\lim_{n\to\infty} F(-t_n) = 0$ dans le cas univarié, c'est-à-dire du point de vue centre-extérieur.

- pour tout $r \in (0,1)$ et tout $\mathbf{y}$ appartenant à la frontière de $\mathbf{Q}_{\pm}(r\,\mathbb{S}^d)$, il existe un rayon T émanant de $\mathbf{y}$ pour lequel $\overline{\mathbf{Q}_{\pm}(r\,\mathbb{S}^d)} \cap T = \{\mathbf{y}\}$. En termes familiers, pour chaque point d'un contour quantile, il existe un rayon qui ne touche pas à nouveau l'ensemble. De plus, $\mathbb{C}_P(\tau)$ sont des régions imbriquées et connectées.

- Si, en outre, $\mathcal{X}$ est compact, pour tout $r \to 1$, $\mathbf{Q}_{\pm}(r\,\mathbb{S}^d)$ tend vers $\mathcal{X}$ en distance de Hausdoff.
$$\lim_{r\to 1} \mathrm{d}_H\big(\mathbf{Q}_{\pm}(r\,\mathbb{S}^d), \mathcal{X}\big) = 0. \tag{1.25}$$
Le support de P peut être approximé par les limites des régions quantiles.

1.4.2 Régression quantile non paramétrique avec sortie multivariée

Le chapitre 7 propose une nouvelle et importante extension non paramétrique et multivariée, basée sur le concept de quantiles du centre vers l'extérieur, du célèbre concept de régression quantile de Koenker et Bassett Koenker and Bassett (1978), un outil puissant dans l'étude statistique de la dépendance d'une variable d'intérêt Y par rapport aux covariables $\mathbf{X} = (X_1, \ldots, X_m)$. Contrairement à la régression classique qui, en quelque sorte, se concentre sur les moyennes conditionnelles $\mathrm{E}[Y|\mathbf{X}]$, la régression quantile nécessite la connaissance complète des distributions conditionnelles $\mathrm{P}_{\mathrm{Y}|\mathbf{X}=\mathbf{x}}$ de Y étant donné

$\mathbf{X} = \mathbf{x}$. Sa formulation non paramétrique, dont l'étude commence avec les travaux pionniers de Stone (1977), est devenue partie intégrante de la pratique statistique quotidienne, avec d'innombrables applications dans tous les domaines de la recherche scientifique dans lesquels un nombre fini de paramètres fournit un modèle trop rigide pour expliquer certains comportements observés.

La principale motivation pour l'utilisation de la régression quantile est l'observation que $\mathrm{P}_{\mathrm{Y}|\mathbf{X}=\mathbf{x}}$ est beaucoup plus informatif que la simple connaissance d'un paramètre d'intérêt, comme la moyenne ou la médiane conditionnelle. La figure 1.2 montre un exemple de régression quantile univariée, avec X =âge et Y =pli cutané du triceps. En regardant uniquement la médiane (ligne rouge), nous ne pourrons obtenir qu'une description de la tendance du modèle, cependant, en prenant en compte les informations données par les tubes quantiles, nous obtenons l'hétéroscédasticité du modèle.

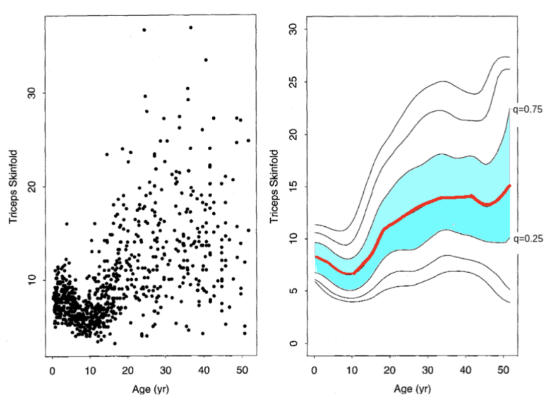

Figure 1.2: Exemple de régression quantile univariée : X =âge et Y =pli cutané du triceps. Figure extraite de Yu and Jones (1998). La ligne rouge représente la médiane et la région bleue la région du quantile centre-extérieur d'ordre $0,5$.

La fonction de tendance dans la régression quantile est la médiane conditionnelle, qui correspond à $Q(\frac{1}{2}|\,\mathbf{x})$. Cependant, la propriété fondamentale de la régression quantile, qui la différencie de tous les autres modèles de régression, *"contrôle probabiliste"* des tubes et des quantiles, c'est à dire

$$\mathbb{P}\left[Y \in \left[Q(\frac{1}{2} - \frac{\tau}{2}|\,\mathbf{x}), Q(\frac{1}{2} + \frac{\tau}{2}|\,\mathbf{x})\right] |\,\mathbf{X} = \mathbf{x}\right] = \tau,$$

sans tenir compte $\mathbb{P}$. Grâce à cette propriété, nous savons, par exemple, que la région bleue de la figure 1.2 a une probabilité de $0,5$. C'est le principe de base de la régression quantile. Tout modèle de régression qui ne le satisfait pas ne peut pas être considéré comme une régression quantile, et s'il était destiné à être une régression quantile, il a certainement manqué sa cible.

Bien sûr, nous ne pouvons pas nous contenter d'une définition de la régression quantile qui n'est valable que pour les sorties univariées; considérons par exemple le modèle suivant:

$$\mathbf{Y} = \begin{pmatrix} Y \\ Z \end{pmatrix} = \begin{pmatrix} V \\ \frac{U\sqrt{|X-\frac{1}{2}|}+V\,(1-\sqrt{|X-\frac{1}{2}|})}{\sqrt{|X-\frac{1}{2}|+(1-\sqrt{|X-\frac{1}{2}|})^2}} \end{pmatrix}, \quad X \sim \mathcal{U}_{(0,1)},\ U, V \sim \mathcal{N}(0,1), \quad (1.26)$$

étant X, U et V mutuellement indépendants. La régression uinivariée sur les marginaux ne nous donne aucune information utile, les deux sont indépendants de X. Toute l'information se trouve dans la distribution conjointe. Le lecteur peut se demander si une projection dans toutes les directions nous donnerait plus d'informations. En effet, plus de projections donneront plus d'informations, puisque celles-ci caractérisent la probabilité du vecteur $\mathbf{Y}$. Cela va dans le sens des profondeurs intégrées (voir Cuevas and Fraiman (2009)) qui, comme toute régression basée sur la profondeur (pour un traitement plus approfondi, voir (Serfling and Zuo, 2000) ou (Serfling, 2002, 2019) et les références qui y sont données), perd le contrôle probabiliste des régions.

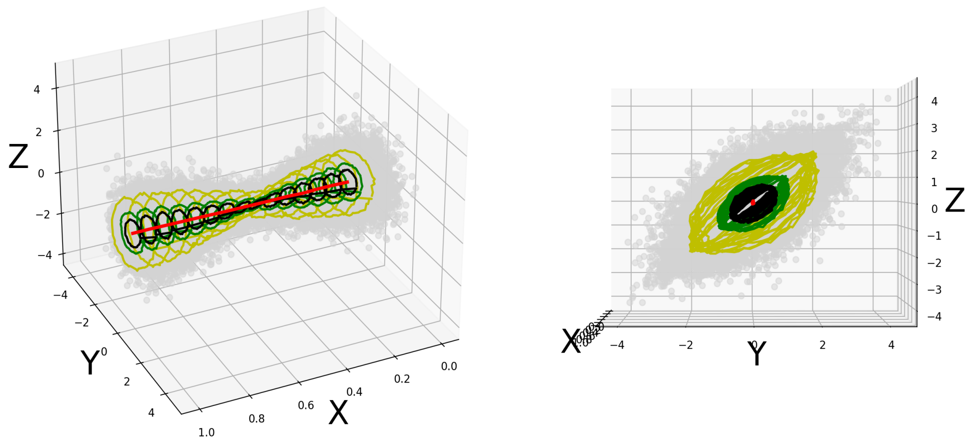

Figure 1.3: Exemple de régression quantile univariée; avec X et $\mathbf{Y} = (Y, Z)$ comme dans (1.26). Le nombre de points observés est $n = 7000$. Les contours du quantile pour $\tau = 0,2, 0,4, 0,8$ sont représentés en noir ($\tau = 0,2$), vert ($\tau = 0,4$) et jaune ($\tau = 0,8$). Le centre est représenté en rouge.

La méthodologie proposée dans le chapitre 7 définit des régions et des tubes quantiles, capables d'analyser des modèles non paramétriques tout en maintenant un contrôle probabiliste asymptotique des régions–le contrôle probabiliste empirique converge vers celui de la population, qui est connu. La figure 1.3 analyse le modèle (1.26), donnant du même coup une solution très visuelle, la tendance du modèle est -comme prévu- clairement constante; cependant la covariable X atteint son pic d'influence sur le vecteur d'intérêt dans la direction $(1, 1)$. *La fonction quantile centre-extérieur* $\mathbf{Q}_{\pm}$ *de* $\mathbf{Y}$ *étant donné* $\mathbf{X} = \mathbf{x}$ est définie comme suit

$$\mathbf{u} \in \mathbb{S}_d \mapsto \mathbf{Q}_{\pm}(\mathbf{u}|\,\mathbf{x}) \in \mathbb{R}^d.$$

Il définit naturellement les régions conditionnelles quantiles comme étant

$$\mathbb{C}_{\pm}(\tau|\mathbf{x}) := \mathbf{Q}_{\pm}(\tau\,\overline{\mathbb{S}}_d|\mathbf{x}) \quad \tau \in (0,1), \;\; \mathbf{x} \in \mathbb{R}^m,$$

qui satisfont la propriété fondamentale

$$\mathbb{P}\left[\mathbf{Y} \in \mathbb{C}_{\pm}(\tau|\,\mathbf{x})\big|\,\mathbf{X} = \mathbf{x}\right] = \tau \quad \text{ pour tous } \mathbf{x} \in \mathbb{R}^m, \tau \in (0,1), \text{ et } \mathbb{P}.$$

Pour $\tau = 0$, la médiane conditionnelle est

$$\mathbb{C}_{\pm}(0\,|\,\mathbf{x}) := \bigcap_{\tau\in(0,1)} \mathbb{C}_{\pm}(\tau|\,\mathbf{x})$$

Il caractérise également le phénomène de nichage (non *"quantile-crossing"*) *"tubes quantiles de régression d'ordre* $\tau \in (0,1)$*"* (dans $\mathbb{R}^{m+d}$)

$$\mathbb{T}_{\pm}(\tau) := \left\{\left(\mathbf{x}, \mathbf{Q}_{\pm}(\tau\,\overline{\mathbb{S}}_d|\,\mathbf{x})\right)\,\middle|\mathbf{x} \in \mathbb{R}^m\right\}, \quad \tau \in (0,1)$$

qui satisfont la propriété fondamentale

$$\mathbb{P}\left[(\mathbf{X}, \mathbf{Y}) \in \mathbb{T}_{\pm}(\tau)\right] = \tau \quad \text{ irrespective of } \mathbb{P}, \tau \in (0,1).$$

Pour $\tau = 0$, on définit

$$\mathbb{T}_{\pm}(0) := \left\{(\mathbf{x}, \mathbf{y}))\,\middle|\mathbf{x} \in \mathbb{R}^m, \, \mathbf{y} \in \mathbb{C}_{\pm}(0\,|\,\mathbf{x})\right\} = \bigcap_{\tau\in(0,1)} \mathbb{T}_{\pm}(\tau)$$

(comme le graphe de $\mathbf{x} \mapsto \mathbb{C}_{\pm}(\tau|\,\mathbf{x})$); par un léger abus de langage, nous l'appelons également *médiane de régression* de $\mathbf{Y}$ par rapport à $\mathbf{X}$.

Lorsque ce qui est observé est un échantillon $(\mathbf{X}, \mathbf{Y})^{(n)} := \big((\mathbf{X}_1, \mathbf{Y}_1), \ldots, (\mathbf{X}_n, \mathbf{Y}_n)\big)$ de n copies i.i.d. de $(\mathbf{X}, \mathbf{Y}) \sim= \mathrm{P}_{\mathbf{XY}}$, le chapitre 7 propose un estimateur de $\mathbf{u} \mapsto \mathbf{Q}_{\pm}(\mathbf{u}|\mathbf{X} = \mathbf{x})$, $\mathbf{x} \in \mathbb{R}^m$. Notre estimateur est obtenu en deux étapes : à l'étape 1, nous construisons une distribution empirique de $\mathbf{Y}$ conditionnelle à $\mathbf{X} = \mathbf{x}$ et, à l'étape 2, nous calculons le quantile centre-extérieur correspondant.

La distribution empirique de $\mathbf{Y}$ conditionnelle à $\mathbf{X} = \mathbf{x}$ est approximée par une fonction de poids, donc (si nous ne voulons pas nous restreindre aux fonctions de poids constantes par morceaux) ce sera une probabilité atomique de poids non égaux. La solution que nous obtenons nécessite la résolution du problème de transport optimal suivant:

$$\begin{aligned}
&\min_{\pi:=\{\pi_{i,j}\}} \sum_{i=1}^{N}\sum_{j=1}^{n} \frac{1}{2}\left|\mathbf{Y}_j - \mathfrak{G}_i\right|^2 \pi_{i,j}\ ,\\
\text{s.t.}\ &\sum_{j=1}^{n} \pi_{i,j} = \frac{1}{N}, \quad i \in \{1,2,\ldots,N\},\\
&\sum_{i=1}^{N} \pi_{i,j} = w_j^{(n)}(\mathbf{x};\mathbf{X}^{(n)}), \quad j \in \{1,2,\ldots,n\},\\
&\pi_{i,j} \geq 0, \quad i \in \{1,2,\ldots,N\},\ j \in \{1,2,\ldots,n\},
\end{aligned} \tag{1.27}$$

où $\mathfrak{G}_1^{(n)}, \ldots, \mathfrak{G}_N^{(n)}$ est une grille régulière de $\mathbb{U}_d$. Dans ce cas, N est arbitraire, on lui demandera seulement de tendre vers l'infini. Le problème est que la solution de (1.27) est un plan de transport et non une application en tant que telle. Pour créer une application, nous appliquons le critère

$$\mathbf{T}^*(\mathfrak{G}_i|\ \mathbf{x}) := \arg\inf\left\{\|\mathbf{y}\| :\ \mathbf{y} \in \operatorname{conv}\left(\{\mathbf{Y}_J :\ J \in \arg\max_j \pi^*_{i,j}(\mathbf{x})\}\right)\right\}, \tag{1.28}$$

où $\operatorname{conv}(A)$ désigne l'enveloppe convexe de A. Nous procédons comme dans Hallin et al. (2021a) en choisissant l'interpolation lisse cycliquement monotone avec la plus grande constante de Lipschitz. Désignons-la par $\mathbf{u} \mapsto \mathbf{Q}^{(n)}_{w,\pm}(\mathbf{u}|\mathbf{x})$ et appelons-la la *fonction quantile conditionnelle empirique centre-extérieur.*

Nous montrerons que si la fonction de poids est cohérente (au sens de Stone (1977)), pour tout $\mathbf{u} \in \mathbb{S}_d$ et $\epsilon > 0$,

$$\mathbb{P}\left(\ \mathbf{Q}^{(n)}_{w,\pm}(\mathbf{u}\,|\mathbf{X}) \notin \mathbf{Q}_{\pm}(\mathbf{u}\,|\mathbf{X}) + \epsilon\mathbb{S}_d\ \right) \longrightarrow 0 \quad \text{as } n \text{ y } N \to \infty,$$

et, pour tous $\tau \in (0,1)$,

$$\mathbb{P}\left(\ \mathbb{C}^{(n)}_{\pm}(\tau\,|\mathbf{X}) \not\subset \mathbb{C}_{\pm}(\tau\,|\mathbf{X}) + \epsilon\mathbb{S}_d\ \right) \to 0 \text{ and } \mathbb{P}\left(\ \mathcal{C}^{(n)}_{\pm}(\tau\,|\mathbf{X}) \not\subset \mathcal{C}^{(n)}_{\pm}(\tau\,|\mathbf{X}) + \epsilon\mathbb{S}_d\ \right) \to 0$$

en tant que n et $N \to \infty$.

Remark 1.4.2. *La preuve de ce résultat suit la ligne de convergence des mappings empiriques multivalués, initiée par del Barrio and Loubes (2019) pour prouver la convergence des potentiels de transport optimaux dans le cas quadratique, poursuivie par Hallin et al. (2021a) pour la preuve du théorème de Glivenko-Cantelli multivarié et finalement formalisée dans Segers (2022). Voir également del Barrio et al. (2021) pour le cas général des coûts. Dans ce cas, nous avons une difficulté supplémentaire; la convergence est en probabilité, et, par conséquent, nous devons passer de nombreuses fois par des sous-séquences. C'est une procédure fastidieuse.*

Le chapitre 6 trouve son application dans le 7; il permet de donner les conditions que $(\mathbf{X}, \mathbf{Y})$ doivent satisfaire afin d'obtenir des notions de convergence plus fortes. Dans ce cas, si l'on suppose que $\mathbf{A}$ tient p.s. pour la probabilité conditionnelle, alors, pour tout sous-ensemble compact $K \subset \mathbb{S}_d \setminus \{\mathbf{0}\}$, lorsque n y $N \to \infty$,

$$\sup_{\mathbf{u} \in K} |\mathbf{Q}^{(n)}_{w,\pm}(\mathbf{u} \,|\mathbf{X}) - \mathbf{Q}_{\pm}(\mathbf{u} \,|\mathbf{X})| \xrightarrow{\mathbb{P}} 0$$

et, pout tous $\tau \in (0,1)$ y $\epsilon > 0$,

$$\mathbb{P}\left(d_H\left(\mathcal{C}^{(n)}_{\pm}(\tau \,|\mathbf{X}), \mathcal{C}_{\pm}(\tau \,|\mathbf{X})\right) > \epsilon\right) \to 0.$$

De plus, sous ces conditions, nous obtenons le contrôle de probabilité asymptotique

$$\mathbb{P}\left(\mathbf{Y} \in \mathbb{C}^{(n)}_{\pm}(\tau \,|\mathbf{X}) \big| \mathbf{X}\right) \xrightarrow{\mathbb{P}} \tau \quad \text{pour tous } \tau \in (0,1),$$

qui justifie la méthodologie proposée.

Le chapitre 7 se termine par une série d'expériences qui montrent le bon comportement des modèles incluant l'hétéroscédasticité et les tendances non linéaires; sa puissance en tant qu'outil d'analyse de données est également illustrée sur quelques ensembles de données réelles.

Bibliography

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 214–223. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/arjovsky17a.html.

Emily Black, Samuel Yeom, and Matt Fredrikson. Fliptest: Fairness testing via optimal transport. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, page 111–121, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450369367. doi: 10.1145/3351095.3372845. URL https://doi.org/10.1145/3351095.3372845.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 02 2013. ISBN 9780199535255. doi: 10.1093/acprof:oso/9780199535255.001.0001. URL https://doi.org/10.1093/acprof:oso/9780199535255.001.0001.

Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991. doi: https://doi.org/10.1002/cpa.3160440402. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160440402.

Luis Caffarelli. Some regularity properties of solutions of monge ampère equation. *Communications on Pure and Applied Mathematics*, 44:965–969, 1991.

Luis A. Caffarelli. A localization property of viscosity solutions to the monge-ampere equation and their strict convexity. *Annals of Mathematics*, 131(1):129–134, 1990. ISSN 0003486X. URL http://www.jstor.org/stable/1971509.

Luis A. Caffarelli. The regularity of mappings with a convex potential. *Journal of the American Mathematical Society*, 5(1):99–104, 1992. ISSN 08940347, 10886834. URL http://www.jstor.org/stable/2152752.

Javier Cárcamo, Antonio Cuevas, and Luis-Alberto Rodríguez. Directional differentiability for supremum-type functionals: Statistical applications. *Bernoulli*, 26(3):2143 – 2175, 2020. doi: 10.3150/19-BEJ1188. URL https://doi.org/10.3150/19-BEJ1188.

Victor Chernozhukov, Alfred Galichon, Marc Hallin, and Marc Henry. Monge-Kantorovich depth, quantiles, ranks and signs. *Ann. Statist.*, 45(1):223–256, 2017. ISSN 0090-5364. doi: 10.1214/16-AOS1450. URL https://doi.org/10.1214/16-AOS1450.

Dario Cordero-Erausquin and Alessio Figalli. Regularity of monotone transport maps between unbounded domains. *Discrete and Continuous Dynamical Systems*, 39(12): 7101–7112, 2019.

Juan A. Cuesta and Carlos Matrán. Notes on the wasserstein metric in hilbert spaces. *Annals of Probability*, 17:1264–1276, 1989.

Antonio Cuevas and Ricardo Fraiman. On depth measures and dual statistics. a methodology for dealing with general data. *Journal of Multivariate Analysis*, 100(4):753–766, 2009. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2008.08.002. URL https://www.sciencedirect.com/science/article/pii/S0047259X08001796.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pages 2292–2300, 2013. URL http://papers.nips.cc/paper/4927-sinkhorn-distances-lightspeed-computation-of-optimal-transport.

Lucas de Lara, Alberto González-Sanz, Nicholas Asher, and Jean-Michel Loubes. Transport-based counterfactual models. *Pre-Print*, 2021.

Lucas De Lara, Alberto González-Sanz, and Jean-Michel Loubes. Diffeomorphic registration using sinkhorn divergences. *To appear in SIAM Journal on Imaging Sciences*, 2022+.

Nabarun Deb and Bodhisattva Sen. Multivariate rank-based distribution-free nonparametric testing using measure transportation. *J. Amer. Statist. Assoc.*, 2019. To appear.

Nabarun Deb, Bhaswar B. Bhattacharya, and Bodhisattva Sen. Efficiency lower bounds for distribution-free hotelling-type two-sample tests based on optimal transport, 2021. URL https://arxiv.org/abs/2104.01986.

Eustasio del Barrio and Jean-Michel Loubes. Central limit theorems for empirical transportation cost in general dimension. *The Annals of Probability*, 47(2):926 – 951, 2019. doi: 10.1214/18-AOP1275. URL https://doi.org/10.1214/18-AOP1275.

Eustasio del Barrio, Evarist Giné, and Frederic Utzet. Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli*, 11(1):131–189, 2005. ISSN 1350-7265. doi: 10.3150/bj/1110228245. URL https://doi.org/10.3150/bj/1110228245.

Eustasio del Barrio, Alberto González-Sanz, and Jean-Michel Loubes. Central limit theorems for general transportation costs. *Pre-Print*, 2021.

R. M. Dudley. The speed of mean glivenko-cantelli convergence. *Ann. Math. Statist.*, 40 (1):40–50, 02 1969. doi: 10.1214/aoms/1177697802. URL https://doi.org/10.1214/aoms/1177697802.

Jean Feydy, Benjamin Charlier, François-Xavier Vialard, and Gabriel Peyré. Optimal transport for diffeomorphic registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 291–299. Springer, 2017.

Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trouvé, and Gabriel Peyré. Interpolating between optimal transport and mmd using sinkhorn divergences. In *AISTATS*, 2019.

Alessio Figalli. *The Monge-Ampère Equation and Its Applications*. Zurich Lectures in Advanced Mathematics, European Mathematical Society (EMS), Zurich, 2017. ISBN 978-3-03719-670-0.

Alessio Figalli. On the continuity of center-outward distribution and quantile functions. *arXiv: Analysis of PDEs*, 2018.

Alessio Figalli and Young-Heon Kim. Partial regularity of brenier solutions of the monge-ampère equation. *Discrete and Continuous Dynamical Systems*, 28:559–565, 2010.

N. Fournier and A. Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162:707–738, 2013.

Alfred Galichon. *Optimal Transport Methods in Economics*. Princeton University Press, 09 2016. ISBN 9780691172767. doi: 10.23943/princeton/9780691172767.001.0001. URL https://doi.org/10.23943/princeton/9780691172767.001.0001.

Wilfrid Gangbo and Robert J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996. doi: 10.1007/BF02392620. URL https://doi.org/10.1007/BF02392620.

Aude Genevay, Lénaïc Chizat, Francis Bach, Marco Cuturi, and Gabriel Peyré. Sample complexity of sinkhorn divergences. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pages 1574–1583. PMLR, 16–18 Apr 2019. URL http://proceedings.mlr.press/v89/genevay19a.html.

Javier González-Delgado, Alberto González-Sanz, Juan Cortés, and Pierre Neuvial. Two-sample goodness-of-fit tests on the flat torus based on wasserstein distance and their relevance to structural biology. *arXiv: Probability*, 2021.

Paula Gordaliza, Eustasio Del Barrio, Gamboa Fabrice, and Jean-Michel Loubes. Obtaining fairness using optimal transport theory. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 2357–2365. PMLR, 09–15 Jun 2019a. URL http://proceedings.mlr.press/v97/gordaliza19a.html.

Paula Gordaliza, Eustasio Del Barrio, Gamboa Fabrice, and Jean-Michel Loubes. Obtaining fairness using optimal transport theory. In *International Conference on Machine Learning*, pages 2357–2365. PMLR, 2019b.

Marc Hallin and Gilles Mordant. On the finite-sample performance of measure transportation-based multivariate rank tests, 2021.

Marc Hallin, Eustasio del Barrio, Juan Cuesta-Albertos, and Carlos Matrán. Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *The Annals of Statistics*, 49(2):1139 – 1165, 2021a. doi: 10.1214/20-AOS1996.

Marc Hallin, Gilles Mordant, and Johan Segers. Multivariate goodness-of-fit tests based on Wasserstein distance. *Electron. J. Stat.*, 15(1):1328–1371, 2021b. doi: 10.1214/21-ejs1816. URL https://doi.org/10.1214/21-ejs1816.

Zaïd Harchaoui, Lang Liu, and Soumik Pal. Asymptotics of entropy-regularized optimal transport via chaos decomposition. *arXiv: Probability*, 2020.

Shayan Hundrieser, Marcel Klatt, Thomas Staudt, and Axel Munk. A unifying approach to distributional limits for empirical optimal transport. *Pre-Print*, 02 2022. URL https://arxiv.org/pdf/2202.12790.pdf.

Lev Kantorovich. On the translocation of masses. *Journal of Mathematical Sciences*, 133: 1381–1382, 2006.

Marcel Klatt, Carla Tameling, and Axel Munk. Empirical regularized optimal transport: statistical theory and applications. *SIAM J. Math. Data Sci.*, 2(2):419–443, 2020. doi: 10.1137/19M1278788. URL https://doi.org/10.1137/19M1278788.

Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1913643.

Robert J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80(2):309 – 323, 1995. doi: 10.1215/S0012-7094-95-08013-2. URL https://doi.org/10.1215/S0012-7094-95-08013-2.

Gonzalo Mena and Jonathan Niles-Weed. Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/5acdc9ca5d99ae66afdfe1eea0e3b26b-Paper.pdf.

Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019. ISSN 1935-8237. doi: 10.1561/2200000073. URL http://dx.doi.org/10.1561/2200000073.

Guido Philippis and Alessio Figalli. Partial regularity for optimal transport maps. *Publications mathématiques de l'IHÉS*, 121, 09 2012. doi: 10.1007/s10240-014-0064-7.

Laurent Risser, Alberto González Sanz, Quentin Vincenot, and Jean-Michel Loubes. Tackling algorithmic bias in neural-network classifiers using wasserstein-2 regularization. *J. Math Imaging Vis*, 2021.

R. Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series. Princeton University Press, Princeton, N. J., 1970.

Johan Segers. Graphical and uniform consistency of estimated optimal transport plans. *arXiv*, 2022.

Robert Serfling. Quantile functions for multivariate analysis: approaches and applications. *Statistica Neerlandica*, 56(2):214–232, 2002. doi: https://doi.org/10.1111/1467-9574.00195. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9574.00195.

Robert Serfling. Depth functions on general data spaces, i. perspectives, with consideration of "density" and "local" depths". Available at https://www.utdallas.edu~serfling, 2019.

Robert Serfling and Yijun Zuo. General notions of statistical depth function. *The Annals of Statistics*, 28(2):461–482, 2000. doi: 10.1214/aos/1016218226. URL https://doi.org/10.1214/aos/1016218226.

Hongjian Shi, Mathias Drton, and Fang Han. Distribution-free consistent independence tests via center-outward ranks and signs. *J. Amer. Statist. Assoc.*, 117(537):395–410, 2022. ISSN 0162-1459. doi: 10.1080/01621459.2020.1782223. URL https://doi.org/10.1080/01621459.2020.1782223.

Max Sommerfeld and Axel Munk. Inference for empirical wasserstein distances on finite spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1): 219–238, 2018.

Charles J. Stone. Consistent nonparametric regression. *Ann. Statist.*, 5(4):595–620, 07 1977. doi: 10.1214/aos/1176343886. URL https://doi.org/10.1214/aos/1176343886.

Cédric Villani. *Topics in optimal transportation*. Number 58 in Graduate studies in mathematics. American Mathematical Soc., 2003.

Cédric Villani. *Optimal Transport: Old and New*. Number 338 in Grundlehren der mathematischen Wissenschaften. Springer, Berlin, 2008. ISBN 978-3-540-71049-3. OCLC: ocn244421231.

Richard A Vitale. Lp metrics for compact, convex sets. *Journal of Approximation Theory*, 45(3):280–287, 1985. ISSN 0021-9045. doi: https://doi.org/10.1016/0021-9045(85)90051-6. URL https://www.sciencedirect.com/science/article/pii/0021904585900516.

Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A):2620 – 2648, 2019. doi: 10.3150/18-BEJ1065. URL https://doi.org/10.3150/18-BEJ1065.

Keming Yu and M. C. Jones. Local linear quantile regression. *Journal of the American Statistical Association*, 93(441):228–237, 1998. doi: 10.1080/01621459.1998.10474104. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1998.10474104.

CHAPTER 1

Introducción (versión española)

El transporte óptimo es un problema de asignación de recursos presente en múltiples áreas de las matemáticas; y por ende en sus aplicaciones. Esta versatilidad también se manifiesta en su propio marco teórico: el estudio de su regularidad recurre a técnicas avanzadas de ecuaciones diferenciales (Caffarelli, 1990, 1991, 1992); el desarrollo de métodos computacionales eficientes, el análisis numérico y la combinatoria (Peyré and Cuturi, 2019, capítulo 3); su comportamiento asintótico, el análisis convexo (del Barrio and Loubes, 2019) y los procesos empíricos (del Barrio et al., 2005).

El transporte óptimo consiste en hallar, entre todas las medidas de probabilidad con las mismas marginales fijas, aquella que minimiza el coste medio de transporte. Este valor mínimo medio es conocido como el coste de transporte óptimo. En aquellos casos en los que una de las probabilidades cuente con densidad, la solución es determinista y se da mediante la aplicación de transporte (Gangbo and McCann, 1996). El coste de transporte óptimo proporciona una estructura métrica –la comúnmente conocida como distancia de Wasserstein– al espacio de medidas de probabilidad. Por lo tanto, desde un punto de vista estadístico, el transporte óptimo ofrece una herramienta para la comparación de datos que tiene en consideración la geometría del espacio latente. La cual ha probado su eficacia en problemas como reparación de sesgos en aprendizaje automático (Risser et al., 2021; Gordaliza et al., 2019a; Black et al., 2020), en la modelización de razonamientos contrafactuales (de Lara et al., 2021; Black et al., 2020) o en registro difeomórfico (Feydy et al., 2017; De Lara et al., 2022+).

Una aplicación natural de cualquier distancia significativa entre distribuciones es el problema de bondad de ajuste, es decir, el problema de probar la hipótesis nula de que una muestra $\mathbf{X}_1, \dots, \mathbf{X}_n$ proviene de una población con una distribución completamente especificada P (Hallin et al., 2021b). Efectivamente, en dimensiones moderadas la distancia de Wasserstein puede proporcionar un test estadístico consistente contra toda alternativa fija, eg. González-Delgado et al. (2021) propone un test consistente basado en la distancia de Wasserstein en el toro plano de dimensión 2. Estas aplicaciones, donde la distancia de Wasserstein cuantifica la similitud entre diferentes muestras de datos, necesitan una justificación matemática rigurosa.

La primera parte de esta tesis se enfoca en el estudio asintótico-estadístico del problema de transporte. Veremos que la dimensión del espacio latente influye en el error de la aproximación empírica del coste de transporte óptimo pero no en su varianza, de hecho, los límites de las fluctuaciones (diferencia entre la distancia de Wasserstein empírica y su media) son gaussianos. La influencia de la dimensión en el ratio de convergencia de la versión empírica a la poblacional es conocido como *la maldición de la dimensión*. Las primeras demostraciones rigurosas de este hecho se remontan a 1969 con el trabajo de Dudley (1969). Recientemente, los trabajos de Fournier and Guillin (2013) y Weed and Bach (2019) corroboran el hecho de que la distancia de Wasserstein empírica entre dos distribuciones continuas converge a la poblacional con ratio de $n^{-\frac{1}{d}}$, salvo quizás factores logarítmicos. Esto demuestra que la distancia de Wasserstein no puede ser utilizada para proporcionar un test de bondad de ajuste consistente en dimensión general. Las fluctuaciones tienen un comportamiento asintótico diferente. Es mas, pueden ser acotadas con ratio $n^{-\frac{1}{2}}$, independiente de la dimensión (Weed and Bach, 2019, proposición 20). Los argumentos de Weed and Bach (2019), basados en la desigualdad de McDiarmid, carecen de sentido para probabilidades no soportadas en un conjunto acotado. Un estudio más preciso de las fluctuaciones es el realizado por del Barrio and Loubes (2019) mediante la desigualdad de Efron-Stein. Sin embargo, son específicos para el coste cuadrático –en el cual las aplicaciones de transporte son gradientes de funciones convexas– haciendo no trivial su generalización a otros costes.

La citada maldición de la dimensión aparece para probabilidades con densidad. Cuando las dos probabilidades son discretas el problema de optimización es paramétrico y satisface un teorema central del límite con ratio $n^{-\frac{1}{2}}$, véase Sommerfeld and Munk (2018). Si una de las probabilidades es discreta –el llamado problema semi-discreto–, del Barrio and Loubes (2019) probaron el teorema central del límite centrado en el valor poblacional como aplicación del resultado obtenido para las fluctuaciones. Esta estrategia, como veremos en este trabajo, no es la más a adecuada. La versión funcional del delta-método proporciona una metodología que requiere menos hipótesis. Esto ha sido observado paralelamente por Hundrieser et al. (2022).

El control de las fluctuaciones resulta más útil para analizar el problema de transporte regularizado por la entropía. Propuesto por Cuturi (2013), es, sin lugar dudas, el método más ampliamente utilizado de regularización del problema de transporte. Mena and Niles-Weed (2019) demuestran, usando los argumentos de del Barrio and Loubes (2019), que las fluctuaciones del problema regularizado son asintóticamente gaussianas. En este trabajo veremos que además el sesgo converge más rápido que la varianza, dando como resultado el teorema central del límite para el coste de transporte regularizado por la entropía. Esto no es suficiente para dar una respuesta positiva al problema de bondad de ajuste basado en nociones de transporte. El término añadido como penalti causa un fenómeno conocido como sesgo entrópico, que reduce significativamente la utilidad del coste de transporte regularizado para realizar inferencia estadística. Feydy et al. (2019) propone una modificación del coste de transporte regularizado –la divergencia de Sinkhorn– que repara este problema. Esta tesis proporciona un desarrollo de segundo orden de la divergencia de Sinkhorn con respecto el proceso empírico. Como consecuencia, se obtiene una caracterización precisa del comportamiento asintótico de la misma. Esto puede ser usado para crear un test de bondad de ajuste basado en la divergencia de Sinkhorn.

Es conocido que los rangos, basados en la noción de orden univariante, proporcionan una metodología general para abordar el problema de bondad de ajuste. La generalización multivariante del concepto de rango no ha sido posible debido a la ausencia de noción de función de distribución multivariante. Recientemente, los trabajos de Marc Hallin y coautores (eg. Hallin et al. (2021a)) proponen la aplicación de transporte entre la muestra de datos y la distribución esférica uniforme como candidata a función de distribución multivariante. Resulta que –siempre según Hallin et al. (2021a)– esta propuesta satisface las principales propiedades que hacen de la función de distribución univariante una herramienta util para la inferencia estadística. Esto ha proporcionado, al igual que se ha mencionado para el caso univariante, una metodología general para la creación de test de bondad de ajuste (Deb and Sen, 2019; Deb et al., 2021) o de independencia (Shi et al., 2022; Hallin and Mordant, 2021). La segunda parte de este manuscrito demuestra la continuidad de la función de distribución multivariante para probabilidades soportadas en un conjunto convexo, extendiendo así el resultado de Figalli (2018). La singularidad en el origen de coordenadas de la distribución de probabilidad esférica uniforme hace que los resultados de Caffarelli (see eg.Caffarelli (1990, 1991, 1992)) no cubran este caso. Para concluir, proporcionaremos, usando el nuevo concepto de función de distribución multivariante, una metodología innovadora para el estudio del problema de regresión no paramétrica con salida multivariante. Veremos que esta propuesta es, hasta la fecha, la única que mantiene la propiedad fundamental de la regresión cuantílica, el control probabilístico de las regiones cuantil. Esta propiedad, en el caso univariante, dictamina que la región cuantil de orden $0 < r < 1$ contiene una proporción de r puntos de la muestra.

Las dos partes que componen este trabajo están claramente delimitadas. En mi opinión, el nexo de unión se encuentra en el aspecto matemático, en los razonamientos y en las herramientas utilizadas, y no en el objeto de estudio en sí mismo. Antes de presentar en detalle los resultados obtenidos, con las definiciones técnicas y las herramientas utilizadas, conviene subrayar el hecho de que cada capítulo es autónomo y está adaptado a partir de su versión on-line (o publicada). En otras palabras, cada uno de ellos puede leerse y analizarse por separado. Quienes no estén interesados en los vínculos entre los distintos capítulos pueden saltarse el resto de la introducción.

1.0.1 El transporte óptimo en pocas palabras

Desde finales del siglo pasado, el problema de asignación de Monge se ha convertido en un importante tema de investigación en estadística y probabilidad, con aplicaciones al aprendizaje automático, la economía, la física y la astronomía, por nombrar sólo algunas. Sea $c : \mathbb{R}^d \times \mathbb{R}^d \to [0, +\infty)$ una función continua –conocida como función de coste– el *coste de transporte óptimo* entre dos medidas de probabilidad $P, Q \in \mathcal{P}(\mathbb{R}^d)$ para el coste c se define como la solución del problema de Monge

$$\inf_{T:\, T_{\#}P=Q} \int_{\mathbb{R}^d} c(\mathbf{x}, T(\mathbf{x})) dP(\mathbf{x}), \tag{1.1}$$

donde la notación $T_\# P$ representa la medida *push-forward*[1], es decir, la medida tal que $T_\# P(A) := P(T^{-1}(A))$, para cada conjunto medible A. Hubo que esperar hasta los años 90, con los trabajos paralelos de Brenier (1991) y Cuesta and Matrán (1989), para demostrar la existencia de la solución para el coste cuadrático ($c(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2$), véase también Gangbo and McCann (1996) para costes más generales y Villani (2003) para un estudio completo.

Tiempo atrás, conretamente en 1942, Kantorovich (véase Kantorovich (2006) para una traducción al inglés del artículo original) formuló la famosa relajación del problema de Monge;

$$\mathcal{T}_c(P, Q) = \inf_{\pi \in \Pi(P,Q)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y}), \tag{1.2}$$

donde $\Pi(P, Q)$ es el conjunto de medidas de probabilidad $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ tales que $\pi(A \times \mathbb{R}^d) = P(A)$ y $\pi(\mathbb{R}^d \times B) = Q(B)$, para todo conjunto medible A, B. Una medida de probabilidad $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ es *un plan de transporte óptimo para el coste* c entre P y Q si es una solución de (1.2).

La principal ventaja de la relajación de Kantorovich es la existencia de soluciones óptimas: para (1.2) sólo requiere la integrabilidad del coste (teorema 4. 1 en Villani (2008)), mientras que el problema de Monge necesita algunas hipótesis sobre una de las dos probabilidades y sobre el coste. En particular, condiciones suficientes son la continuidad absoluta de la probabilidad con respecto a la medida de Lebesgue, y las condiciones de Gangbo and McCann (1996) sobre el coste; es decir $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$, donde $h : \mathbb{R}^d \to [0, \infty)$ es una función no negativa que satisface

(A1) convexidad estricta en $\mathbb{R}^d$,

(A2) dado $r \in \mathbb{R}^+$ y un ángulo $\theta \in (0, \pi)$, existe $M := M(r, \theta) > 0$ tal que, para todo $|\mathbf{p}| > M$, se puede encontrar un cono

$$K(r, \theta, \mathbf{z}, \mathbf{p}) := \left\{ \mathbf{x} \in \mathbb{R}^d : |\mathbf{x} - \mathbf{p}||\mathbf{z}| \cos(\theta/2) \leq \langle \mathbf{z}, \mathbf{x} - \mathbf{p} \rangle \leq r|\mathbf{z}| \right\},$$

con vértice $\mathbf{p}$ (y $\mathbf{z} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$) donde h alcanza su máximo en $\mathbf{p}$,

(A3) $\lim_{|\mathbf{x}| \to \infty} \frac{h(\mathbf{x})}{|\mathbf{x}|} = \infty$.

El problema de transporte admite una formulación dual

$$\mathcal{T}_c(P, Q) = \sup_{(f,g) \in \Phi_c(P,Q)} \int f(\mathbf{x}) dP(\mathbf{x}) + \int g(\mathbf{y}) dQ(\mathbf{y}), \tag{1.3}$$

[1]Usamos directamente la terminología anglosajona, se puede traducir como empujar hacia adelante, en ocasiones diremos que T empuja P hacia Q para referirnos a $Q = T_\# P$.

dónde $\Phi_c(P,Q) = \{(f,g) \in L_1(P) \times L_1(Q) : \ f(\mathbf{x}) + g(\mathbf{y}) \leq c(\mathbf{x},\mathbf{y})\}$, véase Teorema 5.10 en Villani (2008). Una función $\psi \in L_1(P)$ es llamada *un potencial de transporte óptimo de P a Q para el coste c* si existe $\varphi \in L_1(Q)$ tal que el par (ψ, φ) es solución de (1.3). Sorprendentemente, la equivalencia de las formulaciones (1.1) y (1.2) pasa por la regularidad de los potenciales, que a su vez se deriva de la del coste (cf. Gangbo and McCann (1996) por ejemplo). En cuanto al contenido tratado en esta tesis, la formulación dual describirá: en los capítulos 2 y 3, la varianza de los límites de las fluctuaciones; y en el capítulo 3 los precios otorgados por una empresa a determinado producto.

Cada formulación tiene un interés diferente, y su uso depende de la aplicación. Por un lado, cuando existe, la solución de (1.1) define una aplicación de transporte entre medidas de probabilidad. Esto permite inferir las propiedades de una medida de probabilidad a través de otra ya conocida (o alguna medida de referencia). Esta es la idea que está detrás de la exitosa función cuantil multivariante de M. Hallin (Hallin et al. (2021a)); de la regresión cuantil multivariante (capítulo 7); de las explicaciones contrafactuales basadas en el transporte de masas (de Lara et al., 2021) o de la reparación de sesgos (Gordaliza et al., 2019b; Black et al., 2020).

Sea $\mathcal{P}_p(\mathbb{R}^d)$ el espacio de probabilidades sobre $\mathbb{R}^d$ con momento de orden $p \geq 1$ finito. Por otro lado, cuando se considera un coste de tipo potencial ($c_p(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|^p$, para $p \geq 1$), la formulación de Kantorovich entre probabilidades en $\mathcal{P}_p(\mathbb{R}^d)$ admite siempre soluciones, y la función

$$\mathcal{P}_p(\mathbb{R}^d) \times \mathcal{P}_p(\mathbb{R}^d) \ni (P,Q) \to \mathcal{W}_p(P,Q) = (\mathcal{T}_p(P,Q))^{\frac{1}{p}} = \left(\mathcal{T}_{c_p}(P,Q)\right)^{\frac{1}{p}}$$

define una métrica en $\mathcal{P}_p(\mathbb{R}^d)$. Esta está caracterizada tanto por la convergencia débil de probabilidades como por la convergencia de momentos:

$$\mathcal{W}_p(\mu_n,\mu) \to 0 \quad \iff \quad \mu_n \xrightarrow{w} \mu \quad \text{y} \quad \int \|\mathbf{x}\|^p d\mu_n(\mathbf{x}) \to \int \|\mathbf{x}\|^p d\mu(\mathbf{x}),$$

véase (Villani, 2003, capítulo 7). Esta métrica es conocida como *«distancia de Wasserstein»*. Se utiliza ampliamente como discriminador en las redes adversariales Generativas Arjovsky et al. (2017), en el registro difeomórfico o como término de penalización en la reparación del sesgo (Risser et al., 2021).

1.1 Comportamiento asintótico

La probabilidad poblacional P no suele estar disponible para los investigadores; lo que observan es una muestra i.i.d. $\mathbf{X}_1, \dots, \mathbf{X}_n$ de tamaño n de P, que define la medida empírica P_n. Supongamos para esta introducción que Q es conocido. Sin embargo, los resultados aquí expuestos siguen siendo válidos en el caso de dos muestras. El valor $\mathcal{T}_c(P_n, Q)$ es pues la contraparte empírica de la poblacional $\mathcal{T}_c(P,Q)$. Por supuesto, $\mathcal{T}_c(P_n,Q)$ tiende a $\mathcal{T}_c(P,Q)$, pero, ¿a qué ritmo? Es decir, si el número de datos que tenemos n tiende a infinito,

¿cómo varía la diferencia $\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q)$? Para dar una respuesta rápida podemos hacer una división sesgo-varianza del error:

$$\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q) = (\mathbb{E}(\mathcal{T}_c(P_n, Q)) - \mathcal{T}_c(P, Q)) + (\mathcal{T}_c(P_n, Q) - \mathbb{E}(\mathcal{T}_c(P_n, Q))) \,. \tag{1.4}$$

Una simulación puede ayudar a inferir, antes de realizar un análisis formal, los comportamientos asintóticos del sesgo $(\mathbb{E}(\mathcal{T}_c(P_n, Q_n)) - \mathcal{T}_c(P, Q))$ y la varianza $(\mathcal{T}_c(P_n, Q) - \mathbb{E}(\mathcal{T}_c(P_n, Q)))$. Para ello, se fija una dimensión *«relativamente alta»* (por ejemplo $d = 10$) y se simula $\sqrt{n}(\mathcal{T}_2(P_n, Q_n))$, for $P, Q \sim \mathcal{U}_{[0,1]^{10}}$. La figura 1.1 muestra, con diferentes colores, la repetición de 10 simulaciones independientes del mismo procedimiento, los diferentes valores de $\sqrt{n}(\mathcal{T}_2(P_n, Q_n))$ en el eje de ordenadas y el tamaño de la muestra n en el eje de abscisas.

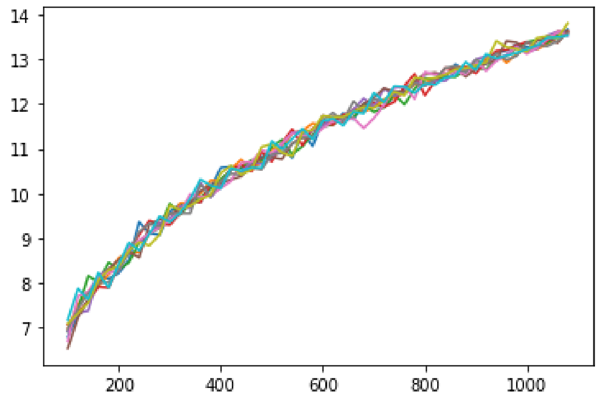

Figure 1.1: Eje de x, tamaño de la muestra n; eje de y, diferencia $\sqrt{n}(\mathcal{T}_2(P_n, Q_n))$, para $P, Q \sim \mathcal{U}_{[0,1]^{10}}$. El experimento se repite 10 veces independientes y los resultados interpolados se representan en diferentes colores.

La observación heurística resulta ser bastante esclarecedora: la tasa de convergencia del sesgo no puede ser la tasa paramétrica $\sqrt{n}$, sin embargo las fluctuaciones $\sqrt{n}(\mathcal{T}_2(P_n, Q_n) - \mathbb{E}\mathcal{T}_2(P_n, Q_n))$ parecen acotadas. Es decir, aparentemente las fluctuaciones no se ven influidas por el hecho de tener una dimensión considerablemente alta ($d = 10$). En efecto, para medidas absolutamente continuas con respecto a la medida de Lebesgue d-dimensional ℓ_d, la tasa de convergencia correcta del sesgo depende de la dimensión del espacio (véase, por ejemplo, Weed and Bach (2019)). Veremos más adelante que es lo que ocurre con las fluctuaciones.

1.1.1 Análisis de las fluctuaciones

Para el análisis de las fluctuaciones (varianza) suponemos que $c(\mathbf{x}-\mathbf{y}) = h(\mathbf{x}-\mathbf{y})$ satisface las condiciones de Gangbo and McCann (1996), P y Q son absolutamente continuas con respecto a la medida de Lebesgue y

$$\int h(2\mathbf{x})^2 dP(\mathbf{x}) < \infty \text{ and } \int h(-2\mathbf{y})^2 dQ(\mathbf{y}) < \infty.$$

El capítulo 2 muestra que

$$\sqrt{n}\left(\mathcal{T}_c(P_n, Q) - \mathbb{E}\mathcal{T}_c(P_n, Q)\right) \xrightarrow{w} N(0, \sigma_c^2(P, Q)), \tag{1.5}$$

donde

$$\sigma_c^2(P, Q) := \int \varphi(\mathbf{x})^2 dP(\mathbf{x}) - \left(\int \varphi(\mathbf{x}) dP(\mathbf{x})\right)^2, \tag{1.6}$$

siendo φ un potencial de transporte óptimo de P a Q para el coste c. En el caso $P = Q$ el potencial de transporte óptimo es constante, por lo que su varianza es nula y el límite (1.5) es degenerado. Volviendo al ejemplo simulado anteriormente (figura 1.1), esto implica que las fluctuaciones no sólo están controladas, sino que tienden en probabilidad a 0. La tasa de convergencia de la fluctuación bajo $P = Q$ sigue siendo un problema abierto.

La demostración utiliza la desigualdad de Efron-Stein (Boucheron et al., 2013, capítulo 3.1), i.e. si $(\mathbf{X}_1', \dots, \mathbf{X}_n')$ y $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ son i.i.d, fijemos $Z := f(\mathbf{X}_1, \dots, \mathbf{X}_n)$ y para cada $i \in \{1, \dots, n\}$ denotemos $Z_i' := f(\mathbf{X}_1, \dots, \mathbf{X}_{i-1}, \mathbf{X}_i', \mathbf{X}_{i+1}, \dots, \mathbf{X}_n)$, entonces

$$Var(Z) \leq \frac{n}{2} E(Z - Z_i')^2 = nE(Z - Z_i')_+^2.$$

donde $(\cdot)_+$ denota la parte postiva. En nuestro problema de transporte, la desigualdad de Efron-Stein implica

$$nVar(\mathcal{T}_c(P_n, Q) - \int \varphi(\mathbf{x}) dP_n(\mathbf{x})) \leq \mathbb{E}\big(\varphi_n(X_1) - \varphi(X_1) - \varphi_n(X_1') + \varphi(X_1')\big)^2, \tag{1.7}$$

ergo $\sqrt{n}\left(\mathcal{T}_c(P_n, Q) - \mathbb{E}\mathcal{T}_c(P_n, Q)\right)$ es ajustado, i.e. admite límites débiles a través de subsucesiones. Los límites se obtienen como consecuencia de la convergencia de los potenciales empíricos hacia su homólogo poblacional, que debe ser único: El corolario 2.2.7 da la unicidad (salvo constantes aditivas) de los potenciales para probabilidades con soporte conectado y el teorema 2.3.4 la estabilidad de la aplicación de transporte. Extraemos aquí el contenido del teorema 2.3.4 debido a su importancia–es el primer resultado que muestra la estabilidad de la aplicación de transporte óptimo y del potencial en dominios no limitados y para costes generales.

Theorem 1.1.1. *Sea $Q \in \mathcal{P}(\mathbb{R}^d)$ tal que $Q \ll \ell_d$ y con soporte conexo y frontera despreciable. Se asume que $Q_n, P_n, P \in \mathcal{P}(\mathbb{R}^d)$ son tales que $P_n \xrightarrow{w} P$, $Q_n \xrightarrow{w} Q$ y*

$$\mathcal{T}_c(P_n, Q_n) < \infty \text{ and } \mathcal{T}_c(P, Q) < \infty$$

para el coste $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$, con h diferenciable y satisfaciendo (A1)-(A3). *Si ψ_n (resp. ψ) es un plan de transporte óptimo de Q_n a P_n (resp. de Q a P) para el coste c. Entonces:*

(i) Existen constantes $a_n \in \mathbb{R}$ *tales que* $\tilde{\psi_n} := \psi_n - a_n \to \psi$ *en la topología de convergencia en los subconjuntos compactos de* $\mathrm{Supp}(Q)$.

(ii) Para cada compacto $K \subset \mathrm{Supp}(Q) \cap \mathrm{dom}(\nabla\psi)$

$$\sup_{\mathbf{x}\in K} \sup_{\mathbf{y}_n \in \partial^c \psi_n(\mathbf{x})} \|\mathbf{y}_n - \nabla^c \psi(\mathbf{x})\| \longrightarrow 0.$$

Sin embargo, este resultado sólo proporciona la convergencia casi seguro de los potenciales empíricos hacia su contraparte poblacional, pero no se puede concluir de esto que (1.7) tiende a 0. Para evitar más suposiciones (como en del Barrio and Loubes (2019)), se puede demostrar que la secuencia $\mathbb{E}\big|\varphi_n(X_1) - \varphi(X_1) - \varphi_n(X_1') + \varphi(X_1')\big|$ está acotada y entonces el teorema de Banach-Alaoglu produce una convergencia débil en $L^2(\mathbb{P})$ de $\big|\varphi_n(X_1) - \varphi(X_1) - \varphi_n(X_1') + \varphi(X_1')\big|$ a 0 a través de subsucesiones. Tomando medias de Cesàro podemos pasar de la convergencia débil a la fuerte.

1.2 Caso semidiscreto

El comportamiento del sesgo es muy diferente en el caso semidiscreto –cuando una de las dos probabilidades está soportada en un conjunto finito de puntos $\mathbb{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$. Aquí, en el capítulo 3 obtenemos

$$\sqrt{n}\,(\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q)) \xrightarrow{w} \sup_{\mathbf{z}\in \mathrm{Opt}_c(P,Q)} \mathbb{G}(\mathbf{z}),$$

donde $\mathrm{Opt}_c(P, Q)$ es el conjunto de potenciales de transporte óptimos y $\mathbb{G}$ un proceso gaussiano que actúa sobre ellos. Resulta entonces evidente que la unicidad (excepto constantes aditivas) de los mismos implica que el límite es gaussiano. Además, bajo esta hipótesis y gracias al trabajo de Cárcamo et al. (2020)[2], sabemos que el bootstrap es consistente como aproximación del proceso límite. El hecho de que la maldición de la dimensionalidad parece no afectar al caso semidiscreto para ambas probabilidades es bastante sorprendente. Pero está parcialmente oculto en la suposición de que el soporte tiene un tamaño finito fijo. Para una mejor comprensión, proporcionamos el límite superior

$$E\,|\mathcal{W}_1(P, Q_m) - \mathcal{W}_1(P, Q)| \le \frac{8\sqrt{2N}}{\sqrt{m}} K(\mathrm{diam}(\mathbb{X}), Q)$$

donde

$$K(\mathrm{diam}(\mathbb{X}), Q) = (4\,\mathrm{diam}(\mathbb{X}) + 2\sqrt{\int |\mathbf{y}|^2 dQ(\mathbf{y})} + 2\,\mathrm{diam}(\mathbb{X}))\left(\log(2) + \sqrt{2\,\mathrm{diam}(\mathbb{X}) + 1}\right)$$

[2] A título personal, agradezco a Luis Alberto Rodríguez, uno de los autores de Cárcamo et al. (2020), las discusiones llevadas a cabo en su estancia en Toulouse.

y $\mathcal{W}_1$ es la distancia 1-Wasserstein para el coste euclídeo.

Las pruebas de los capítulos 2 y 3 se basan en argumentos completamente diferentes; el límite de fluctuaciones utiliza el procedimiento introducido por del Barrio and Loubes (2019) basado en la desigualdad de Efron-Stein, mientras que el problema semidiscreto requiere la derivación funcional en el sentido de Hadamard del coste de transporte. El lema 3.1.1 muestra que $\mathcal{T}_c(P, Q)$ es equivalente a la maximización (sobre una clase paramétrica) de

$$g_c(P, Q, \mathbf{z}) = \sum_{i=1}^{N} z_i p_i + \int \inf_{i=1,\ldots,N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\} dQ(\mathbf{y}). \tag{1.8}$$

Desde luego, de la optimización sobre una clase paramétrica se obtiene un ratio de convergencia paramétrico. La solución $\mathbf{z}^*$, única salvo constantes aditivas en algunas situaciones, de (1.8) define las celdas de Laguerre

$$\mathrm{Lag}_k(\mathbf{z}^*) := \{\mathbf{y} \in \mathbb{R}^d : \ c(\mathbf{x}_k, \mathbf{y}) - z_k < c(\mathbf{x}_i, \mathbf{y}) - z_i^* \text{ for all } i \neq k\}, \quad k = 1, \ldots, N,$$

que son generalizaciones de las celdas de Voronoi–equivalentes a $\mathrm{Lag}_k(\mathbf{0})$ para el coste cuadrático.

El transporte óptimo semidiscreto tiene aplicaciones al *«modelo de localización de Hotelling»*. (p. ej., (Galichon, 2016, capítulo 5.1)). Esta aplicación es un ejemplo típico de problema socioeconómico en el que la ubicación de cierta población está representada por una medida de probabilidad continua Q y las fuentes –negocios que intentan vender un producto– como una probabilidad discreta P. Aquí la ubicación de la fuente i es $\mathbf{x}_i$ y la capacidad es p_i. Cada habitante elegiría la fuente a la vez más cercana y que ofrezca un mejor precio, es decir, la estrategia $\arg\inf_{i=1,\ldots,N}\{c(\mathbf{x}_i, \mathbf{y}) - z_i\}$, donde z_i representa el precio de la fuente i. El conjunto de población que prefiere consumir de la fuente i es en realidad $\mathrm{Lag}_k(\mathbf{z})$. Cuando el mercado está equilibrado –la oferta es igual a la demanda–, cada fuente utiliza a su máxima capacidad y el problema de determinar los precios se reduce a la solución de un problema de transporte óptimo semidiscreto.

EL capítulo 3 proporciona, según tenemos entendido, el primer teorema central del límite para las soluciones del problema dual (1.8) Subrayamos que este resultado no puede generalizarse para distribuciones continuas. En efecto, si ambas probabilidades son continuas y el espacio es de dimensión $d > 4$, no podemos esperar resultados similares, ya que, el valor esperado de la estimación del coste de transporte converge con una tasa, a lo sumo de $O(n^{-\frac{2}{d}})$ y ya no $O(n^{-\frac{1}{2}})$. Cuando las dos muestras son discretas, aunque dicha tasa sea $O(n^{-\frac{1}{2}})$, la falta de unicidad del problema dual no permite probar este tipo de resultados. En consecuencia, el semidiscreto (o tal vez, otros casos en el que una de las distribuciones viva en un espacio de dimensión suficientemente pequeña) es el único caso en el que se pueden esperar tales resultados para los potenciales de transporte en dimensión general.

Como consecuencia de los límites de los potenciales, en el capítulo 3 se proporcionan los límites débiles para las celdas de Laguerre en términos de la métrica L^p (cf. Vitale (1985) por ejemplo) entre las celdas empíricas y poblacionales. Además, en términos de la distancia de Hausdorff (métrica L^∞), veremos como dar regiones de confianza asintótica de las celdas. En todos los casos, se alcanza la velocidad de convergencia paramétrica.

1.3 Transporte óptimo regularizado por la entropía

La influencia de la dimensión es alta en la estimación del coste del transporte a partir de los datos. Encontrar los límites débiles del coste de transporte es un trabajo arduo y poco fructífero incluso para dimensiones moderadas y, aunque se encontraran, el ratio sería lento. Es decir, se necesitarían demasiados datos para aproximar la distribución límite. Como consecuencia, pruebas de bondad de ajuste basada en distancia Wasserstein son prácticamente inabordables en dimensión general. Se necesita, por tanto, otra noción de discrepancia entre distribuciones que, a la vez que capture la geometría del espacio, tenga una aproximación empírica que converja a su homólogo poblacional con la tasa paramétrica $\sqrt{n}$.

Con una motivación computacional, el influyente artículo Cuturi (2013) propone una regularización de (1.2) añadiendo un termino de penalización, la entropía relativa:

$$S_\epsilon(P,Q) = \min_{\pi \in \Pi(P,Q)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \tfrac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 d\pi(x,y) + \epsilon H(\pi|P \times Q), \tag{1.9}$$

donde H denota la entropía relativa, definida, para dos medidas de probabilidad α y β, como o $H(\alpha|\beta) = \int \log(\frac{d\alpha}{d\beta}(\mathbf{x}))d\alpha(\mathbf{x})$ si α es absolutamente continua con respecto a β y $+\infty$ en caso contrario. Sea $\pi^\epsilon_{P,Q}$ la solución de (1.9), que es absolutamente continua con respecto a $P \times Q$ con densidad $\xi^\epsilon_{P,Q}$. Este problema, así como su homólogo no regularizado (1.2), admite una formulación dual.

$$\begin{aligned} S_\epsilon(P,Q) = \sup_{f \in L_1(P), g \in L_1(Q)} \Big\{ & \int_{\mathbb{R}^d} f(\mathbf{x})dP(\mathbf{x}) + \int_{\mathbb{R}^d} g(\mathbf{y})dQ(\mathbf{y}) \\ & - \epsilon \int_{\mathbb{R}^d \times \mathbb{R}^d} e^{\frac{f(\mathbf{x})+g(\mathbf{y})-\frac{1}{2}\|x-y\|^2}{\epsilon}} dP(\mathbf{x})dQ(\mathbf{y}) + \epsilon \Big\} \end{aligned} \tag{1.10}$$

La solución de (1.10) es un par formado por dos funciones, denotémoslo como $(f^\epsilon_{P,Q}, g^\epsilon_{P,Q})$. Sus componentes satisfacen las siguientes *condiciones de optimalidad*:

$$\begin{aligned} \int e^{f_{P,Q}(\mathbf{x})+g_{P,Q}(\mathbf{y})-\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2} dQ(y) = 1, \quad \text{for all } \mathbf{x} \in \mathbb{R}^d, \\ \int e^{f_{P,Q}(x)+g_{P,Q}(y)-\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2} dP(x) = 1, \quad \text{for all } \mathbf{y} \in \mathbb{R}^d. \end{aligned} \tag{1.11}$$

Desde luego que podemos despejar $f^\epsilon_{P,Q}$ en (1.11) y derivar bajo el signo integral. Por lo tanto, cada elemento del par (1.10) tiene un representante en la clase $\mathcal{C}^s$. Esto reduce la complejidad de el problema de optimización (véase Genevay et al. (2019)) y *a fortiori* se obtiene la cota $\sqrt{n}\,\mathbb{E}|S_\epsilon(P_n, Q) - S_\epsilon(P, Q)| \leq C$, válida para medidas subgaussianas Mena and Niles-Weed (2019).

Además, el argumento de del Barrio and Loubes (2019) proporciona también los límites de las fluctuaciones (Mena and Niles-Weed (2019))

$$\sqrt{n}(S_\epsilon(P_n, Q) - \mathbb{E}S_\epsilon(P_n, Q)) \xrightarrow{w} N(0, \mathrm{Var}_{\mathbf{X}\sim P}(f^\epsilon_{P,Q}(\mathbf{X}))). \tag{1.12}$$

Tanto (1.12) como el límite anteriormente mencionado no proporcionan intervalos de confianza asintóticamente válidos para el valor poblacional $S_\epsilon(P, Q)$). Para ello, es necesario conocer el límite exacto de

$$\sqrt{n}\,(\mathbb{E}S_\epsilon(P_n, Q) - S_\epsilon(P, Q)). \tag{1.13}$$

En el capítulo 4 se demuestra que (1.13) tiende a 0. De modo que el valor esperado puede ser sustituido por el valor poblacional en (1.12), dando como resultado

$$\sqrt{n}(S_\epsilon(P_n, Q) - S_\epsilon(P, Q)) \xrightarrow{w} N(0, \mathrm{Var}_{\mathbf{X}\sim P}(f^\epsilon_{P,Q}(\mathbf{X}))). \tag{1.14}$$

Sea z_α el cuantil de α para la distribución normal estándar y

$$\hat{\sigma}^2_n := \mathrm{Var}_{P_n}(f^\epsilon_{P_n,Q}) = \frac{1}{n}\sum_{i=1}^{n}(f^\epsilon_{P_n,Q}(\mathbf{X}_i))^2 - \Big(\frac{1}{n}\sum_{i=1}^{n} f^\epsilon_{P_n,Q}(\mathbf{X}_i)\Big)^2 \tag{1.15}$$

una estimación consistente de la varianza límite $\mathrm{Var}_{\mathbf{X}\sim P}(f^\epsilon_{P,Q}(\mathbf{X}))$, el conjunto

$$\left[S_\epsilon(P_n, Q) \pm \frac{\hat{\sigma}_n}{\sqrt{n}} z_{1-\alpha/2}\right], \tag{1.16}$$

es, por lo tanto, un intervalo de confianza asintótico de nivel α.

Sin embargo, (1.16) no proporciona ninguna consecuencia estadísticamente significativa debido al sesgo entrópico, ¡el hecho de que $S_\epsilon(P, Q) = 0$ no significa que ambas probabilidades sean iguales! La hipótesis $P = Q$ no puede aceptarse ni rechazarse mediante el valor de $S_\epsilon(P, Q)$. La solución, por lo tanto no es totalmente satisfactoria.

La solución más exitosa para cercenar la influencia de la regularización de la entropía es la *divergencia de Sinkhorn*, propuesta por Feydy et al. (2019) y que se expresa como

$$D_\epsilon(P, Q) = S_\epsilon(P, Q) - \tfrac{1}{2}\left(S_\epsilon(P, P) + S_\epsilon(Q, Q)\right).$$

Claramente $D_\epsilon(P, Q)$ es simétrica en P, Q y $D_\epsilon(P, P) = 0$. Moreover, $D_\epsilon(P, Q) \geq 0$, con $D_\epsilon(P, Q) = 0$ si y solamente si $P = Q$ (Feydy et al., 2019, teorema 1).

Los intervalos de confianza asintóticos para $SD_\epsilon(P,Q)$ pueden inferirse de su contrapartida empírica conociendo los límites no triviales de

$$a_n(D_\epsilon(P_n,Q) - D_\epsilon(P,Q)),$$

para una sucesión real $\{a_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$. El capítulo 5 demuestra que la secuencia en cuestión depende de la hipótesis. En particular, obtenemos los siguientes límites:

- Bajo $H_0: \ P = Q$:

$$n\, D_\epsilon(P_n,P) \xrightarrow{w} \frac{1}{4}\sum_{j=1}^{\infty}\lambda_j\left(\sum_{i=1}^{\infty} x_{i,j}N_i\right)^2,$$

  donde $\{N_i\}_{i\in\mathbb{N}}$ es una sucesión de variables aleatorias mutuamente independientes siguiendo una ley normal estándar (i.e. $N_i \sim N(0,1)$) y $\{x_{i,j}\}_{i,j\in\mathbb{N}} \subset \mathbb{R}$, $\{\lambda_j\}_{j\in\mathbb{N}} \subset [0,\infty)$ sucesiones reales (y deterministas) dependiendo de P y ϵ.

- Bajo $H_1: \ P \neq Q$:

$$\sqrt{n}(D_\epsilon(P_n,Q) - D_\epsilon(P,Q)) \xrightarrow{w} N(0, \mathrm{Var}_P(\psi^\epsilon_{P,Q})),$$

  donde $\psi^\epsilon_{P,Q} = f^\epsilon_{P,Q} - \frac{1}{2}(f^\epsilon_{P,P} + g^\epsilon_{P,P})$.

La demostración no usa las mismas técnicas en cada hipótesis. Bajo H_1 la desigualdad de Efron-Stein proporciona el desarrollo de primer orden. Sin embargo, dado que $\psi^\epsilon_{P,Q}$ es cero cuando $P = Q$, esto no es suficiente para obtener límites no triviales bajo H_0. Así que es necesario un desarrollo de segundo orden que, a su vez, es una consecuencia del desarrollo de primer orden de los potenciales–soluciones del problema dual regularizado. La demostración es larga y tediosa; se divide en dos trabajos diferentes. Por un lado, el capítulo 4 muestra la equicontinuidad de los potenciales y la divergencia. Es decir, que existe una constante c_d, que depende sólo de d, tal que

$$\mathbb{E}\|g_{P_n,Q} - g_{P,Q}\|^2_{\mathcal{C}^s(\Omega)},\ \mathbb{E}\|f_{P_n,Q} - f_{P,Q}\|^2_{\mathcal{C}^s(\Omega)} \leq \tfrac{c_d}{n} D_\Omega^{5(d+1)} e^{15D_\Omega^2},$$

y

$$\mathbb{E}D_1(P_n,P) \leq \frac{c_d}{n} D_\Omega^{\frac{3d}{2}+1}\frac{32}{(d+1)^2} e^{\frac{19}{2}D_\Omega^2}.$$

Por otro lado, el capítulo 5 muestra el desarrollo de primer orden de $\sqrt{n}\begin{pmatrix} f^\epsilon_{P_n,Q} - f^\epsilon_{P,Q} \\ g^\epsilon_{P_n,Q} - g^\epsilon_{P,Q}\end{pmatrix}$, in $\mathcal{C}^\alpha(\Omega)\times\mathcal{C}^\alpha(\Omega)$, con respecto al proceso empírico $\sqrt{n}(P_n - P)$:

$$\begin{pmatrix} f^\epsilon_{P_n,Q} - f^\epsilon_{P,Q} \\ g^\epsilon_{P_n,Q} - g^\epsilon_{P,Q}\end{pmatrix} = \begin{pmatrix} (1 - \mathcal{A}^\epsilon_Q\mathcal{A}^\epsilon_P)^{-1}\mathcal{A}^\epsilon_Q\mathbb{G}^n_{P,s} \\ -(1 - \mathcal{A}^\epsilon_P\mathcal{A}^\epsilon_Q)^{-1}\mathbb{G}^n_{P,s}\end{pmatrix} + o_P\left(\frac{1}{\sqrt{n}}\right).$$

Aquí $\mathbb{G}^n_{P,s}$ denota $\frac{1}{n}\sum_{k=1}^n \xi^{\epsilon}_{P,Q}(\mathbf{X}_k,\cdot) - \mathbb{E}\left(\xi^{\epsilon}_{P,Q}(\mathbf{X},\cdot)\right)$, y

$$\begin{aligned} \mathcal{A}^{\epsilon}_P &: L^2(P) \ni f \mapsto \int \xi^{\epsilon}_{P,Q}(\mathbf{x},\cdot) f(\mathbf{x}) dP(\mathbf{x}) \in \mathcal{C}^{\alpha}(\Omega), \\ \mathcal{A}^{\epsilon}_Q &: L^2_0(Q) \ni g \mapsto \int \xi^{\epsilon}_{P,Q}(\cdot,\mathbf{y}) g(\mathbf{y}) dQ(\mathbf{y}) \in \mathcal{C}^{\alpha}(\Omega). \end{aligned} \tag{1.17}$$

Este desarrollo de primer orden de los potenciales permite, mediante la fórmula

$$d\pi^{\epsilon}_{P,Q} = e^{\frac{f_{P,Q}+g_{P,Q}-\frac{1}{2}\|\cdot-\cdot\|^2}{\epsilon}} dPdQ,$$

obtener los límites débiles de las soluciones del problema primario, es decir

$$\sqrt{n}\int \eta\,(d\pi^{\epsilon}_{P_n,Q} - d\pi^{\epsilon}_{P,Q}), \quad \text{with } \eta \in L^2(P\times Q).$$

En particular, obtenemos

$$\sqrt{n}\left(\int \eta d\pi^{\epsilon}_{P_n,Q} - \int \eta d\pi^{\epsilon}_{P,Q}\right) \xrightarrow{w} N(0,\sigma^2_{\lambda,\epsilon}(\eta)), \quad \eta \in L^2(P\times Q), \tag{1.18}$$

donde $\sigma^2_{\lambda}(\eta) = \mathrm{Var}_{\mathbf{X}\sim P}\left((1-\mathcal{A}^{\epsilon}_Q\mathcal{A}^{\epsilon}_P)^{-1}(\eta^{\epsilon}_{\mathbf{x}} - \mathcal{A}^{\epsilon}_Q\eta^{\epsilon}_{\mathbf{y}})(\mathbf{X})\right)$. Esto confirma la veracidad de la conjetura de Harchaoui et al. (2020). (1.18) proporciona intervalos de confianza consistentes para $\pi^{\epsilon}_{P_n,Q}$, lo que permite, entre otras cosas, la posibilidad de realizar inferencia sobre la distancia de Sinkhorn (aplicar (1.18) a la función $(\mathbf{x},\mathbf{y}) \to \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2$) definida en Cuturi (2013)o la medida de colocalización regularizada RCol (aplicar (1.18) a la función $(\mathbf{x},\mathbf{y}) \to \mathbb{1}_{\|\mathbf{x}-\mathbf{y}\|^2\le t}$) definida en Klatt et al. (2020).

1.4 Función de distribución multivariante centro-exterior; regularidad y regresión cuantílica.

La falta de orden canónico de $\mathbb{R}^d$, para $d \geq 2$, impide la generalización de herramientas estadísticas basadas en la relación de orden univariante. Una de las más importantes es la función de distribución de una variable aleatoria univariante $X \sim P$, definida como $F : x \to \mathbb{P}(X \leq x)$.

Se han hecho muchos intentos de definir una función de distribución multivariante. Entre ellos se encuentran los basados en profundidades, en cópulas, en rangos de componentes, en rangos espaciales o en rangos de Mahalanobis; nos remitimos a Hallin et al. (2021a) y a las referencias que contiene. Todas ellas tienen que lidiar con la mencionada ausencia de orden canónico, pero ninguna de ellas es capaz de imitar las propiedades que hacen útil la función de distribución univariante. Es decir, la libertad de distribución (la distribución de $F(X)$ es uniforme en $[0,1]$ mientras que su contraparte empírica $\{F^n(X_i)\}_{i=1}^n$ es uniforme en el conjunto $\{\frac{i}{n}\}_{i=1}^{n}$, independientemente de la distribución de X_i; satisface la propiedad Glivenko-Cantelli (la aproximación empírica converge a. s. uniformemente en $\mathbb{R}$) y mantiene el orden, es decir, es monótona.

La monotonicidad y el hecho de que $F(X)$ sea uniforme en $[0, 1]$ definen de forma única a F, para cualquier variable aleatoria continua $X \sim P$. La monotonicidad es equivalente a ser el gradiente de una función convexa, la libertad de distribución de F, a el hecho de que $F_{\#}P = U_{(0,1)}$ (la distribución uniforme en $(0, 1)$). Por supuesto, la función cuantil Q es el gradiente de una función convexa que empuja $U_{(0,1)}$ hacia P. Estas propiedades llevaron a Chernozhukov et al. (2017) a definir una función cuantil multivariante, la llamada función de distribución cuantil con respecto a una medida de referencia $\mathcal{U}$, definida, para una variable aleatoria $\mathbf{X} \sim P \in \mathcal{P}(\mathbb{R}^d)$, como el único gradiente de una función convexa (definida en casi todos los puntos) que empuja $\mathcal{U}$ hacia P. El célebre teorema de McCann McCann (1995) garantiza su existencia.

El debate actual es sobre la elección de una medida de referencia $\mathcal{U}$. Deb and Sen (2019) aboga por una medida uniforme en el hipercubo (con éxito para probar la independencia), Hallin et al. (2021a) por una uniforme esférica (con éxito para dar una noción de orden centro-exterior) y Deb et al. (2021) por una gaussiana estándar (con éxito para la prueba T^2 de Hotelling). La respuesta más correcta es que depende del propósito (Hallin and Mordant, 2021). Aquí trataremos el caso de la distribución esférica sobre la bola unitaria $\mathbf{U}_d$, es decir, la que se obtiene tomando independientemente un radio $r \sim U_{(0,1)}$ y una dirección θ uniforme sobre la esfera $\mathcal{S}^{d-1}$. La ventaja de utilizar la uniforme esférica como medida de referencia radica en su invariancia a las rotaciones (cambios de base) y su control uniforme de las bolas centradas en el origen y de radio $\tau \in (0, 1)$, es decir

$$\mathbf{U}_d(\tau \mathbb{S}^d) = \tau \quad \tau \in (0, 1). \tag{1.19}$$

Ninguna otra medida de probabilidad en $\mathbb{R}^d$ satisface estas propiedades. (1.19) proporciona una relación orden radial natural y una noción clara de centro–por lo tanto un ordenamiento del centro hacia afuera. La función cuantil $\mathbf{Q}_{\pm}$ se define en este caso como el gradiente único de una función convexa ψ (definida en casi todos los puntos) que empuja $\mathbf{U}_d$ hacia P. La función de distribución se define como

$$\mathbf{F}_{\pm}(\mathbf{x}) := \arg \sup_{\|\mathbf{u}\| \leq 1} \{\langle \mathbf{x}, \mathbf{u} \rangle - \psi(\mathbf{u})\},$$

que coincide con la inversa de $\mathbf{Q}_{\pm}$ en el soporte de P. La importancia dada por $\mathbf{U}_d$ al origen y esa noción de ordenación centro-exterior son heredadas por P a través de $\mathbf{Q}_{\pm}$. Esto da lugar al nombre *distribución centro-exterior y aplicación cuantil* acuñado en Hallin et al. (2021a) para $\mathbf{F}_{\pm}$ y $\mathbf{Q}_{\pm}$. Esta terminología se utilizará a partir de ahora.

Supongamos la observación de una muestra i.i.d. $\mathbf{X}^{[n]} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ de P, su distribución empírica centro-exterior $\mathbf{F}_{\pm}^{n}$ es definida, para cada $\mathbf{X_i}$, como $\mathbf{F}_{\pm}^{n}(\mathbf{X_i}) = \mathfrak{G}_{\sigma_n(i)}^{(n)}$.

Aquí σ_n es la permutación de $\{1,\ldots,n\}$ donde

$$\sum_{i=1}^{n} \|\mathbf{X}_i - \mathfrak{G}_{\sigma(i)}^{(n)}\|^2 \tag{1.20}$$

alcanza su mínimo, y $\mathfrak{G}_1^{(n)},\ldots,\mathfrak{G}_n^{(n)}$ es una malla regular de la distribución esférica uniforme $\mathbb{U}_d$. Consiste en factorizar $n = n_s n_d + n_0$,with $n < \min(n_s, n_d)$ y computar n puntos $\mathfrak{G}_1^{(n)},\ldots,\mathfrak{G}_n^{(n)}$ creados como la intersección entre

– los rayos de la s_S-tupla $\mathbf{u}_1,\ldots,\mathbf{u}_{N_S} \in \mathcal{S}_{d-1}$ de vectores unidad tal que ${N_S}^{-1}\sum_{j=1}^{N_S}\delta_{\mathbf{u}_j}$ converge a la uniforme sobre $\mathcal{S}_{d-1}$ cuando $n_S \to \infty$, y
– las N_R hiper-esferas con centro $\mathbf{0}$ y radios $j/(N_R+1)$, $j = 1,\ldots,N_R$,

los puntos correspondientes a n_0 son identificados con $\mathbf{0}$. La propiedad de Glivenko-Cantelli

$$\max_{1\le k\le n} \|\mathbf{F}_\pm^n(\mathbf{X}_k) - \mathbf{F}_\pm(\mathbf{X}_k)\| \xrightarrow{a.s.} 0 \tag{1.21}$$

es válida si $\mathbf{F}_\pm$ está bien definida en todo el espacio (recordemos que por el momento solo sabemos que lo está excepto en un conjunto despreciable). Esto es algo bastante común cuando se trata de probar la convergencia de gradientes de funciones convexas. En particular, el caso del transporte óptimo requiere tres probar las siguientes afirmaciones; la convergencia débil de los planes de transporte; la convergencia de los potenciales; finalmente, la convergencia de los sub-diferenciales en el sentido de conjuntos. Cuando la función es regular el límite es uniforme en los compactos, la forma particular del soporte de $\mathbb{U}_d$, permite extender la regularidad al supremo en $\mathbb{R}^d$.

Si $\mathbf{F}_\pm$ está bien definida en todo el espacio, como consecuencia de (1.19), las funciones cuantil de centro-exterior definen regiones cerradas encajadas $\mathbb{C}_P(\tau) = \mathbf{Q}_\pm(r\,\mathbb{S}^d)$ y contornos continuos $\mathcal{C}_P(\tau) = \mathbf{Q}_\pm(r\,\mathcal{S}^{d-1})$ indexados por $\tau \in ([0,1)$, tales que, para cualquier P absolutamente continuo, $P\big[\mathbb{C}_P(\tau)\big] = \tau$, independientemente de P. Ninguna otra elección de medida de referencia define una función de distribución multivariante basada en el transporte que comparta la última propiedad, que es la clave de la regresión cuantílica multivariante (véase el capítulo 7).

1.4.1 Regularidad de la función de distribución centro-exterior

La función de distribución centro-exterior $\mathbf{F}_\pm$ está bien definida en todas partes si y sólo si es continua (cf. Rockafellar (1970)). Por lo tanto, el estudio de su regularidad es importante. Se sabe que en una dimensión la función de distribución es continua para todas las probabilidades absolutamente continuas (con respecto a la medida de Lebesgue). Sería deseable obtener resultados similares para su homóloga multivariante. La regularidad de las funciones convexas que satisfacen una condición de transporte suele tratarse por medio de la ecuación de Monge-Amp‘ere (cf. Figalli (2017) por ejemplo). Veremos que en este caso la continuidad de $\mathbf{F}_\pm$ se consigue suponiendo que P tiene una densidad p que se satisface

Hipótesis A. Para cualquier $R > 0$, existe $0 < \lambda_R \leq \Lambda_R < \infty$ tal que

$$\lambda_R \leq p(\mathbf{x}) \leq \Lambda_R \quad \text{para todo } \mathbf{x} \in \mathcal{X} \cap R\,\mathbb{S}^d. \tag{1.22}$$

En particular, $\mathbf{F}_\pm$ satisface la condición de empujar P hacia Q y es el gradiente de una función convexa φ, por lo que su subdiferencial[3] satisface

$$\int_{\partial\varphi(A)} u_d(\mathbf{y})d\mathbf{y} = \int_A p(\mathbf{x})d\mathbf{x}. \tag{1.23}$$

Esta propiedad permite estudiar la continuidad de la función de distribución centro-exterior mediante la conocida ecuación de Monge-Amp'ere –una ecuación diferencial no lineal basada en el determinante de la derivada de segundo orden. El principal resultado del capítulo 6 es el siguiente.

Theorem 1.4.1. *Sea P una probabilidad con densidad p soportada en el conjunto convexo $\mathcal{X} \subseteq \mathbb{R}^d$.*

(i) *Si p satisface (1.23), entonces $K := \partial\psi(\mathbf{0})$ es un conjunto compacto y convexo con medida de Lebesgue 0 tal que $\mathbf{Q}_\pm := \nabla\psi$ y $\mathbf{F}_\pm := \nabla\psi^*$ son homeomorfismos entre $\mathbb{S}^d \setminus \{\mathbf{0}\}$ y $\mathcal{X} \setminus K$, inversos entre sí .*

(ii) *Si, además, $p \in \mathcal{C}^{k,\alpha}_{\text{loc}}(\mathcal{X})$ para cierto $k \in \mathbb{N}$ y $\alpha \in (0,1)$, entonces $\mathbf{Q}_\pm$ y $\mathbf{F}_\pm$ son difeomorfismos de clase $\mathcal{C}^{k+1,\alpha}_{\text{loc}}$ entre $\mathbb{S}^d \setminus \{\mathbf{0}\}$ y $\mathcal{X} \setminus K$.*

La demostración se basa en la teoría de Caffarelli Caffarelli (1990, 1991, 1992) que ha sido ampliamente estudiada por Figalli en Figalli and Kim (2010); Philippis and Figalli (2012); Cordero-Erausquin and Figalli (2019). Se recomienda el libro Figalli (2017) como guía introductoria a este tópico y sus implicaciones a la regularidad del transporte. Hasta Figalli (2018), la regularidad de $\mathbf{F}_\pm$ no estaba cubierta por ninguno de estos trabajos, véase Cordero-Erausquin and Figalli (2019) para un estudio del caso más general. Sin embargo las hipótesis de Figalli (2018) (se asume **A.** y que P está soportada en todo el espacio) son demasiado fuertes, en vista del teorema 1.4.1, pueden ser relajadas .

Como consecuencia, si P es una probabilidad con densidad p soportada en el conjunto convexo $\mathcal{X} \subseteq \mathbb{R}^d$ tal que p satisface (1.23), las siguientes propiedades se satisfacen.

- La propiedad de Glivenko-Cantelli (1.21) (cf. (Hallin et al., 2021a, Proposición 2.4)) es cierta. Lo cual implica que la función de distribución del centro hacia fuera puede ser uniformemente estimada.

- Para todo $\mathbf{u}$ en la esfera unidad $\mathcal{S}^{d-1}$, toda sucesión $(t_n)_{n\in\mathbb{N}}$ de números reales tal que $t_n \to \infty$, se verifica el límite

$$\lim_{n\to\infty} \mathbf{F}_\pm(t_n\mathbf{u}) = \mathbf{u}. \tag{1.24}$$

[3]Una generalización del concepto de derivada para funciones no regulares (véase Rockafellar (1970))

Esta propiedad es la análoga a $\lim_{n\to\infty} F(t_n) = 1$ y $\lim_{n\to\infty} F(-t_n) = 0$ en el caso univariante, eso si, desde la perspectiva centro hacia fuera.

- Para todo $r \in (0, 1)$ y todo $\mathbf{y} \in \mathcal{C}_{\mathbf{P}}(\tau)$, existe un rayo T que emana de $\mathbf{y}$ para el cual $\overline{\mathbf{Q}_{\pm}(r\,\mathbb{S}^d)} \cap T = \{\mathbf{y}\}$. En palabras coloquiales, para todo punto de un contorno cuantil existe un rayo que no vuelve a tocar el conjunto. Además $\mathbb{C}_P(\tau)$ son regiones encajadas y conexas.

- Si, además, $\mathcal{X}$ es compacto, para todo $r \to 1$, $\mathbf{Q}_{\pm}(r\,\mathbb{S}^d)$ tiende a $\mathcal{X}$ en distancia de Hausdoff.

$$\lim_{r\to 1} \mathrm{d}_H\big(\mathbf{Q}_{\pm}(r\,\mathbb{S}^d), \mathcal{X}\big) = 0. \tag{1.25}$$

Por lo que el soporte de P puede ser aproximado por los límites de las regiones cuantiles.

1.4.2 Regresión cuantílica no paramétrica de salida multivariante

El capítulo 7 propone una novedosa y significante extensión no paramética y multivariante, basada en el concepto de cuantiles del centro hacia afuera, del célebre concepto de regresión cuantílica de Koenker y Bassett Koenker and Bassett (1978), herramienta poderosa en el estudio estadístico de la dependencia de una variable de interés Y con respecto a las covariables $\mathbf{X} = (X_1, \ldots, X_m)$ que, a diferencia de la regresión clásica que, en cierto modo, se centra en las medias condicionales $\mathrm{E}[Y|\mathbf{X}]$, la regresión cuantílica necesita el conocimiento completo de la distribuciones condicionales $\mathrm{P}_{\mathrm{Y}|\mathbf{X}=\mathbf{x}}$ de Y dado $\mathbf{X} = \mathbf{x}$. Su formulación no paramétrica, cuyo estudio comienza con el pionero trabajo de Stone (1977), se ha convertido en parte de la práctica estadística diaria, con innumerables aplicaciones en todos los ámbitos de la investigación científica en los cuales un número finito de parámetros proporcionan un modelo demasiado rígido como para explicar cierto comportamiento observado.

La motivación principal del uso de la regresión cuantílica es la observación de que el conocimiento de $\mathrm{P}_{\mathrm{Y}|\mathbf{X}=\mathbf{x}}$ es mucho más informativo que el simple cognición de un parámetro de interés, tales como media o mediana condicional. La figura 1.2 muestra un ejemplo de egresión cuantílica univariante, con X =edad e Y =pliegue cutáneo del tríceps. Observando solo la mediana (línea roja) solo seremos capaces de obtener una descripción de la tendencia del modelo, sin embargo, teniendo en cuenta la información dada por los tubos cuantílicos, se obtiene la heterocedasticidad del mismo.

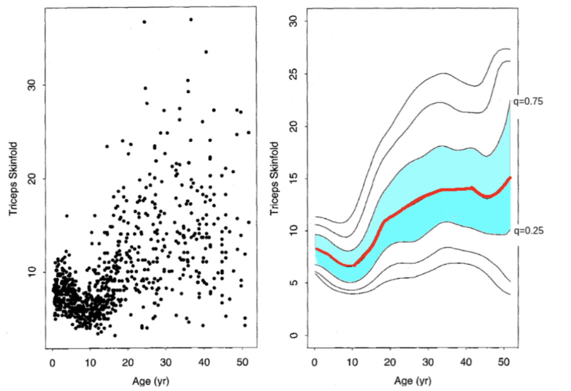

Figure 1.2: Example of univariate quantile regression X =age and Y =triceps skinfold. Figure extracted from Yu and Jones (1998). Red line represents the median and the blue region the center-outward quantile region of order 0.5.

La función que marca la tendencia en la regresión cuantílica es la mediana condicional, que corresponde con el valor $Q(\frac{1}{2}|\,\mathbf{x})$. Sin embargo, la propiedad fundamental de la regresión cuantílica, que la diferencia del resto de modelos de regresión, es el *«control probabilístico»* de los tubos y de las regiones cuantile, i.e.

$$\mathbb{P}\left[Y \in \left[Q(\frac{1}{2} - \frac{\tau}{2}|\,\mathbf{x}), Q(\frac{1}{2} + \frac{\tau}{2}|\,\mathbf{x})\right] |\, \mathbf{X} = \mathbf{x}\right] = \tau,$$

independiente de la probabilidad $\mathbb{P}$. Gracias a esta propiedad sabemos, por ejemplo, que la región azul la figura 1.2 tiene una probabilidad de 0.5. Este es el principio básico de la regresión cuantílica. Todo modelo de regresión que no lo satisfaga no puede considerarse como cuantílico, y si su intención fuera esta, desde luego que ha errado su objetivo.

Desde luego, no podemos conformarnos con una definición de regresión cuantílica válida únicamente para salidas univariantes; consideremos por ejemplo el siguiente modelo:

$$\mathbf{Y} = \begin{pmatrix} Y \\ Z \end{pmatrix} = \begin{pmatrix} V \\ \frac{U\sqrt{|X-\frac{1}{2}|} + V\,(1-\sqrt{|X-\frac{1}{2}|})}{\sqrt{|X-\frac{1}{2}| + (1-\sqrt{|X-\frac{1}{2}|})^2}} \end{pmatrix}, \quad X \sim \mathcal{U}_{(0,1)},\ U, V \sim \mathcal{N}(0,1), \quad (1.26)$$

siendo X, U e V mutuamente independientes. La regresión univariante en las marginales no nos da ninguna información útil, ambas son independientes de X. Toda la información está en la distribución conjunta. El lector se preguntará si proyectando a través de todas las direcciones podríamos sacar más información. Efectivamente más proyecciones darán más información, ya que estas caracterizan la probabilidad del vector $\mathbf{Y}$. Esto va en la línea de las profundidades integradas (véase Cuevas and Fraiman (2009)) que, como toda regresión basada en profundidades (véase (Serfling and Zuo, 2000) o (Serfling, 2002, 2019) para exposiciones de carácter general), pierde el control probabilístico de las regiones.

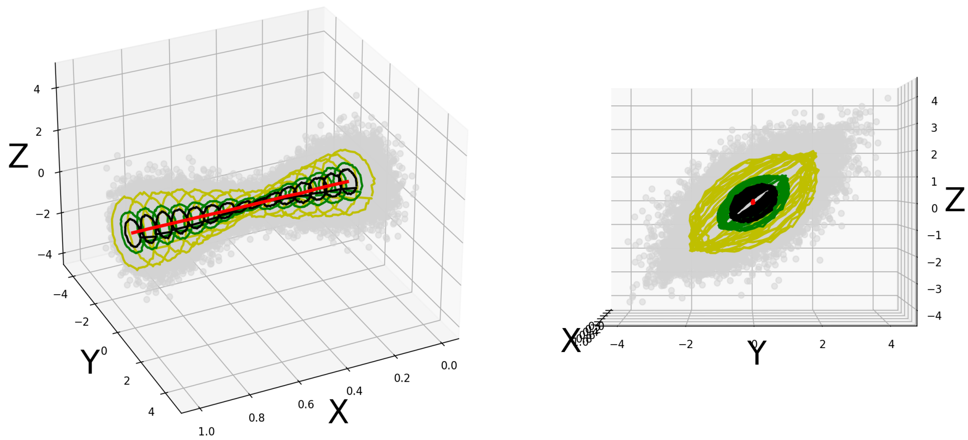

Figure 1.3: Ejemplo de egresión cuantílica univariante; con X e $\mathbf{Y} = (Y, Z)$ como en (1.26). El número de puntos observados es $n = 7{,}000$. Los contornos cuantiles para $\tau = 0.2, 0.4, 0.8$ están representados en negro ($\tau = 0.2$), verde ($\tau = 0.4$) y amarillo ($\tau = 0.8$). El centro viene representado en rojo.

La metodología propuesta en el capítulo 7 define regiones y tubos cuantiles, es capaz de analizar modelos no paramétricos mientras mantiene un control probabilístico asintótico de las regiones (el control de probabilidad empírico converge hacia el poblacional, que es conocido). La figura 1.3 analiza el modelo (1.26), dando al mismo tiempo una solución muy visual del mismo, la tendencia del modelo es claramente constante sin embargo la covariable X alcanza su pico de influencia sobre el vector de interés en la dirección $(1, 1)$. *La aplicación cuantil centro a fuera* $\mathbf{Q}_\pm$ *de* $\mathbf{Y}$ *condicionada a* $\mathbf{X} = \mathbf{x}$ está definida como

$$\mathbf{u} \in \mathbb{S}_d \mapsto \mathbf{Q}_\pm(\mathbf{u}|\, \mathbf{x}) \in \mathbb{R}^d.$$

Esta define, de manera natural las regiones condicionadas cuantilílicas como

$$\mathbb{C}_\pm(\tau|\mathbf{x}) := \mathbf{Q}_\pm(\tau\, \overline{\mathbb{S}}_d|\mathbf{x}) \quad \tau \in (0,1), \;\; \mathbf{x} \in \mathbb{R}^m,$$

que satisfacen la propiedad fundamental

$$\mathbb{P}\left[\mathbf{Y} \in \mathbb{C}_\pm(\tau|\, \mathbf{x}) \,\middle|\, \mathbf{X} = \mathbf{x}\right] = \tau \quad \text{ para todo } \mathbf{x} \in \mathbb{R}^m, \tau \in (0,1), \text{ y } \mathbb{P}.$$

Para $\tau = 0$, la mediana de la regresión es

$$\mathbb{C}_\pm(0\,|\,\mathbf{x}) := \bigcap_{\tau \in (0,1)} \mathbb{C}_\pm(\tau|\, \mathbf{x}).$$

La misma aplicación cuantil condicional caracteriza a los encajados (sin fenómeno de "cruce de cuantiles") *"tubos cuantílicos de regresión de orden τ"* (en $\mathbb{R}^{m+d}$)

$$\mathbb{T}_{\pm}(\tau) := \left\{ \left(\mathbf{x}, \mathbf{Q}_{\pm}(\tau \, \overline{\mathbb{S}}_d \big| \, \mathbf{x})\right) \big| \mathbf{x} \in \mathbb{R}^m \right\}, \quad \tau \in (0,1)$$

que satisfacen la propiedad fundamental

$$\mathbb{P}\left[(\mathbf{X}, \mathbf{Y}) \in \mathbb{T}_{\pm}(\tau)\right] = \tau \quad \text{independientemente de } \mathbb{P}, \, \tau \in (0,1).$$

Para $\tau = 0$, se define

$$\mathbb{T}_{\pm}(0) := \left\{ (\mathbf{x}, \mathbf{y}) \big| \mathbf{x} \in \mathbb{R}^m, \, \mathbf{y} \in \mathbb{C}_{\pm}(0 \, | \, \mathbf{x}) \right\} = \bigcap_{\tau \in (0,1)} \mathbb{T}_{\pm}(\tau)$$

(como el *grafo* de $\mathbf{x} \mapsto \mathbb{C}_{\pm}(\tau | \, \mathbf{x})$); con un ligero abuso del lenguaje, también la llamamos *mediana de la regresión* de $\mathbf{Y}$ con respecto a $\mathbf{X}$.

Cuando lo observado es una muestra $(\mathbf{X}, \mathbf{Y})^{(n)} := \left((\mathbf{X}_1, \mathbf{Y}_1), \ldots, (\mathbf{X}_n, \mathbf{Y}_n)\right)$ de n i.i.d. copias de $(\mathbf{X}, \mathbf{Y}) \sim= \mathrm{P}_{\mathbf{XY}}$, el capítulo 7 desarrolla un estimador de $\mathbf{u} \mapsto \mathbf{Q}_{\pm}(\mathbf{u} | \mathbf{X} = \mathbf{x})$, $\mathbf{x} \in \mathbb{R}^m$. Este estimador se obtiene en dos pasos: en el paso 1, construimos una distribución empírica de $\mathbf{Y}$ condicionada a $\mathbf{X} = \mathbf{x}$, en el paso 2, calculamos la correspondiente aplicación cuantil del centro hacia fuera.

La distribución empírica de $\mathbf{Y}$ condicionada a $\mathbf{X} = \mathbf{x}$ se aproxima mediante una función de pesos, por lo tanto (si no queremos limitarnos a funciones de peso constantes a trozos) esta será una probabilidad atómica con diferentes pesos. La solución que obtenemos requiere de la solución del problema de transporte óptimo siguiente

$$\begin{aligned} &\min_{\pi := \{\pi_{i,j}\}} \sum_{i=1}^{N} \sum_{j=1}^{n} \frac{1}{2} \left| \mathbf{Y}_j - \mathfrak{G}_i \right|^2 \pi_{i,j} \, , \\ \text{s.t.} \; &\sum_{j=1}^{n} \pi_{i,j} = \frac{1}{N}, \quad i \in \{1, 2, \ldots, N\}, \\ &\sum_{i=1}^{N} \pi_{i,j} = w_j^{(n)}(\mathbf{x}; \mathbf{X}^{(n)}), \quad j \in \{1, 2, \ldots, n\}, \\ &\pi_{i,j} \geq 0, \quad i \in \{1, 2, \ldots, N\}, \, j \in \{1, 2, \ldots, n\}, \end{aligned} \tag{1.27}$$

donde $\mathfrak{G}_1^{(n)}, \ldots, \mathfrak{G}_N^{(n)}$ es una interpolación regular de la uniforme esférica of $\mathbb{U}_d$. En este caso N es arbitrario, solo le pediremos que tienda a infinito. El problema es que la solución de (1.27) es un plan de transporte y no una aplicación como tal. Para crear una aplicación aplicamos el criterio

$$\mathbf{T}^*(\mathfrak{G}_i | \, \mathbf{x}) := \arg\inf \left\{ \|\mathbf{y}\| : \; \mathbf{y} \in \operatorname{conv}\left(\{ \mathbf{Y}_J : \; J \in \arg\max_j \pi^*_{i,j}(\mathbf{x}) \} \right) \right\}, \tag{1.28}$$

donde $\mathrm{conv}(A)$ denota la envolvente convexa de A. Procedemos como en Hallin et al. (2021a) eligiendo la interpolación suave cíclicamente monótona con la mayor constante de Lipschitz. Esta aplicación continua $\mathbf{u} \mapsto \mathbf{Q}^{(n)}_{w,\pm}(\mathbf{u}|\mathbf{x})$ de $\mathbb{S}_d$ a $\mathbb{R}^d$ se llamará la *función cuantitativa condicional empírica del centro hacia afuera.*

Demostraremos que si $w^{(n)}$ es consistente (en el sentido de Stone (1977)) entonces, para todo $\mathbf{u} \in \mathbb{S}_d$ y $\epsilon > 0$,

$$\mathbb{P}\left(\mathbf{Q}^{(n)}_{w,\pm}(\mathbf{u}\,|\mathbf{X}) \notin \mathbf{Q}_{\pm}(\mathbf{u}\,|\mathbf{X}) + \epsilon\mathbb{S}_d \right) \longrightarrow 0 \quad \text{cuando } n \text{ y } N \to \infty,$$

y para todo $\tau \in (0,1)$,

$$\mathbb{P}\left(\mathbb{C}^{(n)}_{\pm}(\tau\,|\mathbf{X}) \not\subset \mathbb{C}_{\pm}(\tau\,|\mathbf{X}) + \epsilon\mathbb{S}_d \right) \to 0 \text{ y } \mathbb{P}\left(\mathcal{C}^{(n)}_{\pm}(\tau\,|\mathbf{X}) \not\subset \mathcal{C}^{(n)}_{\pm}(\tau\,|\mathbf{X}) + \epsilon\mathbb{S}_d \right) \to 0$$

cuando n y $N \to \infty$.

Remark 1.4.2. *La demostración de este resultado sigue la línea de convergencia de aplicaciones multi-valuadas empíricas, iniciada por del Barrio and Loubes (2019) para probar la convergencia de los potenciales del transporte óptimo en el caso cuadrático, continuada por Hallin et al. (2021a) para la demostración del teorema Glivenko-Cantelli y formalizada finalmente en Segers (2022). Véase también del Barrio et al. (2021) para el caso de coste general. En este caso tenemos una dificultad añadida; la convergencia es en probabilidad, y, por lo tanto, tenemos que pasar numerosas veces a través de sub-sucesiones. Es un procedimiento tedioso.*

El capítulo 6 tiene su aplicación en el 7; es capaz de dar las condiciones que $(\mathbf{X}, \mathbf{Y})$ han de cumplir para poder obtener nociones más fuertes de convergencia. En este caso si **A** vale a.s. para la probabilidad condicionada, entonces, para todo compacto $K \subset \mathbb{S}_d \setminus \{\mathbf{0}\}$, cuando n y $N \to \infty$,

$$\sup_{\mathbf{u}\in K} |\mathbf{Q}^{(n)}_{w,\pm}(\mathbf{u}\,|\mathbf{X}) - \mathbf{Q}_{\pm}(\mathbf{u}\,|\mathbf{X})| \xrightarrow{\mathbb{P}} 0$$

y, para todo $\tau \in (0,1)$ y $\epsilon > 0$,

$$\mathbb{P}\left(d_H\left(\mathcal{C}^{(n)}_{\pm}(\tau\,|\mathbf{X}), \mathcal{C}_{\pm}(\tau\,|\mathbf{X})\right) > \epsilon\right) \to 0.$$

Es más, bajo estas condiciones, obtenemos el control de probabilidad asintótico

$$\mathbb{P}\left(\mathbf{Y} \in \mathbb{C}^{(n)}_{\pm}(\tau\,|\mathbf{X})|\mathbf{X}\right) \xrightarrow{\mathbb{P}} \tau \quad \text{para todo } \tau \in (0,1),$$

que justifica la metodología propuesta.

El capítulo 7 acaba con una serie de experimentos donde se muestra el comportamiento del estimador en modelos que incluyen heteroscedasticidad y tendencias no lineales; su poder como herramienta de análisis de datos se ilustra también en algunos conjuntos de datos reales.

Bibliography

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 214–223. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/arjovsky17a.html.

Emily Black, Samuel Yeom, and Matt Fredrikson. Fliptest: Fairness testing via optimal transport. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, page 111–121, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450369367. doi: 10.1145/3351095.3372845. URL https://doi.org/10.1145/3351095.3372845.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 02 2013. ISBN 9780199535255. doi: 10.1093/acprof:oso/9780199535255.001.0001. URL https://doi.org/10.1093/acprof:oso/9780199535255.001.0001.

Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991. doi: https://doi.org/10.1002/cpa.3160440402. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160440402.

Luis Caffarelli. Some regularity properties of solutions of monge ampère equation. *Communications on Pure and Applied Mathematics*, 44:965–969, 1991.

Luis A. Caffarelli. A localization property of viscosity solutions to the monge-ampere equation and their strict convexity. *Annals of Mathematics*, 131(1):129–134, 1990. ISSN 0003486X. URL http://www.jstor.org/stable/1971509.

Luis A. Caffarelli. The regularity of mappings with a convex potential. *Journal of the American Mathematical Society*, 5(1):99–104, 1992. ISSN 08940347, 10886834. URL http://www.jstor.org/stable/2152752.

Javier Cárcamo, Antonio Cuevas, and Luis-Alberto Rodríguez. Directional differentiability for supremum-type functionals: Statistical applications. *Bernoulli*, 26(3):2143 – 2175, 2020. doi: 10.3150/19-BEJ1188. URL https://doi.org/10.3150/19-BEJ1188.

Victor Chernozhukov, Alfred Galichon, Marc Hallin, and Marc Henry. Monge-Kantorovich depth, quantiles, ranks and signs. *Ann. Statist.*, 45(1):223–256, 2017. ISSN 0090-5364. doi: 10.1214/16-AOS1450. URL https://doi.org/10.1214/16-AOS1450.

Dario Cordero-Erausquin and Alessio Figalli. Regularity of monotone transport maps between unbounded domains. *Discrete and Continuous Dynamical Systems*, 39(12): 7101–7112, 2019.

Juan A. Cuesta and Carlos Matrán. Notes on the wasserstein metric in hilbert spaces. *Annals of Probability*, 17:1264–1276, 1989.

Antonio Cuevas and Ricardo Fraiman. On depth measures and dual statistics. a methodology for dealing with general data. *Journal of Multivariate Analysis*, 100(4):753–766, 2009. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2008.08.002. URL https://www.sciencedirect.com/science/article/pii/S0047259X08001796.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pages 2292–2300, 2013. URL http://papers.nips.cc/paper/4927-sinkhorn-distances-lightspeed-computation-of-optimal-transport.

Lucas de Lara, Alberto González-Sanz, Nicholas Asher, and Jean-Michel Loubes. Transport-based counterfactual models. *Pre-Print*, 2021.

Lucas De Lara, Alberto González-Sanz, and Jean-Michel Loubes. Diffeomorphic registration using sinkhorn divergences. *To appear in SIAM Journal on Imaging Sciences*, 2022+.

Nabarun Deb and Bodhisattva Sen. Multivariate rank-based distribution-free nonparametric testing using measure transportation. *J. Amer. Statist. Assoc.*, 2019. To appear.

Nabarun Deb, Bhaswar B. Bhattacharya, and Bodhisattva Sen. Efficiency lower bounds for distribution-free hotelling-type two-sample tests based on optimal transport, 2021. URL https://arxiv.org/abs/2104.01986.

Eustasio del Barrio and Jean-Michel Loubes. Central limit theorems for empirical transportation cost in general dimension. *The Annals of Probability*, 47(2):926 – 951, 2019. doi: 10.1214/18-AOP1275. URL https://doi.org/10.1214/18-AOP1275.

Eustasio del Barrio, Evarist Giné, and Frederic Utzet. Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli*, 11(1):131–189, 2005. ISSN 1350-7265. doi: 10.3150/bj/1110228245. URL https://doi.org/10.3150/bj/1110228245.

Eustasio del Barrio, Alberto González-Sanz, and Jean-Michel Loubes. Central limit theorems for general transportation costs. *Pre-Print*, 2021.

R. M. Dudley. The speed of mean glivenko-cantelli convergence. *Ann. Math. Statist.*, 40 (1):40–50, 02 1969. doi: 10.1214/aoms/1177697802. URL https://doi.org/10.1214/aoms/1177697802.

Jean Feydy, Benjamin Charlier, François-Xavier Vialard, and Gabriel Peyré. Optimal transport for diffeomorphic registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 291–299. Springer, 2017.

Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trouvé, and Gabriel Peyré. Interpolating between optimal transport and mmd using sinkhorn divergences. In *AISTATS*, 2019.

Alessio Figalli. *The Monge-Ampère Equation and Its Applications*. Zurich Lectures in Advanced Mathematics, European Mathematical Society (EMS), Zurich, 2017. ISBN 978-3-03719-670-0.

Alessio Figalli. On the continuity of center-outward distribution and quantile functions. *arXiv: Analysis of PDEs*, 2018.

Alessio Figalli and Young-Heon Kim. Partial regularity of brenier solutions of the monge-ampère equation. *Discrete and Continuous Dynamical Systems*, 28:559–565, 2010.

N. Fournier and A. Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162:707–738, 2013.

Alfred Galichon. *Optimal Transport Methods in Economics*. Princeton University Press, 09 2016. ISBN 9780691172767. doi: 10.23943/princeton/9780691172767.001.0001. URL https://doi.org/10.23943/princeton/9780691172767.001.0001.

Wilfrid Gangbo and Robert J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996. doi: 10.1007/BF02392620. URL https://doi.org/10.1007/BF02392620.

Aude Genevay, Lénaïc Chizat, Francis Bach, Marco Cuturi, and Gabriel Peyré. Sample complexity of sinkhorn divergences. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pages 1574–1583. PMLR, 16–18 Apr 2019. URL http://proceedings.mlr.press/v89/genevay19a.html.

Javier González-Delgado, Alberto González-Sanz, Juan Cortés, and Pierre Neuvial. Two-sample goodness-of-fit tests on the flat torus based on wasserstein distance and their relevance to structural biology. *arXiv: Probability*, 2021.

Paula Gordaliza, Eustasio Del Barrio, Gamboa Fabrice, and Jean-Michel Loubes. Obtaining fairness using optimal transport theory. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 2357–2365. PMLR, 09–15 Jun 2019a. URL http://proceedings.mlr.press/v97/gordaliza19a.html.

Paula Gordaliza, Eustasio Del Barrio, Gamboa Fabrice, and Jean-Michel Loubes. Obtaining fairness using optimal transport theory. In *International Conference on Machine Learning*, pages 2357–2365. PMLR, 2019b.

Marc Hallin and Gilles Mordant. On the finite-sample performance of measure transportation-based multivariate rank tests, 2021.

Marc Hallin, Eustasio del Barrio, Juan Cuesta-Albertos, and Carlos Matrán. Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *The Annals of Statistics*, 49(2):1139 – 1165, 2021a. doi: 10.1214/20-AOS1996.

Marc Hallin, Gilles Mordant, and Johan Segers. Multivariate goodness-of-fit tests based on Wasserstein distance. *Electron. J. Stat.*, 15(1):1328–1371, 2021b. doi: 10.1214/21-ejs1816. URL https://doi.org/10.1214/21-ejs1816.

Zaïd Harchaoui, Lang Liu, and Soumik Pal. Asymptotics of entropy-regularized optimal transport via chaos decomposition. *arXiv: Probability*, 2020.

Shayan Hundrieser, Marcel Klatt, Thomas Staudt, and Axel Munk. A unifying approach to distributional limits for empirical optimal transport. *Pre-Print*, 02 2022. URL https://arxiv.org/pdf/2202.12790.pdf.

Lev Kantorovich. On the translocation of masses. *Journal of Mathematical Sciences*, 133: 1381–1382, 2006.

Marcel Klatt, Carla Tameling, and Axel Munk. Empirical regularized optimal transport: statistical theory and applications. *SIAM J. Math. Data Sci.*, 2(2):419–443, 2020. doi: 10.1137/19M1278788. URL https://doi.org/10.1137/19M1278788.

Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1913643.

Robert J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80(2):309 – 323, 1995. doi: 10.1215/S0012-7094-95-08013-2. URL https://doi.org/10.1215/S0012-7094-95-08013-2.

Gonzalo Mena and Jonathan Niles-Weed. Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/5acdc9ca5d99ae66afdfe1eea0e3b26b-Paper.pdf.

Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019. ISSN 1935-8237. doi: 10.1561/2200000073. URL http://dx.doi.org/10.1561/2200000073.

Guido Philippis and Alessio Figalli. Partial regularity for optimal transport maps. *Publications mathématiques de l'IHÉS*, 121, 09 2012. doi: 10.1007/s10240-014-0064-7.

Laurent Risser, Alberto González Sanz, Quentin Vincenot, and Jean-Michel Loubes. Tackling algorithmic bias in neural-network classifiers using wasserstein-2 regularization. *J. Math Imaging Vis*, 2021.

R. Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series. Princeton University Press, Princeton, N. J., 1970.

Johan Segers. Graphical and uniform consistency of estimated optimal transport plans. *arXiv*, 2022.

Robert Serfling. Quantile functions for multivariate analysis: approaches and applications. *Statistica Neerlandica*, 56(2):214–232, 2002. doi: https://doi.org/10.1111/1467-9574.00195. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9574.00195.

Robert Serfling. Depth functions on general data spaces, i. perspectives, with consideration of "density" and "local" depths". Available at https://www.utdallas.edu~serfling, 2019.

Robert Serfling and Yijun Zuo. General notions of statistical depth function. *The Annals of Statistics*, 28(2):461–482, 2000. doi: 10.1214/aos/1016218226. URL https://doi.org/10.1214/aos/1016218226.

Hongjian Shi, Mathias Drton, and Fang Han. Distribution-free consistent independence tests via center-outward ranks and signs. *J. Amer. Statist. Assoc.*, 117(537):395–410, 2022. ISSN 0162-1459. doi: 10.1080/01621459.2020.1782223. URL https://doi.org/10.1080/01621459.2020.1782223.

Max Sommerfeld and Axel Munk. Inference for empirical wasserstein distances on finite spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):219–238, 2018.

Charles J. Stone. Consistent nonparametric regression. *Ann. Statist.*, 5(4):595–620, 07 1977. doi: 10.1214/aos/1176343886. URL https://doi.org/10.1214/aos/1176343886.

Cédric Villani. *Topics in optimal transportation*. Number 58 in Graduate studies in mathematics. American Mathematical Soc., 2003.

Cédric Villani. *Optimal Transport: Old and New*. Number 338 in Grundlehren der mathematischen Wissenschaften. Springer, Berlin, 2008. ISBN 978-3-540-71049-3. OCLC: ocn244421231.

Richard A Vitale. Lp metrics for compact, convex sets. *Journal of Approximation Theory*, 45(3):280–287, 1985. ISSN 0021-9045. doi: https://doi.org/10.1016/0021-9045(85)90051-6. URL https://www.sciencedirect.com/science/article/pii/0021904585900516.

Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A):2620 – 2648, 2019. doi: 10.3150/18-BEJ1065. URL https://doi.org/10.3150/18-BEJ1065.

Keming Yu and M. C. Jones. Local linear quantile regression. *Journal of the American Statistical Association*, 93(441):228–237, 1998. doi: 10.1080/01621459.1998.10474104. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1998.10474104.

Part I

Weak limits of optimal transport

CHAPTER 2

Central Limit theorem for general transport costs

The content of this chapter has been officially accepted for publication in Annales de l'Institut Henri Poincaré. It is fully available online in del Barrio et al. (2021b).

Contents

We consider the problem of optimal transportation with general cost between an empirical measure and a general target probability on $\mathbb{R}^d$, with $d \geq 1$. We provide results on asymptotic stability of optimal transport potentials under minimal regularity assumptions on the costs or the underlying probability. This stability is combined with a refined linearization technique based on the sequential compactness of the closed unit ball in $L^2(P)$ for the weak topology and the strong convergence of Cesàro means along subsequences. As a result we obtain a CLT for the transportation cost under sharp smoothness and moment assumptions, giving a positive answer to a conjecture in del Barrio and Loubes (2017) for the quadratic costs.

2.1 Introduction

In the last few years new techniques based on the optimal transportation problem have become popular to handle statistical and machine learning problems over the space of probability distributions. Dealing with distributions has shed light on the need for probabilistic tools that are well adapted to the intrinsic geometry of the data, and the theory of optimal transport provides a natural framework to tackle such issues. In particular the transportation cost distance is a convenient metric in many problems encountered in data science and the range of application fields is huge, including for instance computational statistics, biology, image analysis, economy, finance or fairness in machine learning. We refer for instance to Bachoc et al. (2018); Black et al. (2020); Courty et al. (2018); Peyré and Cuturi (2019); Peyré and Cuturi (2019); Gordaliza et al. (2019); Risser et al. (2021); Schiebinger et al. (2019) and references therein. Understanding the approximations done when dealing with empirical distributions and providing better controls on the asymptotic distribution of optimal transport cost is of importance for further research on this subject.

In all this work, we will be concerned with probabilities on the measurable space $\mathbb{R}^d$, endowed with the Borel σ-field, denoted as $\mathcal{P}(\mathbb{R}^d)$. In this setting, the optimal transport problem is formulated as follows. Let P, Q be probability measures in $\mathcal{P}(\mathbb{R}^d)$ and $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ be a function referred to as the *cost*. We say that a measurable map $T : \mathbb{R}^d \to \mathbb{R}^d$ is an *optimal transport map* from P to Q if it is a minimizer in the problem

$$\mathcal{T}_c(P,Q) := \inf_{T:\, T_{\#}P=Q} \int_{\mathbb{R}^d} c(\mathbf{x}, T(\mathbf{x}))dP(\mathbf{x}), \tag{2.1}$$

where the notation $T_{\#}P$ represents the *push-forward* measure, that is, the measure such that for each measurable set A we have $T_{\#}P(A) := P(T^{-1}(A))$.
This previous formulation of the problem is known as the Monge formulation and is closely related to the following problem known as the Kantorovich optimal transportation problem. A probability measure $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is said to be an *optimal transport plan for the cost c* between P and Q if it is a minimizer in the problem

$$\mathcal{T}_c(P,Q) = \inf_{\gamma\in\Pi(P,Q)} \int_{\mathbb{R}^d\times\mathbb{R}^d} c(\mathbf{x},\mathbf{y})d\pi(\mathbf{x},\mathbf{y}), \tag{2.2}$$

where $\Pi(P,Q)$ is the set of probability measures $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ such that $\pi(A \times \mathbb{R}^d) = P(A)$ and $\pi(\mathbb{R}^d \times B) = Q(B)$ for all A, B measurable sets. We have used the same notation for the minimum value in both (2.1) and (2.2), and this and the existence of optimal transport maps indeed hold for rather general costs, as shown in Gangbo and McCann (1996), including the *potential* costs $c_p(\mathbf{x},\mathbf{y}) = |\mathbf{x}-\mathbf{y}|^p$, $p \geq 1$, where $|\cdot|$ denotes the Euclidean distance in $\mathbb{R}^d$. We write in this case $\mathcal{T}_p(P,Q)$ for the minimal value in (2.1) or (2.2) and $\mathcal{W}_p(P,Q) := (\mathcal{T}_p(P,Q))^{1/p}$. Note that $\mathcal{W}_p(P,Q)$ is a distance on the subset in $\mathcal{P}(\mathbb{R}^d)$ of distributions with finite moment of order p, denoted as $\mathcal{P}^p(\mathbb{R}^d)$, referred to as the $p-$Wasserstein or Monge-Kantorovich distance. This distance is closely related to the weak topology of $\mathcal{P}(\mathbb{R}^d)$, in the sense that $P_n \xrightarrow{w} P$ and $\int |\mathbf{x}|^p dP_n(\mathbf{x}) \longrightarrow \int |\mathbf{x}|^p dP(\mathbf{x})$ is equivalent to $\mathcal{W}_p(P_n, P) \longrightarrow 0$.
In this work we prove, under minimal assumptions, a Central Limit Theorem for the empirical transport cost, in general dimension and for a large class of costs, including potential costs c_p with $p > 1$,

$$\sqrt{n}\left(\mathcal{T}_c(P_n,Q) - E\mathcal{T}_c(P_n,Q)\right) \xrightarrow{w} N(0, \sigma_c^2(P,Q)),$$

with

$$\sigma_c^2(P,Q) := \int \varphi(\mathbf{x})^2 dP(\mathbf{x}) - \left(\int \varphi(\mathbf{x})dP(\mathbf{x})\right)^2,$$

where φ is an optimal transport potential for the cost c, from P to Q (see (2.3) below) and P_n denotes the empirical measure on a sample $X_1, \ldots, X_n$ of i.i.d. observations following the distribution P. Note that this result also holds for $\mathcal{T}_c(P_n, Q_n)$, where Q_n is the corresponding empirical measure of a sample with distribution Q.

A large number of papers tackle the problem of understanding the asymptotic behaviour of the empirical transport cost. Early work on this topic, starting with Ajtai et al. (1984) (see also Talagrand (1992, 1994); Talagrand and Yukich (1993) and the more recent Fournier and Guillin (2013)), focused on the case $P = Q$. Here $\mathcal{T}_p(P_n, P)$ converges to 0 a.s. if P has a finite moment of order p, so authors focus on rates of decay of $\mathcal{T}_p(P_n, P)$, which depend on the dimension of the sample space. In the one dimensional case, expression of the transport cost using quantiles function is exploited in del Barrio et al. (1999) and Barrio et al. (2005) for proving distributional limit theorems for $\mathcal{T}_p(P_n, P)$, $p = 1, 2$. We refer to del Barrio and Loubes (2017) and references therein for a more detailed history of the problem. The problem has received a renewed interest in the last few years, both in the setup $P = Q$ (see Ambrosio et al. (2019); Ledoux (2019); Talagrand (2018)) or for general P and Q (see Sommerfeld and Munk (2018) and Tameling et al. (2019) for finitely and countably supported probabilities, del Barrio and Loubes (2017) for the case $p = 2$ and general probabilities and dimension and del Barrio et al. (2019); Berthet et al. (2020) for dimension $d = 1$ and general costs). Yet, for dimension greater than 1, the problem becomes more difficult.

The central limit theorems for $\mathcal{T}_c(P_n, Q)$ or $\mathcal{T}_c(P_n, Q_m)$ provided in this paper are valid for general cost functions and general dimension, under minimal moment and regularity assumptions on P and Q. Our contribution covers the strictly convex costs in Gangbo and McCann (1996) for which existence of the optimal transport is guaranteed since strict convexity of the cost is a minimal requirement for a general central limit theorem with a Gaussian limiting distribution. Actually for the non strictly convex cost $p = 1$ and in a univariate setup, del Barrio et al. (1999) shows that $\{\sqrt{n}\mathcal{T}_1(P_n, P)\}_{n\in\mathbb{N}}$ converges to a non-Gaussian distribution under some regularity assumptions. Our moment assumptions improve upon those in del Barrio and Loubes (2017). In that work, the central limit theorem holds for a quadratic cost, mild regularity assumptions on P and Q (these are assumed to be absolutely continuous probabilities on $\mathbb{R}^d$ with convex supports) and finite moments of order $4 + \delta$ for some $\delta > 0$. If finite moment of order 4 is necessary (and sufficient in this case) for a CLT, we prove in this paper that, as in the case of dimension 1 (see in del Barrio et al. (2019)), for the potential cost c_p a necessary and sufficient condition is the existence of only finite moments of order $2p$.

The key to prove a CLT for the transportation cost in del Barrio and Loubes (2017) is a linearization technique based on the Efron-Stein inequality for variances coupled with stability results for optimal transportation *potentials*. Actually, for continuous costs the Kantorovich problem (2.2) admits an equivalent dual form, namely,

$$\mathcal{T}_c(P, Q) = \sup_{(f,g)\in\Phi_c(P,Q)} \int f(\mathbf{x})dP(\mathbf{x}) + \int g(\mathbf{y})dQ(\mathbf{y}), \tag{2.3}$$

where $\Phi_c(P,Q) = \{(f,g) \in L^1(P) \times L^1(Q) : \ f(\mathbf{x}) + g(\mathbf{x}) \leq c(\mathbf{x},\mathbf{y})\}$. It is said that $\psi \in L^1(P)$ is an optimal transport potential from P to Q for the cost c if there exists $\varphi \in L^1(Q)$ such that the pair (ψ, φ) solves (2.3).

The present contribution provides a completely new tool to prove a central limit theorem for transportation costs. This approach can be summarised as follows. As before, we approximate the empirical transportation cost by a linear term. The linearization error R_n (see (2.21) for details) has a variance that can be bounded using the Efron-Stein inequality. The upper bound is the expected value of a random variable denoted by U_n, which converges to 0 a.s. (using new stability results for optimal transport potentials), but one cannot conclude from this that $E(U_n) \to 0$ without further conditions (as in del Barrio and Loubes (2017), for instance). Yet, one can show that the sequence EU_n is bounded and then the Banach-Alaoglu theorem yields weak convergence in $L^2(\mathbb{P})$ of U_n along subsequences. By taking Cesàro means we can go from weak to strong convergence. The major improvement we provide, comes at this point, where a detailed analysis of the variances of the Cesàro means allows to conclude that $\sqrt{n}(R_n - ER_n) \to 0$ in probability, which immediately yields a CLT. We refer to Section 2.4, the proofs in the Appendix and Remark 2..2 for all the details. While the linearization based directly on the Efron-Stein inequality requires technical assumptions, the proof we provide here enables to obtain optimal sharp assumptions improving previous results that will benefit to all works using such method such as Mena and Niles-Weed (2019) where the asymptotic behaviour of entropically regularized Wasserstein distances is proved using such arguments.

A second relevant contribution in this paper are new results on the convergence and, in some sense, the uniqueness of both optimal transport potentials and maps between P_n and Q_n when these sequences converge weakly to some probabilities P and Q. There is a large amount of literature working on these topics. Convergence of optimal maps is a topic of general interest, beyond our application to CLTs, and results on this issue have a long history, tracing back at least to Cuesta-Albertos et al. (1997). To our knowledge, interest on the convergence of potentials is more recent and requires some additional guarantee on the uniqueness of the potentials. Seminal results on it can be found in Theorem 2.8 in del Barrio and Loubes (2017) for the quadratic cost. Corollary 5.23 in Villani (2008) deals with this problem for general costs but one of both probabilities is supposed to be fixed. Some results are provided in Theorem 1.52. in Santambrogio (2015) when the involved probabilities are compactly supported.

Then problem of uniqueness of optimal transport potentials is linked to the smoothness of the probabilities and also to the topology of their supports. For a probability Q is the smallest closed set R_Q such $Q(R_Q) = 1$. Yet, with a slight abuse of notation, we will write

$$\text{Supp}(Q) := \text{int}\,(R_Q) \tag{2.4}$$

for the interior of R_Q. Moreover, we say that a probability Q has *negligible boundary* if $\ell_d(R_Q \setminus \text{Supp}(Q)) = 0$, where ℓ_d denotes Lebesgue measure on $\mathbb{R}^d$. A probability with a convex support has a negligible boundary, but the condition is far from necessary. When the cost is of the form $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$ with h satisfying some regularity assumptions (see (A1)-(A3) and the related discussion in Section 2) and Q has a density with respect to Lebesgue measure (in the sequel, when $Q \ll \ell_d$) and a connected support with negligible boundary then we prove (Corollary 2.2.7) that optimal transport potentials are unique up to an additive constant (during the time our work was under review, Staudt et al. (2022) extends it unconnected domains, motivated, among other things, by results such as those presented in this paper). From this uniqueness we move on to give general stability results for optimal transport potentials under only the following assumption

Assumption 1. *$Q \in \mathcal{P}(\mathbb{R}^d)$ is such that $Q \ll \ell_d$ and has connected support with negligible boundary; $Q_n, P_n, P \in \mathcal{P}(\mathbb{R}^d)$ are such that $P_n \stackrel{w}{\to} P$, $Q_n \stackrel{w}{\to} Q$,*

$$\mathcal{T}_c(P_n, Q_n) < \infty \ \text{ and } \ \mathcal{T}_c(P, Q) < \infty,$$

for a cost $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$ with h differentiable and satisfying (A1)-(A3), defined below.

If ψ_n (resp. ψ) are the c-optimal transport potentials from Q_n to P_n (resp. from Q to P), then we prove in Theorem 2.3.4 that

(a) There exist constants $a_n \in \mathbb{R}$ such that $\tilde{\psi_n} := \psi_n - a_n \to \psi$ in the sense of uniform convergence on the compacts sets.

(b) For each compact $K \subset \text{Supp}(Q) \cap \text{dom}(\nabla\psi)$

$$\sup_{\mathbf{x}\in K} \sup_{\mathbf{y}_n \in \partial^c \psi_n(\mathbf{x})} |\mathbf{y}_n - \nabla^c \psi(\mathbf{x})| \longrightarrow 0,$$

where $\text{dom}(\nabla\psi)$ denotes the set of points where ψ is differentiable, $\partial^c \psi_n$ and $\nabla^c \psi(\mathbf{x})$ are defined in section 2.2.

The paper falls into the following sections. Section 2.2 present the main results of analysis that will be used in the paper. Stability results for the optimal transport potential under Assumption 1 are given in Section Section 2.3. In Section 2.4 we consider the asymptotic behaviour of $\{\sqrt{n}(\mathcal{T}_c(P_n, Q) - E\mathcal{T}_c(P_n, Q))\}_{n\in\mathbb{N}}$ and provide our main result in the Central Limit Theorem 2.4.5. This CLT holds assuming only that $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$ with h differentiable and satisfying (A1)-(A3) and $P, Q \in \mathcal{P}(\mathbb{R}^d)$ satisfying

Assumption 2. *$P \ll \ell_d$ and $Q \ll \ell_d$ have connected supports with negligible boundary; moreover*

$$\int h(2\mathbf{x})^2 dP(\mathbf{x}) < \infty, \quad \int h(-2\mathbf{y})^2 dQ(\mathbf{y}) < \infty,$$

and

$$\inf_{q_1, q_2 \in [1,\infty]:\ \frac{1}{q_1} + \frac{1}{q_2} = 1} E|X_1 - X_1'|^{2q_1} E\left(\int_{\mathbb{R}^d} |\nabla h(X_1 - \mathbf{y})|^{2q_2} dQ(\mathbf{y})\right) < \infty,$$

where X_1 and X_1' are i.i.d. according to P.

As noted above, the linearization technique that we use yields CLTs for the transportation cost under minimal assumptions. We discuss this with detail in the case of potential costs in Section 4. As a minor price to pay, the approach does not yield moment convergence. We show in Theorem 2.4.6 that moment convergence holds under some additional moment assumptions. Finally, we derive a CLT for the empirical transportation cost in a two-sample setup and a further CLT for the empirical p-Wasserstein distance. Proofs are postponed to the Appendix

We end this introduction with some details about our setup and notation. We assume all the involved random variables (we use this term for both $\mathbb{R}$ and $\mathbb{R}^d$-valued random elements) to be defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We write $L^2(\mathbb{P})$ for the Hilbert space of square integrable random variables on the former space. $\xrightarrow{w}$ denotes weak convergence of probability measures, while we write $\stackrel{L^2}{\rightharpoonup}$ for weak convergence (in the usual sense in Functional Analysis) in the space $L^2(\mathbb{P})$. At some points we write $A \subset\subset B$ to mean that there is some compact set, K, such that $A \subset K \subset B$.

2.2 Preliminary results on optimal transport maps and potentials

This section presents some results related to optimal transport potentials and maps for general costs. The main reference on the topic is Gangbo and McCann (1996). We give two main results, which are necessary tools for the study of stability in section 3: we prove uniqueness, up to an additive constant, of the optimal transport potential (Corollary 2.2.7) and a weak continuity result for a version of the optimal transport maps Lemma 2.2.10.

We consider the optimal transport problem formulated in its dual form (2.3). Convexity plays a key role in the optimal transportation problem with quadratic cost. This idea can be adapted to general costs through the notion of c-concavity. Recall that $f : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ is said to be *c-concave* if there exist a set $\mathcal{T} \subset \mathbb{R}^d \times \mathbb{R}$ such that

$$f(\mathbf{x}) = \inf_{(\mathbf{y},\mathbf{t}) \in \mathcal{T}} \{c(\mathbf{x}, \mathbf{y}) - \mathbf{t}\}. \tag{2.5}$$

For a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ the *c-conjugate of f* (see Gangbo and McCann (1996)) is defined as

$$f^c(\mathbf{y}) = \inf_{\mathbf{x} \in \mathbb{R}^d} \{c(\mathbf{x}, \mathbf{y}) - f(\mathbf{x})\} \text{ for all } \mathbf{y} \in \mathbb{R}^d. \tag{2.6}$$

c-conjugation can be seen as a generalization of the Legendre's transform in convex analysis, see Rockafellar (1970). Obviously, f^c is c-concave and it is easy to check that its own c-conjugate, f^{cc}, satisfies $f^{cc} \geq f$, with equality if f is c-concave. This means that we can restrict the collection of pairs (f, g) in (2.3) to pairs (f, f^c), with f c-concave, without changing the optimal value.

For a $c-$concave function $f : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ the *c-superdifferential* of f, $\partial^c f$, is the set of pairs $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^d$ such that

$$f(\mathbf{z}) \leq f(\mathbf{x}) + [c(\mathbf{z}, \mathbf{y}) - c(\mathbf{x}, \mathbf{y})] \quad \text{for all } \mathbf{z} \in \mathbb{R}^d$$

(see, e.g., Definition 1.1 in Gangbo and McCann (1996)). We write $\partial^c f(\mathbf{x})$ for the set of $\mathbf{y}$ such that $(\mathbf{x},\mathbf{y}) \in \partial^c f$ and, more generally, $\partial^c f(U) = \cup_{\mathbf{x}\in U} \partial^c f(\mathbf{x})$ for $U \subset \mathbb{R}^d$. Under mild assumptions (implied by (A1)-(A3) below; see Propositions C.3 and C.4 in Gangbo and McCann (1996)) $\partial^c f(\mathbf{x})$ is nonempty if f is finite in a neighborhood of $\mathbf{x}$. When $\partial^c f(\mathbf{x})$ is a singleton we denote this point as $\nabla^c f(\mathbf{x})$. It is easy to see, for a c-concave function f, that $f(\mathbf{x}) + f^c(\mathbf{y}) \le c(\mathbf{x},\mathbf{y})$, with equality if and only if $\mathbf{y} \in \partial^c f(\mathbf{x})$. As a consequence of these key observations, $\pi \in \Pi(P,Q)$ is an optimal transport plan (a minimizer in (2.2)) and the c-concave function f is an optimal transport potential ((f, f^c) is a maximizer in (2.3)) if and only if π is concentrated on the set $\partial^c f$. This yields a characterization of optimal transport plans, provided a maximizer in (2.3) exists. In that case we can get an equivalent description of optimal transport plans in terms of cyclical monotonocity (see Smith and Knott (1992); Rüschendorf (1995)). A set $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ is said to be *c-cyclically monotone* if for all $n \in \mathbb{N}$ and $\{(\mathbf{x}_k, \mathbf{y}_k)\}_{k=1}^n \subset \Gamma$

$$\sum_{k=1}^n c(\mathbf{x}_k, \mathbf{y}_k) \le \sum_{k=1}^n c(\mathbf{x}_{\sigma(k)}, \mathbf{y}_k), \tag{2.7}$$

for every permutation σ in $\{1,\dots,n\}$. Optimal transport plans are supported in c-cyclically monotone sets (see Theorem 2.2.4 below). In the convex case (which corresponds to the quadratic cost $c_2(\mathbf{x},\mathbf{y}) = |\mathbf{x}-\mathbf{y}|^2$) cyclically monotone sets are those contained in the subdifferential of a convex function and the subdifferential of a convex function is maximal cyclically monotone (this is known as Rockafellar's Theorem, see for instance, Rockafellar (1966)). For general costs a similar result holds. We quote it for convenience in the next Lemma. A proof can be found in Rüschendorf (1996) (Lemma 2.1). Note that Lemma 2.2.1 is weaker than Rockafellar's Theorem for convex functions, since it does not claim that the set $\partial^c f$ is maximal.

Lemma 2.2.1. *If $c \ge 0$ is a continuous cost then a set $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ is c-cyclically monotone if and only if there exists a c-concave function f such that $\Gamma \subset \partial^c f$.*

Existence of maximizing pairs in (2.3) (which, as noted above, would yield a characterisation of optimal transport plans) is not guaranteed without some assumptions on the cost. Hence, we restrict our study to regular costs in the sense of Gangbo and McCann (1996), as follows: we will assume $c(\mathbf{x},\mathbf{y}) = h(\mathbf{x}-\mathbf{y})$, where $h : \mathbb{R}^d \to [0,\infty)$ is a non negative function satisfying

(A1) h is strictly convex on $\mathbb{R}^d$,

(A2) given a height $r \in \mathbb{R}^+$ and an angle $\theta \in (0,\pi)$, there exists some $M := M(r,\theta) > 0$ such that for all $|\mathbf{p}| > M$, one can find a cone

$$K(r,\theta,\mathbf{z},\mathbf{p}) := \left\{ \mathbf{x} \in \mathbb{R}^d : |\mathbf{x}-\mathbf{p}||\mathbf{z}|\cos(\theta/2) \le \langle \mathbf{z}, \mathbf{x}-\mathbf{p}\rangle \le r|\mathbf{z}| \right\},$$

with vertex at $\mathbf{p}$ (and $\mathbf{z} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$) on which h attains its maximum at $\mathbf{p}$, (A3) $\lim_{|\mathbf{x}|\to\infty} \frac{h(\mathbf{x})}{|\mathbf{x}|} = \infty$.

Remark 2.2.2. *The potential cost* $c_p(\mathbf{x},\mathbf{y}) := |\mathbf{x}-\mathbf{y}|^p$ *satisfies conditions* (A1)-(A3) *for* $p > 1$*, see* Gangbo and McCann (1996).

In the case of a quadratic cost the crucial step to turn the characterization of optimal transport plans into a characterization of optimal transport maps relies on the fact that convex functions are locally Lipschitz, hence, by Rademacher's Theorem (see, e.g., Theorem 9.60 in R. Tyrrell Rockafellar (1998)), they are differentiable at almost every point in the interior of their domain. For general costs convexity does not hold, but the Lipschitz property remains with great generality. In fact, if g is a c-concave function then for every $(\mathbf{a},\mathbf{b}),(\mathbf{x},\mathbf{y}) \in \partial^c g$ we have

$$|g(\mathbf{x}) - g(\mathbf{a})| \quad \leq \quad |c(\mathbf{x},\mathbf{y}) - c(\mathbf{a},\mathbf{y})| + |c(\mathbf{x},\mathbf{b}) - c(\mathbf{a},\mathbf{b})|. \tag{2.8}$$

When $c(\mathbf{x},\mathbf{y}) = h(\mathbf{x}-\mathbf{y})$ with h convex and differentiable, (2.8) implies that

$$|c(\mathbf{x},\mathbf{y}) - c(\mathbf{a},\mathbf{y})| \quad \leq \quad |\mathbf{x}-\mathbf{a}| \left(|\nabla h(\mathbf{x}-\mathbf{y})| + |\nabla h(\mathbf{a}-\mathbf{y})|\right).$$

As a consequence we obtain that

$$|g(\mathbf{x}) - g(\mathbf{a})| \leq |\mathbf{x}-\mathbf{a}||\zeta(\mathbf{a},\mathbf{b},\mathbf{x},\mathbf{y})|, \text{ for all } (\mathbf{a},\mathbf{b}),(\mathbf{x},\mathbf{y}) \in \partial^c g, \tag{2.9}$$

where $\zeta(\mathbf{a},\mathbf{b},\mathbf{x},\mathbf{y})$ is a continuous function (we recall that a differentiable convex function is, in fact, continuously differentiable, see Corollary 25.5.1 in Rockafellar (1970)). Elaborating on these bounds it can be proved that under (A1)-(A3) c-concave functions are locally Lipschitz, hence, differentiable at almost every point. For convenience we quote here a precise result (see Theorem 3.3 in Gangbo and McCann (1996)).

Lemma 2.2.3. *Let* $c(\mathbf{x},\mathbf{y}) = h(\mathbf{x}-\mathbf{y})$ *be a cost satisfying* (A1)-(A3) *and let* f *be a* c*-concave function, then there exists a convex set* $K \subset \mathbb{R}^d$ *with interior* Ω *such that*

(i) $\Omega \subset \mathrm{dom}(f) = \{\mathbf{x} : f(\mathbf{x}) \in \mathbb{R}\} \subset K$,

(ii) f *is locally Lipschitz in* Ω.

Now we can relate the shape of the gradient of a c-concave function to the shape of the c-superdifferential. We write h^* for the convex conjugate of h, namely, $h^*(\mathbf{y}) = \sup_{\mathbf{x}}(\langle \mathbf{x},\mathbf{y}\rangle - h(\mathbf{x}))$. Then, if f is c-concave (see Proposition 3.4 in Gangbo and McCann (1996)):

a) the relation $\mathbf{s}(\mathbf{x}) = \mathbf{x} - \nabla h^*(\nabla f(\mathbf{x}))$ defines a Borel function in the set where f is differentiable, $\mathrm{dom}(\nabla f)$,

b) for all $\mathbf{x} \in \mathrm{dom}(\nabla f)$ it holds that $\partial^c f(\mathbf{x}) = \nabla^c f(\mathbf{x}) = \{\mathbf{s}(\mathbf{x})\}$,

c) the set $\mathrm{dom}(f) \setminus \mathrm{dom}(\nabla f)$ is of Lebesgue measure zero.

Now, with all the ingredients above, a characterization of optimal transport plans and maps is given the next result, which summarizes Theorems 1.2, 2.3 and 2.7 in Gangbo and McCann (1996).

Theorem 2.2.4. *For any cost* $c(\boldsymbol{x},\boldsymbol{y}) = h(\boldsymbol{x}-\boldsymbol{y})$*, satisfying* (A1)-(A3)*, and Borel probability measures* P, Q *on* $\mathbb{R}^d$ *such that* $\mathcal{T}_c(P,Q) < \infty$:

(i) There exists an optimal transport plan. $\gamma \in \Pi(P,Q)$ *is an optimal transport plan if and only if its support,* $\text{Supp}(\gamma)$*, is a c-cyclically monotone set, or, equivalently, if there exists a c-concave function* ψ *such that* $\text{Supp}(\gamma) \subset \partial^c\psi$*. In this case* ψ *is an optimal transport potential.*

(ii) If $P \ll \ell_d$*, then there exists a unique optimal transport plan* $\gamma := (id \times T)\#P$*, where* $T(\boldsymbol{x}) := \boldsymbol{x} - \nabla h^*(\nabla\psi(\boldsymbol{x})) = \nabla^c\psi(\boldsymbol{x})$ *is* P*-a.s. unique and the c-concave function* ψ *is an optimal transport potential.*

The approach in this work to CLT's for the empirical transportation cost relies on the stability results for optimal transport potentials that we prove in Section 2.3. There cannot be any result in that sense without some kind of uniqueness of this potential. Of course, a look at (2.3) shows that if ψ is an optimal transport potential and $C \in \mathbb{R}$ then $\psi + C$ is also an optimal transport potential. With the next results we show that, under some minimal assumptions, the optimal transport potential is unique up to the addition of a constant.

Lemma 2.2.5. *Let* $\Omega \subset \mathbb{R}^d$ *be an open, bounded convex set,* $f : \Omega \longrightarrow \mathbb{R}$ *be a Lipschitz function such that* $\nabla f = \mathbf{0}$ *almost everywhere in* Ω*, then there exists a constant* $C \in \mathbb{R}$ *such that* $f = C$ *in* L*.*

Proof. This is a straightforward consequence of Poincaré's inequality in convex domains (see, e.g., Theorem 3.2. in Acosta and Duran (2004)). □

Theorem 2.2.6. *Assume* $c(\mathbf{x},\mathbf{y}) = h(\mathbf{x}-\mathbf{y})$ *satisfies* (A1)-(A3) *and* f_1, f_2 *are c-concave functions such that* $\nabla f_1 = \nabla f_2$ *almost everywhere in an open connected set* Ω*, then there exists a constant* $C \in \mathbb{R}$ *such that* $f_2 = f_1 + C$ *in* Ω*.*

Proof. Assume $\mathbf{p} \in \Omega \subset \text{dom}(f_1) \cap \text{dom}(f_2)$. By Lemma 2.2.3 ϕ, ψ are locally Lipschitz, hence, there exist $\epsilon_{\mathbf{p}} > 0$ such that f_1, f_2 are Lipschitz in $B(\mathbf{p}, \epsilon_{\mathbf{p}})$. Then the function $f_2 - f_1$ satisfies the assumptions of Lemma 2.2.5. As a consequence, there exists $C_{\mathbf{p}} \in \mathbb{R}$ such that $f_2 = f_1 + C_{\mathbf{p}}$ in $\mathbb{B}(\mathbf{p}, \epsilon_{\mathbf{p}})$ for each $\mathbf{p} \in L$. The proof will be complete if we show that the previous constant does not depend on $\mathbf{p}$. But this follows from connectedness of the Ω, since if we set

$$\Gamma := \{\mathbf{q} \in \Omega : C_{\mathbf{q}} = C_{\mathbf{p}}\}$$

then Γ is obviously open and, by continuity, Γ is also closed in the relative topology on Ω. Hence, being both open and closed implies $\Omega = \Gamma$ due to the assumption of connectedness. □

Let us assume now that P and Q are probabilities on $\mathbb{R}^d$ with P absolutely continuous and ψ_1, ψ_2 are optimal transport potentials. By Theorem 2.2.4 we have $\nabla h^*(\nabla\psi_1(\mathbf{x})) = \nabla h^*(\nabla\psi_2(\mathbf{x}))$ P-a.s. If h is differentiable then $\nabla h^*(\nabla\psi_1(\mathbf{x})) = \nabla h^*(\nabla\psi_2(\mathbf{x})) = \mathbf{y}$ implies $\nabla\psi_1(\mathbf{x}) = \nabla\psi_2(\mathbf{x}) = \nabla h(\mathbf{y})$ (cf. Corollary 23.5.3 and Theorem 25.1 in Rockafellar (1970)). Hence, P-a.s., $\nabla\psi_1(\mathbf{x}) = \nabla\psi_2(\mathbf{x})$. If, additionally, P is supported in an open, connected set we can apply Theorem 2.2.6 and conclude that $\psi_2 = \psi_1 + C$ on the support of P for some constant $C \in \mathbb{R}$. This proves the following uniqueness result for optimal transport potentials.

Corollary 2.2.7. *If $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$, where h is differentiable and satisfies* (A1)-(A3)*, $P \ll \ell_d$ and is supported on an open, connected set, A, and ψ_1, ψ_2 are optimal transport potentials from P to Q for the cost c, then, there exists a constant, $C \in \mathbb{R}$, such that $\psi_2(\mathbf{x}) = \psi_1(\mathbf{x}) + C$ for every $\mathbf{x} \in A$.*

In the next section we will state and prove results related to the stability of optimal transport maps and potentials, namely, we will prove convergence in different senses of optimal transport potentials (φ_n) or maps $(\nabla^c\varphi_n)$ from P_n to Q_n under the assumption that (at least) $P_n \xrightarrow{w} P$ and $Q_n \xrightarrow{w} Q$. Results of this kind have a long history, tracing back at least to Cuesta-Albertos et al. (1997) for the case of optimal transport maps under quadratic costs. Stability of the potentials is crucial for the Efron-Stein approach to CLTs in del Barrio and Loubes (2017) or in section 4 in this paper, and has only been investigated recently. For smooth probabilities, optimal transport potentials are a.s. differentiable, and there is a simple relation between their gradients and the optimal transport maps, as noted above. Hence, it is natural to try to go from stability results for optimal maps to stability results for optimal potentials. We should note, additionally, that the points of nondifferentiability of the potentials are those points in which the superdifferentials are not singletons and that, for this reason, the better way to deal with stability of the optimal plans is to think of them as multivalued maps ($\mathbf{x} \mapsto \partial^c\varphi_n(\mathbf{x}) \subset \mathbb{R}^d$) or, equivalently, as subsets ($\partial^c\varphi_n = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^d : \mathbf{y} \in \partial^c\varphi_n(\mathbf{x})\}$, the *graph* of $\partial^c\varphi_n$). The notion of convergence that fits our goals is the commonly called Painlevé-Kuratowski convergence (see R. Tyrrell Rockafellar (1998)), which is defined as follows: for a sequence $\{\Gamma_n\}_{n\in\mathbb{N}}$ of subsets of $\mathbb{R}^m$

- the outer limit, $\limsup_n \Gamma_n$, is the set of $\mathbf{x} \in \mathbb{R}^m$ for which there exists a sequence $\{\mathbf{x}_n\}$ with $\mathbf{x}_n \in \Gamma_n$ such that there exists a subsequence which converges to $\mathbf{x}$,
- the inner limit, $\liminf_n \Gamma_n$, is the set of $\mathbf{x} \in \mathbb{R}^m$ for which there exists a sequence $\{\mathbf{x}_n\}$ with $\mathbf{x}_n \in \Gamma_n$ which converges to $\mathbf{x}$.

When the outer and inner limit sets are equal the sequence is said to converge in the Painlevé-Kuratowski sense and the common set is the limit. This notion of convergence is automatically transferred easily to multivalued maps. In this case $\{T_n\}_{n\in\mathbb{N}}$, where $T_n : \mathbb{R}^d \to 2^{\mathbb{R}^d}$, is said to *converge graphically* to another multivalued map T if

$$\mathrm{Gph}(T_n) := \{(\mathbf{x}, \mathbf{y}) : \mathbf{y} \in T_n(\mathbf{x})\} \to \mathrm{Gph}(T)$$

in the Painlevé-Kuratowski sense. A very convenient feature of the Painlevé-Kuratowski sense is that sequential compactness can be easily described in terms of a simple condition. To be precise, it is said that a sequence of sets $\Gamma_n \subset \mathbb{R}^d$, $n \geq 1$, does not *escape to the horizon* if there exist $\epsilon > 0$ and some subsequence $\{n_j\}$ such that $\Gamma_{n_j} \cap B(\mathbf{0}, \epsilon) \neq \emptyset$ for all $j \geq 1$. For convenience we quote next a version of Theorem 4.18 in R. Tyrrell Rockafellar (1998).

Theorem 2.2.8. *Let $\{\Gamma_n\}_{n\geq 1}$ be a sequence of subsets of $\mathbb{R}^m$ that does not escape to horizon, then there exists a subsequence $\{n_{j_k}\}$ and a nonempty subset $\Gamma \subset \mathbb{R}^m$ such that*

$$\Gamma_{n_{j_k}} \longrightarrow \Gamma, \quad \textit{in the sense of Painlevé-Kuratowski.}$$

In the next theorem we show that when a sequence of c-cyclically monotone sets converges in the sense of Painlevé-Kuratowski to a set, then it is also c-cyclically monotone, generalizing the result for classical convexity in R. Tyrrell Rockafellar (1998).

Lemma 2.2.9. *Assume c is a continuous cost function and $\{\Gamma_n\}_{n\in\mathbb{N}} \subset \mathbb{R}^{2d}$ is a sequence of c-cyclically monotone sets. If $\Gamma_n \to \Gamma$ in the sense of Painlevé-Kuratowski, then Γ is also c-cyclically monotone.*

Proof. We consider $\{(\mathbf{x}_k, \mathbf{y}_k)\}_{k=1}^m \subset \Gamma$. For each pair $(\mathbf{x}_k, \mathbf{y}_k)$ there exists a sequence $(\mathbf{x}_k^n, \mathbf{y}_k^n) \in \Gamma_n$ such that $(\mathbf{x}_k^n, \mathbf{y}_k^n) \to (\mathbf{x}_k, \mathbf{y}_k)$ as $n \to \infty$. Since Γ_n is c-cyclically monotone,

$$\sum_{k=1}^m c(\mathbf{x}_k^n, \mathbf{y}_k^n) \leq \sum_{k=1}^m c(\mathbf{x}_{\sigma(k)}^n, \mathbf{y}_k^n),$$

for every permutation σ of $\{1, \dots, m\}$. Continuity of c guarantees that

$$\sum_{k=1}^m c(\mathbf{x}_k, \mathbf{y}_k) \leq \sum_{k=1}^m c(\mathbf{x}_{\sigma(k)}, \mathbf{y}_k).$$

□

Combining the last two results we see that if a sequence of c-superdifferentials does not escape to the horizon, then there exists a converging subsequence to a set and this set is also c-cyclically monotone.

We finish the section with a weak continuity result for the multivalued map $\partial^c\psi$, which will be very useful in the following section.

Lemma 2.2.10. *Assume $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$ with h satisfying* (A1)-(A3). *Let f be a $c-$concave function and $\mathbf{x} \in \mathrm{dom}(\nabla^c f)$. Then for each sequence $\mathbf{x}_n \to \mathbf{x}$ and $\mathbf{y}_n \in \partial^c f(\mathbf{x}_n)$ we have that $\mathbf{y}_n \to \nabla^c f(\mathbf{x})$. As a consequence, for each $\epsilon > 0$ there exists some $\delta > 0$ such that $\partial^c f(B(\mathbf{x}, \delta)) \subset B(\nabla^c f(\mathbf{x}), \epsilon)$.*

Proof. Let $(\mathbf{x}_n, \mathbf{y}_n)$ be as in the statement. Then for every $\mathbf{z} \in \mathbb{R}^d$ we have

$$f(\mathbf{z}) \leq f(\mathbf{x}_n) + [c(\mathbf{z}, \mathbf{y}_n) - c(\mathbf{x}_n, \mathbf{y}_n)]. \tag{2.10}$$

Since f is differentiable at $\mathbf{x}$, it is bounded in a neighbourhood of $\mathbf{x}$, say U, which can chosen to be compact. By Proposition C.4 in Gangbo and McCann (1996) $\partial^c f(U)$ is bounded. Hence, the sequence $\mathbf{y}_n$ must be bounded and, taking subsequences if necessary, we can assume that it is convergent. Taking limits in (2.10) and noticing that f is continuous in its domain we get the first conclusion. To check the second claim, assume it is false. Then we can choose some $\epsilon > 0$ such that for each $n \in \mathbb{N}$ there exists $|\mathbf{x}_n - \mathbf{x}| \leq \frac{1}{n}$ and some $\mathbf{y}_n \in \partial^c f(\mathbf{x}_n)$ with $|\mathbf{y}_n - \nabla^c f(\mathbf{x})| > \epsilon$. To conclude note that the sequences $\{\mathbf{x}_n\}_{n\in\mathbb{N}}$ and $\{\mathbf{y}_n\}_{n\in\mathbb{N}}$ lead to a contradiction with the first assertion. □

2.3 Stability of Optimal Transport Potential and Map Under General Costs

The main goal of this section is to prove a general result (Theorem 2.3.4) on the stability of optimal maps and potentials for a very large class of costs, using the tools presented in section 2.2. The path to this main result starts by proving stability along subsequences of the c-superdifferentials of optimal transport potentials (Lemma 2.3.1), extending a similar result in del Barrio and Loubes (2017) for the particular setup of classical convexity. We prove then (Lemmas 2.3.2 and 2.3.3) a uniform boundedness result which, once the potentials are conveniently fixed at a convenient point (see (2.12) below) allows to prove the anticipated stability result. For the sake of readability we present here the results and defer most of the proofs to the Appendix.

The first step in the plan above is this result on the stability of c-superdifferentials.

Lemma 2.3.1. *Let $Q \in \mathcal{P}(\mathbb{R}^d)$ be such that $Q \ll \ell_d$ and has connected support and negligible boundary. Let $Q_n, P_n, P \in \mathcal{P}(\mathbb{R}^d)$ be such that $P_n \xrightarrow{w} P$, $Q_n \xrightarrow{w} Q$ and*

$$\mathcal{T}_c(P_n, Q_n) < \infty \text{ and } \mathcal{T}_c(P, Q) < \infty, \text{ for all } n \in \mathbb{N},$$

for a cost $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$ with h differentiable and satisfying (A1)-(A3). *If ψ_n (resp. ψ) are optimal transport c-potentials from Q_n to P_n (resp. from Q to P), then there exists a cyclically monotone set Γ such that*

$$\partial^c \psi_n \to \Gamma \subset \partial^c \psi \tag{2.11}$$

in the sense of Painlevé-Kuratowski along subsequences. Moreover, if $\mathbf{x} \in dom(\nabla^c \psi) \cap Supp(Q)$, then $(\mathbf{x}, \nabla^c \psi(\mathbf{x})) \in \Gamma$.

In our next results we pay attention to the optimal transportation potentials, ψ_n, which are well-defined up to the addition of a constant. The possibility of arbitrarily choosing that

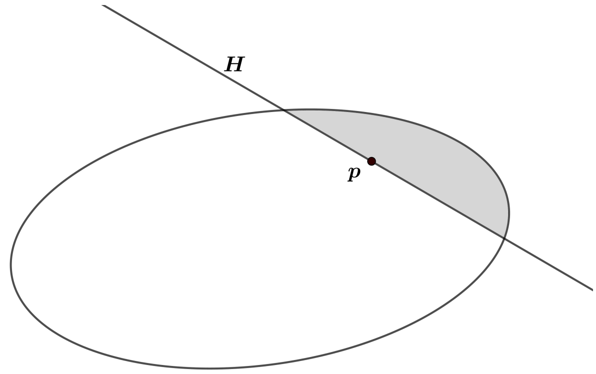

Figure 2.1: Geometric interpretation of Lemma 2.3.2 and Lemma 2.3.3.

constant could lead to some difficulties that we can avoid fixing it as follows. We choose some $\mathbf{p}_0 \in \mathrm{dom}(\nabla\psi) \cap \mathrm{Supp}(Q)$ and assume

$$\psi(\mathbf{p}_0) = 0 \quad \text{and} \quad \psi_n(\mathbf{p}_0) = 0 \text{ for large } n. \tag{2.12}$$

Of course, we can ensure that the potential ψ vanishes at any $\mathbf{p}_0$ where it is finite by taking $\tilde{\psi}(\mathbf{x}) = \psi(\mathbf{x}) - \psi(\mathbf{p}_0)$. Under the assumptions of Lemma 2.3.1 (see the proof for further details) we must have $\mathbf{p}_0 \in \mathrm{dom}(\nabla\psi_n)$ for large enough n, hence, $\mathbf{p}_0 \in \mathrm{dom}(\psi_n)$ and we can choose the potentials as in (2.12).

Next, we present two technical lemmas in which the assumptions (A2) and (A3) play the main roles. These results, crucial in the proof of Theorem 2.3.4, are proved elaborating on the arguments in Gangbo and McCann (1996) to prove that a c-concave function is locally Lipschitz. The geometric interpretation of these results is shown in Figure 2.1. Lemma 2.3.2 shows that for any point $\mathbf{p}$ for which the boundedness condition fails, there is a hyperplane H passing trough $\mathbf{p}$ and splitting the space into two parts such that in one of both, the grey one in Figure 2.1, this property holds for any other point.

Lemma 2.3.2. *Under the same assumptions as in Lemma 2.3.1, let $\mathbf{p} \in \mathbb{R}^d$ be such that there exists a sequence $\{\mathbf{p}_n\}_{n\in\mathbb{N}} \subset \mathbb{R}^d$ such that $\mathbf{p}_n \to \mathbf{p}$ and $\psi_n(\mathbf{p}_n)$ is not bounded. Then there exists $\mathbf{z} \in \mathbb{R}^d$ such that, for every bounded sequence $\{\mathbf{x}_n\}_{n\in\mathbb{N}} \subset\subset \{\mathbf{x} : \langle \mathbf{z}, \mathbf{x} - \mathbf{p}\rangle > 0\}$, the sequence $\psi_n(\mathbf{x}_n)$ is not bounded.*

Lemma 2.3.2 is the key to the next technical result, which proves boundedness of both $\bigcup_{k\in\mathbb{N}} \psi_{n_k}(K)$ and $\bigcup_{k\in\mathbb{N}} \partial^c\psi_{n_k}(K)$ for compact $K \subset \mathrm{Supp}(Q)$.

Lemma 2.3.3. *Let P, Q, P_n, Q_n be probability measures satisfying the assumptions of Lemma 2.3.1. Assume that $\mathbf{p}_0 \in \mathrm{Supp}(Q)$ and $\psi_n(\mathbf{p}_0) \to 0$. Then for each compact $K \subset \mathrm{Supp}(Q)$ there exists a subsequence $\{\psi_{n_k}\}_{k\in\mathbb{N}}$ such that $\bigcup_{k\in\mathbb{N}} \psi_{n_k}(K)$ and $\bigcup_{k\in\mathbb{N}} \partial^c\psi_{n_k}(K)$ are bounded sets.*

Now, as an application of the uniform boundedness results in Lemma 2.3.3, we are ready to apply the classical Arzelà-Ascoli theorem to prove the main theorem of the section.

Theorem 2.3.4. *Let $Q \in \mathcal{P}(\mathbb{R}^d)$ be such that $Q \ll \ell_d$ and has a connected support with negligible boundary. Assume $Q_n, P_n, P \in \mathcal{P}(\mathbb{R}^d)$ are such that $P_n \stackrel{w}{\to} P$, $Q_n \stackrel{w}{\to} Q$ and*

$$\mathcal{T}_c(P_n, Q_n) < \infty \ \text{ and } \ \mathcal{T}_c(P, Q) < \infty$$

for a cost $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$, with h differentiable and satisfying (A1)-(A3). *If ψ_n (resp. ψ) are optimal transport potentials from Q_n to P_n (resp. from Q to P) for the cost c. Then:*

(i) There exist constants $a_n \in \mathbb{R}$ such that $\tilde{\psi}_n := \psi_n - a_n \to \psi$ in the sense of uniform convergence on the compacts sets of $\text{Supp}(Q)$.

(ii) For each compact $K \subset \text{Supp}(Q) \cap \text{dom}(\nabla\psi)$

$$\sup_{\mathbf{x} \in K} \sup_{\mathbf{y}_n \in \partial^c \psi_n(\mathbf{x})} |\mathbf{y}_n - \nabla^c \psi(\mathbf{x})| \longrightarrow 0. \tag{2.13}$$

We note that Theorem 2.3.4 generalizes Theorem 2.8 in del Barrio and Loubes (2017) to a more general class of costs. Moreover, it also generalizes the results of stability of optimal transport maps, as Corollary 5.23 in Villani (2008). An important improvement of Theorem 1.52. in Santambrogio (2015) is obtained since we do not require a compact assumption. Finally we will see in the following sections that it is a useful tool to prove a Central Limit Theorem for general Wasserstein distances.

Under stronger assumptions on the way that P_n approaches P and Q_n approaches Q it is possible to prove L^2 convergence of the potentials. We show this next for potential costs. We recall that the hypotheses of Corollary 2.3.5 are fulfilled when we have weak convergence $P_n \stackrel{w}{\to} P$, $Q_n \stackrel{w}{\to} Q$ plus convergence of moments of order $2p$,

$$\int |\mathbf{x}|^{2p} dP_n(\mathbf{x}) \longrightarrow \int |\mathbf{x}|^{2p} dP(\mathbf{x}), \quad \int |\mathbf{y}|^{2p} dQ_n(\mathbf{y}) \longrightarrow \int |\mathbf{y}|^{2p} dQ(\mathbf{y}).$$

Corollary 2.3.5. *Let $Q \in \mathcal{P}_{2p}(\mathbb{R}^d)$ be such that $Q \ll \ell_d$ and has connected support with negligible boundary. Assume $P_n, P \in \mathcal{P}(\mathbb{R}^d)$ are such that*

$$\mathcal{T}_{2p}(P_n, P) \to 0.$$

If ψ_n (resp. ψ) are optimal transport potentials from Q to P_n (resp. from Q to P) for the cost $c_p(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|^p$ and $p > 1$, then there exist constants $a_n \in \mathbb{R}$ such that $\tilde{\psi}_n := \psi_n - a_n \to \psi$ in the sense of $L^2(Q)$.

Proof. We can apply Theorem 2.3.4 to see that there exist constants $a_n \in \mathbb{R}$ such that $\tilde{\psi}_n = \psi_n - a_n \to \psi$ and $\nabla^c \psi_n \to \nabla^c \psi$ Q-a.s. We note also that the assumption $\mathcal{T}_{2p}(P_n, P) \to 0$ implies that

$$\int |\nabla^c \psi_n(\mathbf{y})|^{2p} dQ(\mathbf{y}) = \int |\mathbf{x}|^{2p} dP_n(\mathbf{x}) \longrightarrow \int |\mathbf{x}|^{2p} dP(\mathbf{x})$$

and, therefore $|\nabla^c\psi_n|^{2p}$ is Q-uniformly integrable. We relabel the potentials and write ψ_n instead of $\tilde{\psi}_n$ and assume (with no loss of generality) that $\psi(\mathbf{x}_0) = \psi_n(\mathbf{x}_0) = 0$ for some $\mathbf{x}_0 \in \text{Supp}(Q) \cap \text{dom}(\nabla\psi)$. To conclude, it suffices to show that ψ_n^2 is Q-uniformly integrable. To check this we set $\mathbf{y}_0 = \nabla^c\psi(\mathbf{x}_0)$, take $\mathbf{y}_n \in \partial^c\psi_n(\mathbf{x}_0)$ and recall that, by Theorem 2.3.4 $\mathbf{y}_n \to \mathbf{y}_0$. Now, we observe that

$$\psi_n(\mathbf{x}) \leq \psi_n(\mathbf{x}_0) + |\mathbf{x} - \mathbf{y}_n|^p - |\mathbf{x}_0 - \mathbf{y}_n|^p \leq |\mathbf{x} - \mathbf{y}_n|^p \tag{2.14}$$

for every $\mathbf{x}$. Similarly,

$$\psi_n^c(\mathbf{y}) \leq \psi_n^c(\mathbf{y}_n) + |\mathbf{y} - \mathbf{x}_0|^p - |\mathbf{y}_n - \mathbf{x}_0|^p = |\mathbf{y} - \mathbf{x}_0|^p$$

for every $\mathbf{y}$. Since Q-a.s. we have $\psi_n(\mathbf{x}) + \psi_n^c(\nabla^c\psi_n(\mathbf{x})) = |\mathbf{x} - \nabla^c\psi_n(\mathbf{x})|^p$, we conclude that

$$\psi_n(\mathbf{x}) \geq |\mathbf{x} - \nabla^c\psi_n(\mathbf{x})|^p - |\nabla^c\psi_n(\mathbf{x}) - \mathbf{x}_0|^p, \;\; Q - a.s.$$

This last bound together with (2.14) shows that ψ_n^2 is Q-uniformly integrable and completes the proof. □

2.4 Central Limit Theorem and Variance Bounds

2.4.1 One-sample case

Let $P \in \mathcal{P}(\mathbb{R}^d)$ and for each $n \in \mathbb{N}$ let $X_1, \dots, X_n$ denote a a sample of independent random variables with distribution P. Consider also the correspondent empirical measure $P_n := \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$. We are interested in the behavior of the sequence $\{\sqrt{n}\,(\mathcal{T}_p(P_n, Q) - E\mathcal{T}_p(P_n, Q))\}_{n\in\mathbb{N}}$. We will prove first tightness of this sequence from a suitable variance bound, following similar arguments as those in del Barrio and Loubes (2017). We recall the Efron-Stein inequality and refer for further details to Chapter 3.1 in Boucheron et al. (2013). Let $(X_1', \dots, X_n')$ be an independent copy of $(X_1, \dots, X_n)$, set $Z := f(X_1, \dots, X_n)$ and for each $i \in \{1, \dots, n\}$ denote

$$Z_i' := f(X_1, \dots, X_{i-1}, X_i', X_{i+1}, \dots, X_n).$$

The Efron-Stein inequality states then that

$$\text{Var}(Z) \leq \frac{1}{2}\sum_{i=1}^n E(Z - Z_i')^2 = \sum_{i=1}^n E(Z - Z_i')_+^2,$$

where $(\cdot)_+$ denotes the positive part. Note that when $X_1, \dots, X_n$ are i.i.d, the inequality can be written as

$$\text{Var}(Z) \leq \frac{n}{2}E(Z - Z_i')^2 = nE(Z - Z_i')_+^2.$$

In this work we present a general bound for the variance of $\mathcal{T}_c(P_n, Q)$ assuming only that one of both probabilities is absolutely continuous with respect to Lebesgue measure and assuming also that the cost is convex. We note that for X with law P the set of points where $h(X - \cdot)$ is not differentiable is a set of Lebesgue measure 0, hence if $Q \ll \ell_d$ then it is differentiable Q-a.s. As a consequence $\nabla h(X - \mathbf{y})$ is well defined Q- a.s., and also $E|\nabla h(X - Y)|^{2q_2}$ in the next statement.

Lemma 2.4.1. *Assume $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$, with h satisfying (A1)-(A3). Let $P, Q \in \mathcal{P}(\mathbb{R}^d)$ be such that $Q \ll \ell_d$. Assume X, X', Y are independent random variables with $X \sim P$, $X' \sim P$ and $Y \sim Q$. Then*

$$n\text{Var}(\mathcal{T}_c(P_n, Q)) \leq \inf_{(q_1,q_2)\in\alpha} \left[\left(E|X - X'|^{2q_1}\right)^{\frac{1}{q_1}} \left(E|\nabla h(X - Y)|^{2q_2}\right)^{\frac{1}{q_2}} \right], \tag{2.15}$$

where $\alpha = \{(q_1, q_2) : q_i \in [1, \infty], \frac{1}{q_1} + \frac{1}{q_2} = 1\}$.

We remark that assumptions (A1)-(A3) are only used in Lemma 2.4.1 to ensure the existence of an optimal transport map.

Remark 2.4.2. *As a consequence of Lemma 2.4.1, under the same assumptions, if*

$$\inf_{(q_1,q_2)\in\alpha} \left[\left(E|X - X'|^{2q_1}\right)^{\frac{1}{q_1}} \left(E|\nabla h(X - Y)|^{2q_2}\right)^{\frac{1}{q_2}} \right] < \infty, \tag{2.16}$$

then the sequence $\{\sqrt{n}\,(\mathcal{T}_c(P_n, Q) - E\mathcal{T}_c(P_n, Q))\}_{n\in\mathbb{N}}$ is tight.

We show next that we can replace assumption (2.16) with a simpler version in the case of potential costs. It should be noted that absolute continuity of Q is not needed for the following result.

Corollary 2.4.3. *If $c(\boldsymbol{x}, \boldsymbol{y}) = |\boldsymbol{x} - \boldsymbol{y}|^p$ and $p > 1$ then*

$$n\text{Var}(\mathcal{T}_p(P_n, Q)) \leq \left(E|X - X'|^{2p}\right)^{\frac{1}{p}} \left(pE|X - Y|^{2p}\right)^{\frac{p}{p-1}}$$

Proof. We assume that the right hand side in the last bound is finite (there is nothing to prove otherwise). Since $|\nabla h(X_1 - \mathbf{y})| = p|X_1 - \mathbf{y}|^{p-1}$, the result follows by taking $q_1 = p$, $q_2 = \frac{p}{p-1}$ in (2.15) if $Q \ll \ell_d$. For general Q we can take random variables $Y \sim Q$, $Y_m \sim Q_m$, $m \in \mathbb{N}$ with $Q_m \ll \ell_d$ and $E|Y_m - Y|^{2p} \to 0$. Without loss of generality we can assume that (X, X') is independent of $(Y, \{Y_m\}_{m\geq 1})$. For fixed $n \in \mathbb{N}$ we have that $\mathcal{T}_p(P_n, Q_m)$ converges to $\mathcal{T}_p(P_n, Q)$ a.s. as $m \to \infty$. Also, for each $m \in \mathbb{N}$, we have

$$n\text{Var}(\mathcal{T}_p(P_n, Q_m)) \leq \left(E|X - X'|^{2p}\right)^{\frac{1}{p}} \left(pE|X - Y_m|^{2p}\right)^{\frac{p}{p-1}} =: A_m.$$

We observe that $A_m \to A := \left(E|X - X'|^{2p}\right)^{\frac{1}{p}} \left(pE|X - Y|^{2p}\right)^{\frac{p}{p-1}}$. Finally, Fatou's lemma enables us to conclude that

$$n\text{Var}(\mathcal{T}_p(P_n, Q)) \leq n \liminf_m \text{Var}(\mathcal{T}_p(P_n, Q_m)) \leq \liminf_m A_m = A.$$

□

Remark 2.4.4. *As in Remark 2.4.2, Corollary 2.4.3 yields the conclusion that* $\{\sqrt{n}\,(\mathcal{T}_p(P_n,Q)-E\mathcal{T}_p(P_n,Q))\}_{n\in\mathbb{N}}$ *is tight if P and Q have finite moments of order $2p$. This assumption is sharp in the sense that if P is such that* $\{\sqrt{n}\,(\mathcal{T}_p(P_n,Q)-E\mathcal{T}_p(P_n,Q))\}_{n\in\mathbb{N}}$ *for $Q=\delta_{\mathbf{0}}$ then P must have finite moment of order $2p$. In fact, the optimal transport map from P_n to Q is $T(\mathbf{x})=\mathbf{0}$, hence,* $\mathcal{T}_p(P_n,Q)=\int|\mathbf{x}|^p dP_n(\mathbf{x})$ *and*

$$\sqrt{n}\,(\mathcal{T}_p(P_n,Q)-E\mathcal{T}_p(P_n,Q))=\tfrac{1}{\sqrt{n}}\textstyle\sum_{j=1}^{n}(|X_j|^p-E|X_1|^p). \tag{2.17}$$

It is well known (see, e.g., Chapter 10 in Ledoux and Talagrand (1991)) that the sequence of (2.17) is tight if and only if $E(|X_1|^{2p})<\infty$. Hence, as claimed, a finite moment of order $2p$ is a minimal requirement for P to guarantee that $\{\sqrt{n}\,(\mathcal{T}_p(P_n,Q)-E\mathcal{T}_p(P_n,Q))\}_{n\in\mathbb{N}}$ *is tight for, say, every Q with bounded support.*

Condition (2.16) is enough to achieve tightness with a cost c satisfying assumptions (A1)-(A3). In the following theorem we show that, with this assumptions on the cost, there exists a unique weak cluster point of the sequence $\{\sqrt{n}\,(\mathcal{T}_c(P_n,Q)-E\mathcal{T}_c(P_n,Q))\}_{n\in\mathbb{N}}$, which is Gaussian. Similar work, in the particular case of the cost $|\cdot|^2$, was done in del Barrio and Loubes (2017), where a version of Efron-Stein inequality is used to prove that the empirical transport cost is approximately linear. This approach has also been used for the entropic regularization of the empirical transport cost in Mena and Niles-Weed (2019). This tool based on Efron-Stein inequality requires to have some sort of uniform integrability, which can be guaranteed assuming finite moments of order $4+\delta$. Following arguments developed in Remark 2.4.4, the following result proves that the moment assumption can be relaxed.

Theorem 2.4.5. *Assume $c(\mathbf{x},\mathbf{y})=h(\mathbf{x}-\mathbf{y})$ with h differentiable and satisfying* (A1)-(A3). *Let $P,Q\in\mathcal{P}(\mathbb{R}^d)$ be such that $P\ll\ell_d$, $Q\ll\ell_d$, and P has connected support and negligible boundary. Assume further that*

$$\int h(2\mathbf{x})^2dP(\mathbf{x})<\infty \quad and \quad \int h(-2\mathbf{y})^2dQ(\mathbf{y})<\infty, \tag{2.18}$$

and (2.16) holds. Then

$$\sqrt{n}\,(\mathcal{T}_c(P_n,Q)-E\mathcal{T}_c(P_n,Q))\xrightarrow{w}N(0,\sigma_c^2(P,Q)), \tag{2.19}$$

where

$$\sigma_c^2(P,Q):=\int\varphi(\mathbf{x})^2dP(\mathbf{x})-\left(\int\varphi(\mathbf{x})dP(\mathbf{x})\right)^2, \tag{2.20}$$

and φ is an optimal transport potential for the cost c from P to Q.

It should be noted at this point that the optimal transport potential in Theorem 2.4.5 is unique, up to the addition of a constant, as a consequence of Corollary 2.2.7. It follows from the proof of Theorem 2.4.5 that $\varphi\in L^2(P)$. This implies that the limiting variance,

$\sigma_c^2(P, Q)$, is well-defined and finite. It is worth noting that the assumption (2.18) can be relaxed to

$$h(X_1 - \nabla^c \varphi(X_1)) \leq t\, h(\frac{1}{t} X_1) + (1-t)\, h(-\frac{1}{(1-t)} \nabla^c \varphi(X_1)),$$

for any choice of $0 < t < 1$, see Remark 2..1.

The proof of Theorem 2.4.5 initially follows the path in del Barrio and Loubes (2017). This means that we look at

$$R_n := \mathcal{T}_c(P_n, Q) - \int \varphi(\mathbf{x}) dP_n(\mathbf{x}), \tag{2.21}$$

where φ is an optimal transport potential from P to Q for the cost c. We write R_n' for the version of R_n computed from $X_1', X_2, \ldots, X_n$. Using the stability results for optimal transport potentials one can prove that $n(R_n - R_n') \xrightarrow{a.s.} 0$. If $n^2 E(R_n - R_n')^2 \to 0$ then the conclusion in Theorem 2.4.5 follows inmediately. Variance bounds obtained from the Efron-Stein inequality yield $n^2 E(R_n - R_n')^2 \leq M$ under mild moment assumptions. However, the convergence $n^2 E(R_n - R_n')^2 \to 0$ may fail wihtout some stronger assumptions (such as the $4 + \delta$ moment assumption in del Barrio and Loubes (2017)). Our proof of Theorem 2.4.5 avoids these stronger assumptions by using the following workaround. First, the bound $n^2 E(R_n - R_n')^2 \leq M$ and the Banach-Alaoglu Theorem (see, e.g., Theorem 3.16 in Brezis (2011)) show that, along subsequences, $n(R_n - R_n')$ converges weakly to 0 in the Hilbert (hence reflexive) space $L^2(\mathbb{P})$. Then, the Banach-Saks property of Hilbert spaces (see, e.g., Exercise 5.34 in Brezis (2011)) shows that (taking further subsequences if necessary) there exists a Cesàro mean of $\{n|R_n - R_n'|\}_{n\in\mathbb{N}}$ convergent to 0 in $L^2(\mathbb{P})$ in the strong sense. We show then that the same holds with the Cesàro means of the sequence $\sqrt{n}(R_n - ER_n)$ and from this we conclude that $\sqrt{n}(R_n - ER_n) \to 0$ in probability, which yields, as a consequence, (2.19). All the details are given in the proof postponed to the Appendix.

In general it is not possible to guarantee moment convergence in (2.19) under the minimal assumptions of Theorem 2.4.5. The following theorem guarantees convergence of variances under slightly stronger assumptions.

Theorem 2.4.6. *Assume $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$ with h differentiable and satisfying* (A1)-(A3). *Let $P, Q \in \mathcal{P}(\mathbb{R}^d)$ be such that $P \ll \ell_d$, $Q \ll \ell_d$ and P has connected support and negligible boundary. Suppose that (2.18) holds and assume R_n is as in (2.21). Assume further that X, X' and Y are independent random variables with $X \sim P$, $X' \sim P$ and $Y \sim Q$. If there exists some $\delta > 0$ such that ,*

$$\inf_{q_1, q_2 \in [1,\infty]:\ \frac{1}{q_1} + \frac{1}{q_2} = 1} \left[E|X - X'|^{(2+\delta)q_1} E|\nabla h(X - Y)|^{(2+\delta)q_2} \right] < \infty, \tag{2.22}$$

then $n\mathrm{Var}(R_n) \longrightarrow 0$. As a consequence,

$$n\mathrm{Var}(\mathcal{T}_c(P_n, Q)) \longrightarrow \sigma_c^2(P, Q). \tag{2.23}$$

To get a more clear picture about the sharpness of the assumptions in Theorems 2.4.5 and 2.4.6, we include the particular version for potential costs, $c_p(\mathbf{x},\mathbf{y}) = |\mathbf{x}-\mathbf{y}|^p$ for $p > 1$ (recall from Remark 2.2.2 that c_p satisfies (A1)-(A3) for $p > 1$).

Corollary 2.4.7. *Assume $p > 1$. Let $P, Q \in \mathcal{P}(\mathbb{R}^d)$ be such that $P \ll \ell_d$ and has connected support and negligible boundary. If P and Q have finite moments of order $2p$, then*

$$\sqrt{n}\left(\mathcal{T}_p(P_n,Q) - E\mathcal{T}_p(P_n,Q)\right) \xrightarrow{w} N(0,\sigma_p^2(P,Q)), \tag{2.24}$$

where

$$\sigma_p^2(P,Q) := \int \varphi(\mathbf{x})^2 dP(\mathbf{x}) - \left(\int \varphi(\mathbf{x}) dP(\mathbf{x})\right)^2, \tag{2.25}$$

and φ is an optimal transport potential from P to Q for c_p. Moreover if P has a finite moment of order $2p+\epsilon$ for some $\epsilon > 0$, then

$$n\mathrm{Var}(\mathcal{T}_p(P_n,Q)) \longrightarrow \sigma_p^2(P,Q). \tag{2.26}$$

Proof. A look at the proof of Corollary 2.4.3 shows that finite $2p$ moments guarantee that (2.16) holds. Clearly, (2.18) holds too, and we can apply Theorem 2.4.5 to conclude (2.24) (the fact that absolute continuity of Q is not necessary follows using the approximation argument in the proof of Corollary 2.4.3). For (2.26) we take in (2.22) the conjugate pair $q_1 = \frac{2p+\epsilon}{\epsilon-\delta p+2+\delta}$ and

$$q_2 = \frac{q_1}{q_1-1} = \frac{2p+\epsilon}{2p+\delta p-2-\delta} = \frac{2p+\epsilon}{(2+\delta)(p-1)},$$

where $\delta = \frac{\epsilon}{p}$. With this choices we observe that

$$\begin{aligned}\left(E|X_1 - X_1'|^{(2+\delta)q_1}\right)&\left(E\left(\int_{\mathbb{R}^d}|X_1-\mathbf{y}|^{q_2(2+\delta)(p-1)}dQ(\mathbf{y})\right)\right)\\ &= \left(E|X_1-X_1'|^{(2p+\epsilon)}\right)\left(E\left(\int_{\mathbb{R}^d}|X_1-\mathbf{y}|^{2p+\epsilon}dQ(\mathbf{y})\right)\right) < \infty,\end{aligned}$$

see that(2.22) becomes

and we apply Theorem 2.4.6. The case of finite moment of order $2p+\epsilon$ for Q follows similarly. □

Remark 2.4.8. *As noted in Remark 2.4.4, the assumption of finite moments of order $2p$ (ar least for P) cannot be relaxed for tightness and, in that sense, the moment assumptions in Theorem 2.4.7 are sharp and cannot be improved. On the other hand, in the case $p = 2$, Corollary 2.4.7 improves Theorem 4.1 in* del Barrio and Loubes (2017)*, not only by proving that finite fourth moments are enough (the original assumption was finite moments of order $4+\epsilon$ in* del Barrio and Loubes (2017)*), but also by assuming milder regularity assumptions on P and Q. In this new setting, P must have a connected support with a negligible boundary, relaxing the assumption of a convex support. The only price to pay is that variance convergence may fail under this relaxed assumptions.*

So far we have considered CLTs for $\mathcal{T}_p(P,Q)$. Its $p-$root $\mathcal{W}_p(P,Q) := (\mathcal{T}_p(P,Q))^{1/p}$ defines a well-known metric in the space of probabilities with finite moments of order p, the p-Wasserstein distance. Proving a CLT for the empirical Wasserstein distance is not a straightforward application of a delta-method and Corollary 2.4.7, since we do not have a fixed centering constant in Theorem 2.4.10. Yet, we can circumvent this issue and prove the following result.

Theorem 2.4.9. *Let $P \neq Q \in \mathcal{P}(\mathbb{R}^d)$ be such that $P \ll \ell_d$ and has connected support and negligible boundary. Assume P and Q have finite moments of order $2p$ and $p > 1$. Then, if $\sigma_p^2(P,Q)$ is defined as in Theorem 2.4.7.*

$$\sqrt{n}\left(\mathcal{W}_p(P_n,Q) - \left(E[\mathcal{W}_p^p(P_n,Q)]\right)^{\frac{1}{p}}\right) \xrightarrow{w} N(0, \beta_p^2(P,Q)),$$

where $\beta_p^2(P,Q) := \left(\frac{1}{p\mathcal{W}_p^p(P_n,Q)^{p-1}}\right)^2 \sigma_p^2(P,Q)$.

Proof. Setting

$$A_n := \mathcal{W}_p(P_n,Q) \ \text{ and } \ B_n := \left(E[\mathcal{W}_p^p(P_n,Q)]\right)^{\frac{1}{p}},$$

we know from Corollary 2.4.7 that

$$\sqrt{n}\,(A_n^p - B_n^p) \xrightarrow{w} N(0, \sigma_p^2(P,Q)). \tag{2.27}$$

Moreover, the bound

$$\mathcal{W}_p^p(P_n,Q) \leq 2^{p-1}\int |\mathbf{x}|^p dP_n(\mathbf{x}) + 2^{p-1}\int |\mathbf{y}|^p dQ(\mathbf{y}),$$

together with the assumption of finite moments of order $2p$, imply that $\mathcal{W}_p^p(P_n,Q)$ is uniformly integrable. It follows that

$$A_n \xrightarrow{a.s.} \mathcal{W}_p(P,Q), \text{ and } B_n \to \mathcal{W}_p(P,Q). \tag{2.28}$$

By the mean value theorem applied to the function $t \mapsto t^p$, there exists $\varepsilon_n \in (0,1)$ such that

$$A_n^p - B_n^p = (A_n - B_n)p(A_n\varepsilon_n + B_n(1-\varepsilon_n))^{p-1}. \tag{2.29}$$

The limits of (2.28) imply that necessarily $p(A_n\varepsilon_n + B_n(1-\varepsilon_n))^{p-1} \xrightarrow{a.s.} p\mathcal{W}_p(P,Q)^{p-1} > 0$. This fact, together with the limit (2.27) and Slutsky's theorem applied in (2.29) conclude the proof. □

2.4.2 Two-sample case

For $n, m \in \mathbb{N}$ let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ be independent i.i.d. random samples with distributions P and Q. Consider the correspondent empirical measures $P_n := \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$ and $Q_m := \frac{1}{m}\sum_{k=1}^m \delta_{Y_k}$. At first sight one may conjecture that the approach leading to Theorems 2.4.5 and 2.4.6 trivially extends to the two-sample setup, yielding a CLT for $\mathcal{T}_c(P_n, Q_m)$. However, a closer look at the proof shows that major issues appear when extending Claim 3—the treatment of the two-sample counterpart of the inequalities of the proof of Theorem 2.4.5 are quite difficult and we could not overcome this obstacle. For this reason an adaptation of Theorem 2.4.5 to the two-sample setup is left for further work. On the other hand, under stronger moment assumptions, such as (2.22), the extension is straightforward. We present the result avoiding additional details.

Theorem 2.4.10. *Assume $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$ with h differentiable and satisfying (A1)-(A3). Let $P, Q \in \mathcal{P}(\mathbb{R}^d)$ be such that $P \ll \ell_d$, $Q \ll \ell_d$ and both have connected support and negligible boundary. Assume that (2.18) holds and also that there exists some $\delta > 0$ such that (2.22) holds, as well as the corresponding conditions exchanging the roles of P and Q. Then, if $\frac{n}{n+m} \to \lambda \in (0,1)$ as $n, m \to \infty$,*

$$\sqrt{\tfrac{nm}{n+m}}\left(\mathcal{T}(P_n, Q_m) - E\mathcal{T}(P_n, Q_m)\right) \xrightarrow{w} N\left(0, (1-\lambda)\sigma_c^2(P, Q) + \lambda\sigma_c^2(Q, P)\right),$$

with $\sigma_c^2(Q, P)$ as in (2.20). Furthermore,

$$n\text{Var}(\mathcal{T}(P_n, Q_m)) \to (1-\lambda)\sigma_c^2(P, Q) + \lambda\sigma_c^2(Q, P).$$

2.4.3 Variance estimation

Theorem 2.4.6 provides the consistence of the variance in the terms of $nVar(\mathcal{T}(P_n, Q)) \to \sigma_c^2(P, Q)$. This allows, for instance, to obtain an estimator by using Monte Carlo methods. But we can also compute it by using the explicit formula

$$\sigma_c^2(P_n, Q) = \int \varphi_n(\mathbf{x})^2 dP_n(\mathbf{x}) - \left(\int \varphi_n(\mathbf{x}) dP_n(\mathbf{x})\right)^2,$$

which is consistent in the sense that $\sigma_c^2(P_n, Q) \xrightarrow{P} \sigma_c^2(P, Q)$. To prove this last claim, it is enough to see that

$$\int (\varphi_n(\mathbf{x})^2 - \varphi(\mathbf{x})^2) dP_n(\mathbf{x}) \xrightarrow{P} 0,$$

which can be derived from

$$\int (\varphi_n(\mathbf{x}) - \varphi(\mathbf{x}))^2 dP_n(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n (\varphi_n(X_i) - \varphi(X_i))^2 \xrightarrow{P} 0. \tag{2.30}$$

The consistency of (2.30) is a direct consequence of the following equality

$$E\left(\frac{1}{n}\sum_{i=1}^{n}(\varphi_n(X_i)-\varphi(X_i))^2\right)=\frac{1}{n}\sum_{i=1}^{n}E\left(\varphi_n(X_i)-\varphi(X_i)\right)^2=E\left(\varphi_n(X_1)-\varphi(X_1)\right)^2,$$

Hence, we can thus conclude as in the proof of Theorem 2.4.6 that $\sigma_c^2(P_n,Q)\xrightarrow{P}\sigma_c^2(P,Q)$. Similar augmentation proves the two sample case.

2.5 Considerations and further work

In the recent years, researchs on weak limits for the transport problem has become a very hot topic in statistics and machine learning. Many of them focus on the analysis of the centered fluctuations $\sqrt{n}\,(\mathcal{T}_c(P_n,Q)-E\mathcal{T}_c(P_n,Q))$. For instance, González-Delgado et al. (2021) extends Theorem 2.4.5 to probabilities on the flat torus with applications in structural biology.
Nevertheless, one of the main problem tackled in this recent literature is about finding cases where such Central Limit Theorem also holds when the centering constant $E\mathcal{T}_c(P_n,Q)$ can be replaced by the true transportation cost, enabling to obtain a CLT for $\sqrt{n}\,(\mathcal{T}_c(P_n,Q)-\mathcal{T}_c(P,Q))$ as in Manole and Niles-Weed (2021) and references therein. A first setting, developed in del Barrio et al. (2021a), deals with the semi-discrete case, where one of the distribution is discrete. In this case, arguments from functional derivation of the supremum, proves that CLT holds, which suggests that the complexity of the optimal transport problem adapts to the one of the discrete probability. This adaptability of the complexity is fulfilled in a much broader way, see Hundrieser et al. (2022b). This culminates in the work of Hundrieser et al. (2022a), where, using the above observations, a complete description of the possible bounds of the transport problem with convergence ratio $\sqrt{n}$ is given.

For regularized transportation, del Barrio et al. (2022) uses the fluctuations result of Mena and Niles-Weed (2019) to derive the central limit theorem of the entropy regularized optimal transport cost centered at the population cost and Manole et al. (2021) for the plugin estimatior.

It is worth noting that both results for the one-sample and for the two-sample cases could be generalized to probabilities, supported in more general spaces—not necessarily connected. This property is necessary only to guarantee the uniqueness (up to additive constants) of the potentials. During the review process of this paper, Staudt et al. (2022) has proved a generalization of Corollary 2.2.7 with yet different conditions. Extending the results we provide to this new case could be the topic of future research.

Acknowledgment

We thank an anonymous reviewer, whose commentaries helps to improve the paper and provides the counterexample described in Remark 2..2.

Appendix

Contents

2..1 Proofs of main results

PROOF OF THEOREM 2.3.4. We prove each claim separately. To prove *(i)* we take, without loss of generality, $\mathbf{p}_0 \in \text{Supp}(Q) \cap \text{dom}(\nabla\psi)$ as in (2.12) (hence, $\psi_n(\mathbf{p}_0) = 0$). From (2.9) and Lemma 2.3.3 we see that for each compact $K \subset \text{Supp}(Q)$ there exists a subsequence ψ_{n_k} and a constant $R = R(K) > 0$ such that for $\mathbf{a}, \mathbf{x} \in K$

$$|\psi_{n_k}(\mathbf{x}) - \psi_{n_k}(\mathbf{a})| \leq |\mathbf{x} - \mathbf{a}|R.$$

Hence, the functions of the sequence $\{\psi_{n_k}\}$ are R-Lipschitz on each compact set and $\psi_{n_k}(\mathbf{p}_0) = 0$ and we can apply Arzelà-Ascoli theorem in each compact set to conclude that there exists a continuous function f such that $\psi_{n_{k_m}} \to f$ uniformly on the compact sets of $\text{Supp}(Q)$ for some subsequence.
We claim that $f = \psi + C$. To prove it we consider $\mathbf{x} \in \text{Supp}(Q)$ and any sequence $\mathbf{y}_n \in \partial^c \psi_n(\mathbf{x})$, by Lemma 2.3.3 we know that there exists a sub-sequence $\{\mathbf{y}_{n_k}\}_{k\in\mathbb{N}}$ which is bounded. Hence, by Lemma 2.3.1, there exists $\mathbf{y} \in \partial\psi(\mathbf{x})$ such that $\mathbf{y}_{n_k} \to \mathbf{y} \in \partial^c\psi(\mathbf{x})$ along a subsequence. We keep the notation for this sub-sequence and note that it satisfies

$$\psi_{n_k}(\mathbf{z}) \leq \psi_{n_k}(\mathbf{x}) + [c(\mathbf{z}, \mathbf{y}_{n_k}) - c(\mathbf{x}, \mathbf{y}_{n_k})] \ \text{ for all } \ \mathbf{z} \in \mathbb{R}^d,$$

and by taking limits,

$$f(\mathbf{z}) \leq f(\mathbf{x}) + [c(\mathbf{z}, \mathbf{y}) - c(\mathbf{x}, \mathbf{y})] \ \text{ for all } \ \mathbf{z} \in \text{dom}(f).$$

Therefore, $\partial^c f(\mathbf{x})$ is non-empty for every $\mathbf{x} \in \text{Supp}(Q)$. This entails that f is $c-$concave and, as a consequence, almost surely differentiable. Moreover, $\mathbf{y} \in \partial^c f(\mathbf{x}) \cap \partial^c\psi(\mathbf{x})$. We conclude that $\nabla^c f = \nabla^c \psi$ a.s. in $\text{Supp}(Q)$ and *(i)* follows by Corollary 2.2.7.

We turn now to *(ii)* and assume, on the contrary, that there exists a sequence $\{\mathbf{x}_n\} \subset K$ and $\mathbf{y}_n \in \partial^c\psi_n(\mathbf{x}_n)$ such that

$$|\mathbf{y}_n - \nabla^c\psi(\mathbf{x}_n)| > \epsilon \text{ for some } \epsilon > 0 \text{ and all } n. \tag{2.31}$$

Compactness of K implies that there exists $\mathbf{x} \in K$ such that $\mathbf{x}_n \to \mathbf{x}$ along a subsequence, which, to ease notation, we denote also as $\mathbf{x}_n$. Lemma 2.3.3 implies that $\mathbf{y}_n$ also converges to some $\mathbf{y}$ along a subsequence. But then Lemma 2.3.1 shows that $\mathbf{y} = \nabla^c\psi(\mathbf{x})$ which contradicts (2.31). □ PROOF OF THEOREM 2.4.5. We write

$(X_1', \ldots, X_n')$ for an independent copy of $(X_1, \ldots, X_n)$ and denote by $P_n^{(i)}$ the empirical measure on $(X_1, \ldots, X_i', \ldots, X_n)$. As in (2.21),

$$R_n = \mathcal{T}_c(P_n, Q) - \int \varphi(\mathbf{x}) dP_n(\mathbf{x}),$$

where φ is an optimal transport potential from P to Q. We write $R_n^{(i)}$ for the version of R_n computed from $P_n^{(i)}$ instead of P_n. To ease notation it will be convenient to write P_n' rather that $P_n^{(1)}$ and R_n' instead of $R_n^{(1)}$ at some points.

The guideline of the proof is to show that $n(R_n - R_n') \xrightarrow{a.s.} 0$ and $n^2 E(R_n - R_n')^2 \leq M$. From this we can obtain, using the Banach-Alaoglu theorem and the Banach-Saks property (see details below), that there exists a Cesàro mean of $\{n|R_n - R_n'|\}_{n\in\mathbb{N}}$ convergent to 0 in $L^2(\mathbb{P})$. Finally the same holds with the Cesàro means of the sequence $\sqrt{n}(R_n - ER_n)$. To conclude we will prove that these three claims imply the central limit theorem. We follow this path in the following complete proof, which we split into three main steps:

Claim 1: $n(R_n - R_n') \xrightarrow{a.s.} 0$ and $n^2 E(R_n - R_n')^2 \leq M$.

We write φ_n for an optimal transport potential between P_n and Q. Since

$$\begin{aligned} \mathcal{T}_c(P_n', Q) &= \sup_{(f,g)\in\Phi_c(P,Q)} \int f(\mathbf{x}) dP_n'(\mathbf{x}) + \int g(\mathbf{y}) dQ(\mathbf{y}) \\ &\geq \int \varphi_n(\mathbf{x}) dP_n'(\mathbf{x}) + \int \varphi_n^c(\mathbf{y}) dQ(\mathbf{y}), \end{aligned}$$

then we have

$$R_n' \geq \frac{1}{n}\varphi_n(X_1') + \frac{1}{n}\sum_{k=2}^n \varphi_n(X_k) - \frac{1}{n}\sum_{k=2}^n \varphi(X_k) - \frac{1}{n}\varphi(X_1') + \int \varphi_n^c(\mathbf{y}) dQ(\mathbf{y}).$$

This implies that

$$R_n - R_n' \leq \frac{1}{n}\big(\varphi_n(X_1) - \varphi(X_1) - \varphi_n(X_1') + \varphi(X_1')\big). \tag{2.32}$$

By Theorem 2.3.4 we can assume, without loss of generality, that, almost surely, $\varphi_n \to \varphi$, uniformly on compact subsets of $\mathrm{Supp}(P)$. This entails that $n(R_n - R_n')_+ \stackrel{a.s.}{\to} 0$. By symmetry, $n(R_n' - R_n)_+ \stackrel{a.s.}{\to} 0$ and we conclude that $n(R_n' - R_n) \stackrel{a.s.}{\to} 0$.

For the second part of this claim we recall that

$$n(R_n - R_n') = n(\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P_n', Q)) - (\varphi(X_1) - \varphi(X_1')).$$

It follows from (2.16) and the proof of Lemma 2.4.1 that $n^2 E(\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P_n', Q))^2$ is a bounded sequence and, therefore, it suffices to show that $E\varphi(X_1)^2 < \infty$. To check this, we fix $\mathbf{x}_0 \in \text{Supp}(P) \cap \text{dom}(\nabla\varphi)$. From (2.8) we get that

$$\begin{aligned} |\varphi(X_1)| &\leq |\varphi(\mathbf{x}_0)| + |c(X_1, \mathbf{y}) - c(\mathbf{x}_0, \mathbf{y})| + |c(X_1, \mathbf{b}) - c(\mathbf{x}_0, \mathbf{b})|, \\ &\leq |\varphi(\mathbf{x}_0)| + c(X_1, \mathbf{y}) + c(\mathbf{x}_0, \mathbf{y}) + c(X_1, \mathbf{b}) + c(\mathbf{x}_0, \mathbf{b}), \end{aligned}$$

for all $(\mathbf{x}_0, \mathbf{b}), (X_1, \mathbf{y}) \in \partial^c\varphi$. Since φ is differentiable at $\mathbf{x}_0$ then if $X_1 \in \text{dom}(\nabla\varphi)$ we have

$$\begin{aligned} |\varphi(X_1)| \quad \leq \quad & |\varphi(\mathbf{x}_0)| + c(X_1, \nabla^c\varphi(X_1)) + c(\mathbf{x}_0, \nabla^c\varphi(X_1)) \\ & + c(X_1, \nabla^c\varphi(\mathbf{x}_0)) + c(\mathbf{x}_0, \nabla^c\varphi(\mathbf{x}_0)). \end{aligned}$$

Recalling that $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$ and that h is convex, we see that

$$c(X_1, \nabla^c\varphi(X_1)) = h(X_1 - \nabla^c\varphi(X_1)) \leq \tfrac{1}{2}h(2X_1) + \tfrac{1}{2}h(-2\nabla^c\varphi(X_1)).$$

Remark 2..1. *Note that, for any choice of* $0 < t < 1$*, we have still the bound*

$$h(X_1 - \nabla^c\varphi(X_1)) \leq t\, h(\frac{1}{t}X_1) + (1-t)\, h(-\frac{1}{(1-t)}\nabla^c\varphi(X_1))$$

and the proof continues in the same way

Hence, using the fact that $Q = \nabla^c\varphi \# P$ and (2.18) we deduce that

$$E(c(X_1, \nabla^c\varphi(X_1)))^2 \leq \int h(2\mathbf{x})^2 dP(\mathbf{x}) + \int h(-2\mathbf{y})^2 dQ(\mathbf{y}) < \infty.$$

Similarly, we check that $E(c(X_1, \nabla^c\varphi(\mathbf{x}_0))^2) < \infty$ and $E(c(\mathbf{x}_0, \nabla^c\varphi(X_1))^2) < \infty$. This shows that $\varphi(X_1)$ has a finite second moment, as claimed.

Claim 2: From every subsequence of $\{n|R_n - R_n'|\}_{n\in\mathbb{N}}$ we can extract a subsequence for which the Cesàro mean converges to 0 in $L^2(\mathbb{P})$.

From Claim 1 and the Banach-Alaoglu theorem (see Theorem 3.16 in Brezis (2011)) applied on the Hilbert space $L^2(\mathbb{P})$, we see that, along subsequences, $n|R_n - R_n'| \overset{L^2}{\rightharpoonup} 0$, where $\overset{L^2}{\rightharpoonup}$ denotes the weak convergence in the space $L^2(\mathbb{P})$. By a theorem of Banach and Saks (see the Banach–Saks property, exercise 5.24 in Brezis (2011)), we conclude that there exists a sub-sequence, $\{n_k|R_{n_k} - R_{n_k}'|\}_{k\in\mathbb{N}}$, such that the Cesàro means converge strongly to 0 in $L^2(\mathbb{P})$, that is,

$$E\Big(\frac{1}{m}\sum_{k=1}^{m} n_k|R_{n_k} - R_{n_k}'|\Big)^2 \longrightarrow 0. \tag{2.33}$$

Claim 3: From every subsequence of $\sqrt{n}(R_n - ER_n)$ we can extract a further subsequence for which the Cesàro mean converges to 0 in $L^2(\mathbb{P})$.

To check it we consider a subsequence $\{\sqrt{n_k}(R_{n_k} - ER_{n_k})\}_{k\in\mathbb{N}}$. Taking subsequences if necessary we can assume that (2.33) holds.
We set $G_m := \frac{1}{m}\sum_{k=1}^m \sqrt{n_k}R_{n_k}$. By the Efron-Stein inequality

$$\mathrm{Var}(G_m) \leq \frac{1}{2}\sum_{i=1}^{n_m} E(G_m - G_m^{(i)})^2. \tag{2.34}$$

Next, we observe that

$$\begin{aligned} E(G_m - G_m^{(i)})^2 &= E\Big(\frac{1}{m}\sum_{k=1}^m \sqrt{n_k}(R_{n_k} - R_{n_k}^{(i)})\Big)^2 \\ &= \frac{1}{m^2}\sum_{k=1}^m n_k E\Big(R_{n_k} - R_{n_k}^{(i)}\Big)^2 \\ &+ \frac{2}{m^2}\sum_{k=1}^m \sum_{j=k+1}^m \sqrt{n_k}\sqrt{n_j}E(R_{n_k} - R_{n_k}^{(i)})(R_{n_j} - R_{n_j}^{(i)}). \end{aligned}$$

Since for the terms with $n_k < i$ the difference is 0, we have

$$\begin{aligned} E(G_m - G_m^{(i)})^2 &= \frac{1}{m^2}\sum_{n_k\geq i}^{n_m} n_k E\left(R_{n_k} - R_{n_k}^{(i)}\right)^2 \\ &+ \frac{2}{m^2}\sum_{n_k\geq i}^{n_m}\sum_{j=k+1}^m \sqrt{n_k}\sqrt{n_j}E(R_{n_k} - R_{n_k}^{(i)})(R_{n_j} - R_{n_j}^{(i)}) \\ &= \frac{1}{m^2}\sum_{n_k\geq i}^{n_m} n_k E\left(R_{n_k} - R'_{n_k}\right)^2 \\ &+ \frac{2}{m^2}\sum_{n_k\geq i}^{n_m}\sum_{j=k+1}^m \sqrt{n_k}\sqrt{n_j}E(R_{n_k} - R'_{n_k})(R_{n_j} - R'_{n_j}). \end{aligned}$$

Here, the second equality comes from the fact that $\left(R_{n_k} - R'_{n_k}\right)^2$ has the same distribution as $(R_{n_k} - R_{n_k}^{(i)})^2$ when $i \leq n_k$, and the same happens with $(R_{n_k} - R'_{n_k})(R_{n_j} - R'_{n_k})$ and

$(R_{n_k} - R_{n_k}^{(i)})(R_{n_j} - R_{n_k}^{(i)})$. Now turning back to (2.34) we have

$$\begin{aligned}
\text{Var}(G_m) \;\; &\leq \;\; \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{n_m}\sum_{n_k\geq i}^{n_m} n_k E\left(R_{n_k} - R'_{n_k}\right)^2 \\
&\quad + \frac{2}{m^2}\frac{1}{2}\sum_{i=1}^{n_m}\sum_{n_k\geq i}^{n_m}\sum_{j=k+1}^{m} \sqrt{n_k}\sqrt{n_j}E(R_{n_k} - R'_{n_k})(R_{n_j} - R'_{n_j}) \\
&\leq \;\; \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{n_m}\sum_{n_k\geq i}^{n_m} n_k E\left(R_{n_k} - R'_{n_k}\right)^2 \\
&\quad + \frac{2}{m^2}\frac{1}{2}\sum_{i=1}^{n_m}\sum_{n_k\geq i}^{n_m}\sum_{j=k+1}^{m} \sqrt{n_k}\sqrt{n_j}E(|R_{n_k} - R'_{n_k})||(R_{n_j} - R'_{n_j})| \\
&= \;\; \frac{1}{2}\frac{1}{m^2}\sum_{k=1}^{m} n_k^2 E\left(R_{n_k} - R'_{n_k}\right)^2 \\
&\quad + \frac{1}{m^2}\sum_{i=1}^{n_m}\sum_{n_k\geq i}^{n_m}\sum_{j=k+1}^{m} \sqrt{n_k}\sqrt{n_j}E|(R_{n_k} - R'_{n_k})||(R_{n_j} - R'_{n_j})|, \quad (2.35)
\end{aligned}$$

where the last equality comes from

$$\sum_{i=1}^{n_m}\sum_{n_k\geq i}^{n_m} n_k E\left(R_{n_k} - R'_{n_k}\right)^2 = \sum_{k=1}^{m}\sum_{i=1}^{n_k} n_k E\left(R_{n_k} - R'_{n_k}\right)^2 = \sum_{k=1}^{m} n_k^2 E\left(R_{n_k} - R'_{n_k}\right)^2.$$

Compute the last term of (2.35) to obtain

$$\begin{aligned}
&\frac{1}{m^2}\sum_{i=1}^{n_m}\sum_{n_k\geq i}^{n_m}\sum_{j=k+1}^{m} \sqrt{n_k}\sqrt{n_j}E|(R_{n_k} - R'_{n_k})||(R_{n_j} - R'_{n_j})| \\
&\qquad = \;\; \sum_{j=1}^{m}\sum_{k=1}^{j-1}\sum_{i=1}^{n_k} \sqrt{n_k}\sqrt{n_j}E|(R_{n_k} - R'_{n_k})||(R_{n_j} - R'_{n_j})| \\
&\qquad = \;\; \sum_{j=1}^{m}\sum_{k=1}^{j-1} \sqrt{n_k}\sqrt{n_j}E|(R_{n_k} - R'_{n_k})||(R_{n_j} - R'_{n_j})|n_k \\
&\qquad \leq \;\; \sum_{j=1}^{m}\sum_{k=1}^{j-1} n_k n_j E|(R_{n_k} - R'_{n_k})||(R_{n_j} - R'_{n_j})|.
\end{aligned}$$

We conclude that

$$\begin{aligned}
\text{Var}(G_m) &\leq \frac{1}{2}\frac{1}{m^2}\sum_{k=1}^{m} n_k^2 E\left(R_{n_k} - R'_{n_k}\right)^2 + \frac{1}{m^2}\sum_{j=1}^{m}\sum_{k=1}^{j-1} n_k n_j E|(R_{n_k} - R'_{n_k})||(R_{n_j} - R'_{n_j})| \\
&= \frac{1}{2}E\Big(\frac{1}{m}\sum_{k=1}^{m} n_k|R_{n_k} - R'_{n_k}|\Big)^2,
\end{aligned}$$

which, together with (2.33), shows that

$$E\Big(\frac{1}{m}\sum_{k=1}^{m}\sqrt{n_k}(R_{n_k}-ER_{n_k})\Big)^2 = \mathrm{Var}(G_m) \longrightarrow 0. \tag{2.36}$$

Finally we have proven that for every subsequence of $\{G_m\}_{m\in\mathbb{N}}$ we can find a further subsequence converging to 0 strongly in $L^2(\mathbb{P})$, and Claim 3 follows.

To conclude, we note that claim 1 yields

$$E(n(R_n - ER_n)^2) \leq M, \tag{2.37}$$

which implies that the sequence $\sqrt{n}(R_n - ER_n)$ is contained in the centered ball of radius $\sqrt{M}$ in $L^2(P)$, which is compact with respect to the weak topology in $L^2(P)$. In consequence let $A \in L^2(P)$ be such that $\sqrt{n}(R_n - ER_n)$ converges, along a sub-sequence, say $\sqrt{n_k}(R_{n_k} - ER_{n_k})$, weakly in $L^2(P)$ to A. Now it is true that any Cesàro average of $\sqrt{n_k}(R_{n_k} - ER_{n_k})$ converges weakly in $L^2(P)$ to A. Therefore Claim 3 forces $A = 0$. Then $\sqrt{n}(R_n - ER_n)$ converges weakly in $L^2(P)$ to 0, which means that for all $X \in L^2(P)$ we have $\sqrt{n}E((R_n - ER_n)X) \to 0$. With the notation $B_n = \sqrt{n}(R_n - ER_n)$, (2.37) implies that such a sequence is tight, which implies that for each subsequence B_{n_k} there exists a further subsequence converging in distribution to some Y. For the sake of simplicity we use the same notation B_{n_k} for the subsequence. Skorokhod's representation theorem (cf. Theorem 6.7 in Billingsley (1999) eg.) yields that there exists random variables $\tilde{B}_{n_k} \xrightarrow{a.s.} \tilde{Y}$, where $\tilde{B}_{n_k}$ and $\tilde{Y}$ follows the same law of B_{n_k} and Y respectively. Since $E|\tilde{B}_{n_k}|^2 < \sqrt{M}$ and $L^{\frac{3}{2}}(P)$ is continuously embedded in $L^2(P)$ (by Jensen's inequality), we have strong convergence of $\tilde{B}_{n_k}$ to $\tilde{Y}$ in $L^{\frac{3}{2}}(P)$, which implies the convergence of the norm

$$E|B_{n_k}|^{\frac{3}{2}} = E|\tilde{B}_{n_k}|^{\frac{3}{2}} \to E|\tilde{Y}|^{\frac{3}{2}} = E|Y|^{\frac{3}{2}}, \tag{2.38}$$

and the weak convergence in $L^{\frac{3}{2}}(P)$,

$$E(\tilde{B}_{n_k}Z) \to E(\tilde{Y}Z), \text{ for all } Z \in L^3(P). \tag{2.39}$$

Let π_k^* be the Law of the pair $(\tilde{B}_{n_k}, Z)$ (resp. $\pi^* = \mathcal{L}(Y)$) then the previous limit is equivalent to the following;

$$\int xz d\pi_k^*(x,z) \to \int xz d\pi^*(x,z).$$

In consequence (2.39) becomes equivalent to the following condition

$$\int xz d\pi_k(x,z) \to \int xz d\pi(x,z),$$

for all π_k with marginals $\mathcal{L}(\tilde{B}_{n_k}) = \mathcal{L}(B_{n_k})$ and $\mathcal{L}(Z)$, such that $Z \in L^3(P)$, and π with marginals $\mathcal{L}(\tilde{Y}) = \mathcal{L}(Y)$ and $\mathcal{L}(Z)$. This implies in particular that

$$E(B_{n_k}Z) \to E(YZ), \text{ for all } Z \in L^3(P). \tag{2.40}$$

But we know that B_{n_k} converges weakly in $L^2(P)$ to 0, then $EYZ = 0$, for all $Z \in L^3(P)$, which implies $Y = 0$. Moreover (2.38) yields that $E|B_{n_k}|^{\frac{3}{2}} \to 0$, for every subsequence. Then $E|B_n|^{\frac{3}{2}} \to 0$ which implies convergence of $|B_n|$ in probability to 0. The Central Limit Theorem yields the limit

$$\sqrt{n}\left(\int \varphi(\mathbf{x})dP_n(\mathbf{x}) - E\left(\int \varphi(\mathbf{x})dP_n(\mathbf{x})\right)\right) \xrightarrow{w} N(0, \sigma_c^2(P,Q)),$$

which proves the theorem.

□

Remark 2..2. *The key to claim 3 is the fact that for every sub-sequence there exists a further subsequence convergent in Cesàro mean to the same limit. We remark that, while a.s. convergence of a sequence of random variables entails a.s. convergence of its Cesàro means to the same limit, this is no longer true with convergence in distribution, as the following counterexample shows. Let $\{X_n\}$ the sequence defined $X_{2n} = X$ and $X_{2n+1} = -X$ and $X \sim N(0,1)$. Every mean converges to 0, but the sequence X_n converges in distribution to $N(0,1)$. We thank an anonymous reviewer for pointing out this issue in an old version of the work and provide this counterexample.*

PROOF OF THEOREM 2.4.6. We keep the same notations as in the proof of Theorem 2.4.5, noting that the new assumption (2.22) has no influence on the proof of Claim 1. Hence, we only have to prove that $n^2(R_n - R_n')_+^2$ is uniformly integrable and, in fact, recalling that

$$n(R_n - R_n') = n(\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P_n', Q)) - (\varphi(X_1) - \varphi(X_1'))$$

and that $\varphi(X_1)$ has a finite second moment (as shown in the proof of Theorem 2.4.5), it suffices to prove uniform integrability of $n(\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P_n', Q))$.

To check this we denote $Z := \mathcal{T}_c(P_n, Q)$ and $Z' := \mathcal{T}_c(P_n', Q)$. Arguing as in the proof of Lemma 2.4.1 we see that

$$(Z - Z')_+ \leq |X_1 - X_1'| \int_{C_1'} |\nabla h(X_1 - \mathbf{y})| dQ(\mathbf{y}).$$

Hence, by Hölder's inequality, for every pair $(q_1, q_2) \in \alpha$ it holds that

$$\begin{aligned} E(n(Z - Z')_+)^{2+\delta} &\leq E\Big\{|X_1 - X_1'|^{2+\delta}\Big(\int_{C_1'} n|\nabla h(X_1 - \mathbf{y})|dQ(\mathbf{y})\Big)^{2+\delta}\Big\} \\ &\leq \left(E|X_1 - X_1'|^{(2+\delta)q_1}\right)^{\frac{1}{q_1}} \left(E\Big(\int_{C_1'} n|\nabla h(X_1 - \mathbf{y})|dQ(\mathbf{y})\Big)^{(2+\delta)q_2}\right)^{\frac{1}{q_2}}. \end{aligned}$$

A further use of Hölder's inequality yields that

$$\begin{aligned}
&\int_{C_1'} |\nabla h(X_1-\mathbf{y})| dQ(\mathbf{y}) \\
&\quad\le \left(\int_{C_1'} dQ(\mathbf{y})\right)^{\frac{(2+\delta)q_2-1}{(2+\delta)q_2}} \left(\int_{C_1'} |\nabla h(X_1-\mathbf{y})|^{(2+\delta)q_2} dQ(\mathbf{y})\right)^{\frac{1}{(2+\delta)q_2}} \\
&\quad= \frac{1}{n^{\frac{(2+\delta)q_2-1}{(2+\delta)q_2}}} \left(\int_{C_1'} |\nabla h(X_1-\mathbf{y})|^{q_2(2+\delta)} dQ(\mathbf{y})\right)^{\frac{1}{(2+\delta)q_2}}
\end{aligned}$$

Note that $(X_1', \dots, X_n)$ is independent of X_1, hence, the same holds for C_k', for $k = 1, \dots, n$. By exchangeability, we have that $\int_{C_1'} |\nabla h(X_1-\mathbf{y})|^{(2+\delta)q_2} dQ(\mathbf{y})$ is equally distributed as $\int_{C_k'} |\nabla h(X_1-\mathbf{y})|^{(2+\delta)q_2} dQ(\mathbf{y})$, $k = 2, \dots, n$. This implies

$$\begin{aligned}
E\left\{\int_{C_1'} |\nabla h(X_1-\mathbf{y})|^{(2+\delta)q_2} dQ(\mathbf{y})\right\} &= \frac{1}{n} E\left\{\sum_{i=1}^n \int_{C_i'} |\nabla h(X_1-\mathbf{y})|^{(2+\delta)q_2} dQ(\mathbf{y})\right\} \\
&= \frac{1}{n} E\left\{\int_{\mathbb{R}^d} |\nabla h(X_1-\mathbf{y})|^{(2+\delta)q_2} dQ(\mathbf{y})\right\},
\end{aligned}$$

which, in turn, entails

$$E\left(\int_{C_1'} |\nabla h(X_1-\mathbf{y})| dQ(\mathbf{y})\right)^{(2+\delta)q_2} \le \frac{1}{n^{(2+\delta)q_2}} E\left(\int_{\mathbb{R}^d} |\nabla h(X_1-\mathbf{y})|^{(2+\delta)q_2} dQ(\mathbf{y})\right).$$

Combining the last estimates, we can see that

$$E(n(Z-Z')_+)^{(2+\delta)} \le \left(E|X_1-X_1'|^{(2+\delta)q_1}\right)^{\frac{1}{q_1}} \left(E\left(\int_{\mathbb{R}^d} |\nabla h(X_1-\mathbf{y})|^{(2+\delta)q_2} dQ(\mathbf{y})\right)\right)^{\frac{1}{q_2}}$$

and the proof follows.

□

PROOF OF THEOREM 2.4.10. We set

$$R_{n,m} := \mathcal{T}_c(P_n, Q_m) - \int \varphi(\mathbf{x}) dP_n(\mathbf{x}) - \int \psi(\mathbf{y}) dQ_m(\mathbf{y})$$

with φ an optimal transport potential from P to Q for the cost c and $\psi = \varphi^c$ and observe that it suffices to show that $\frac{nm}{n+m}\mathrm{Var}(R_{n,m}) \to 0$. Once again the key of the proof is Efron-Stein's inequality. Note that $R_{n,m}$ as a function of $X_1, \dots, X_n, Y_1, \dots, Y_m$ is symmetric in its n first variables as well as in the last m. Let X_1'(resp. Y_1') be a copy of X_1 (resp. Y_1) both independent of $X_1, \dots, X_n, Y_1, \dots, Y_m$, finally let P_n' (resp. Q_n') be the empirical distribution of $X_1', X_2 \dots, X_n$ (resp. $Y_1', Y_2 \dots, Y_m$). Hence, if we denote

$$\begin{aligned}
R_{n,m}' &:= \mathcal{T}_c(P_n', Q_m) - \int \varphi(\mathbf{x}) dP_n'(\mathbf{x}) - \int \psi(\mathbf{y}) dQ_m(\mathbf{y}), \\
R_{n,m}'' &:= \mathcal{T}_c(P_n, Q_m') - \int \varphi(\mathbf{x}) dP_n(\mathbf{x}) - \int \psi(\mathbf{y}) dQ_m'(\mathbf{y}),
\end{aligned}$$

by the Efron-Stein inequality we have

$$\frac{nm}{n+m}\mathrm{Var}(R_{n,m}) \le \frac{n^2 m}{n+m}E(R_{n,m}-R'_{n,m})_+^2 + \frac{nm^2}{n+m}E(R_{n,m}-R''_{n,m})_+^2.$$

Now, to conclude, it suffices to prove that

$$n^2 E((R_{n,m}-R'_{n,m})_+^2) \to 0 \quad \text{and} \tag{2.41}$$

$$m^2 E((R_{n,m}-R''_{n,m})_+^2) \to 0. \tag{2.42}$$

We handle (2.41), which will follow if we prove that $n(R_{n,m}-R'_{n,m})_+ \to 0$ a.s. and also that $n^2(R_{n,m}-R'_{n,m})^2$ is uniformly integrable. (2.42) follows by an analogous argument. For the first claim note note that if φ_n (resp. ψ_n) is an optimal transport potential from P_n to Q_m (resp. from Q_m to P_n) then

$$R'_{n,m} \ge \int \varphi_n(\mathbf{x})dP'_n(\mathbf{x}) + \int \psi_m(\mathbf{y})dQ_m(\mathbf{y}) - \int \varphi(\mathbf{x})dP'_n(\mathbf{x}) - \int \psi(\mathbf{y})dQ_m(\mathbf{y}).$$

As a consequence,

$$\begin{aligned} R_{n,m}-R'_{n,m} &\le \int_{\mathbb{R}^d} (\varphi_n(\mathbf{x})-\varphi(\mathbf{x}))\left(dP_n(\mathbf{x})-dP'_n(\mathbf{x})\right) \\ &= \frac{1}{n}\left(\varphi_n(X_1)-\varphi(X_1)-\varphi_n(X'_1)+\varphi(X'_1)\right) \end{aligned}$$

and we see that

$$n\left(R_{n,m}-R'_{n,m}\right)_+ \le |\varphi_n(X_1)-\varphi(X_1)-\varphi_n(X'_1)+\varphi(X'_1)|. \tag{2.43}$$

By Theorem 2.3.4, with a right choice of potentials we can guarantee that, $P-$a.s., $\varphi_n \to \varphi$ and conclude that $n\left(R_{n,m}-R'_{n,m}\right)_+ \to 0$ $P-$a.s.

Finally, it only remains to prove that $n^2 E\left(R_{n,m}-R'_{n,m}\right)_+^2$ is uniformly bounded, which follows arguing as in the proof of Theorem 2.4.6.

□

2..2 Proofs of Lemmas

PROOF OF LEMMA 2.3.1. Set $\mathbf{x}_0 \in \mathrm{dom}(\nabla^c\psi) \cap \mathrm{Supp}(Q)$ and $\mathbf{y}_0 = \nabla^c\psi(\mathbf{x}_0)$. By Lemma 2.2.10 we see that for each $\epsilon > 0$ there exists some $\delta > 0$ such that if $|\mathbf{z}-\mathbf{x}_0| < \delta$ then $\partial^c\psi(\mathbf{z}) \subset B(\mathbf{y}_0,\epsilon)$. Let π be the unique optimal transport plan between Q and P. By Theorem 2.2.4 $\mathrm{Supp}(\pi) \subset \partial^c\psi$. This entails

$$\pi\left(B(\mathbf{x}_0,\delta)\times B(\mathbf{y}_0,\epsilon)\right) = \pi\left(B(\mathbf{x}_0,\delta)\times\mathbb{R}^d\right) = Q(B(\mathbf{x}_0,\delta)) = \eta > 0,$$

where the strict inequality comes from fact $\mathbf{x}_0 \in \mathrm{Supp}(Q)$. Repeating the argument with a decreasing sequence $\epsilon_k \to 0$, we obtain a sequence $\delta_k \le \frac{1}{k}$ such that

$$\begin{aligned} \pi\left(B(\mathbf{x}_0,\delta_k)\times B(\mathbf{y}_0,\epsilon_k)\right) &= \pi\left(B(\mathbf{x}_0,\delta_k)\times\mathbb{R}^d\right) \\ &= Q(B(\mathbf{x}_0,\delta_k)) = \eta_k > 0. \end{aligned}$$

Let π_n be an optimal transport plan between P_n and Q_n. We observe that

(a) $\pi_n \xrightarrow{w} \pi$ along subsequences by Theorem 5.20 in Villani (2008),

(b) $\mathrm{Supp}(\pi_n) \subset \partial^c \psi_n$ by Theorem 2.2.4.

By (a) there exists N_k such that, for $n \geq N_k$, $\pi_n(B(\mathbf{x}_0, \delta_k) \times B(\mathbf{y}_0, \epsilon_k)) \geq \eta_k/2$. Hence, by (b) we can choose a pair

$$(\mathbf{x}_{n_k}, \mathbf{y}_{n_k}) \in \partial^c \psi_n \cap (B(\mathbf{x}_0, \delta_k) \times B(\mathbf{y}_0, \epsilon_k)). \tag{2.44}$$

As a consequence of (2.44), since $\epsilon_k, \delta_k \to 0$, we can extract a sub-sequence of $(\mathbf{x}_n, \mathbf{y}_n) \in \partial^c \psi_n$ converging to $(\mathbf{x}_0, \mathbf{y}_0)$. Define $a_n := \psi_n(\mathbf{x}_n) - \psi(\mathbf{x}_0)$ and $\tilde{\psi}_n := \psi_n - a_n$ (which has the same c-superdifferential as ψ_n). Now, (2.44) implies that $\partial^c \tilde{\psi}_n$ are c-cyclically monotone sets which do not escape to the horizon. By Theorem 2.2.8 and Lemma 2.2.9 we deduce that $\partial^c \tilde{\psi}_n$ converges to a c-cyclically monotone set Γ along a sub-sequence. Necessarily $\Gamma \subset \partial^c f$ for some c-concave function f. We observe that $(\mathbf{x}_0, \mathbf{y}_0) \in \partial^c f$. If we take another arbitrary point $\mathbf{x} \in \mathrm{dom}(\nabla^c \psi)$ and $\partial^c \psi(\mathbf{x}) = \{\mathbf{y}\}$, we can apply the same arguments to check that $(\mathbf{x}, \mathbf{y}) \in \partial^c f$. Hence, $\mathrm{dom}(\nabla^c \psi) \subset \mathrm{dom}(f)$. Since, in view of Lemma 2.2.3, f is differentiable a.s then $\partial^c f$ is a singleton a.s and, therefore, that $\nabla f = \nabla \psi$ a.s. in the support of Q, which is connected. Using Theorem 2.2.6 we conclude that there exists a constant C such that $\psi = f - C$ in Ω. Hence $\partial^c \psi = \partial^c f$ and the result follows. □

PROOF OF LEMMA 2.3.2. We can assume, without loss of generality, that $\mathbf{p}$ is in the interior of the domain of ψ, since otherwise the result is trivial. With this assumption, we check first that we cannot have $\psi_n(\mathbf{p}_n) \to \infty$. In fact, in that case, by c-concavity we would have

$$\psi_n(\mathbf{p}_n) \leq c(\mathbf{p}_n, \mathbf{y}) - \psi_n^c(\mathbf{y})$$

for all $\mathbf{y}$. Hence, we would have $\psi_n^c(\mathbf{y_n}) \to -\infty$ for all $\mathbf{y}_n \to \mathbf{y}$. Now, take $\mathbf{p}_0$ as in (2.12). By Lemma 2.3.1 we can choose $(\tilde{\mathbf{p}}_n, \mathbf{y}_n)$ with $\mathbf{y}_n \in \partial^c \psi_n(\tilde{\mathbf{p}}_n)$, $\tilde{\mathbf{p}}_n \to \mathbf{p}_0$ and $\mathbf{y}_n \to \nabla^c \psi(\mathbf{p}_0) = \mathbf{y}_0$. The same convergence in terms of Painlevé-Kuratowski would state the existence of a subsequence, where we keep the same indexing, $\{\mathbf{y}_n^0\}_{n \in \mathbb{N}}$, with $\mathbf{y}_n^0 \in \partial^c \psi_n(\mathbf{p}_0)$, such that $\mathbf{y}_n^0 \to \nabla^c \psi(\mathbf{p}_0) = \mathbf{y}_0$. From the definition of c-sub-differential;

$$\psi_n(\tilde{\mathbf{p}}_n) \leq \psi_n(\mathbf{p}_0) + [c(\tilde{\mathbf{p}}_n, \mathbf{y}_n^0) - c(\mathbf{p}_0, \mathbf{y}_n^0)] = [c(\tilde{\mathbf{p}}_n, \mathbf{y}_n^0) - c(\mathbf{p}_0, \mathbf{y}_n^0)] \to 0,$$

and the limit

$$\psi_n(\tilde{\mathbf{p}}_n) = c(\tilde{\mathbf{p}}_n, \mathbf{y}_n) - \psi_n^c(\mathbf{y}_n) \to \infty,$$

we would obtain a contradiction.

Now, we can assume, taking subsequences if necessary, that $\psi_n(\mathbf{p}_n) < -n$ for all $n \in \mathbb{N}$. Now, taking $\mathbf{y}_n \in \partial^c \psi(\mathbf{p}_n)$—which is non-empty because $\mathbf{p}_n$ is in the interior of the domain of ψ, for n big enough— we have that

$$\psi_n(\mathbf{x}) \leq c(\mathbf{x}, \mathbf{y}_n) + \lambda_n, \quad \text{for all } \mathbf{x} \in \mathbb{R}^d, \tag{2.45}$$

where $\lambda_n = \psi_n(\mathbf{p}_n) - c(\mathbf{p}_n, \mathbf{y}_n)$. Hence, by assumption we have that $c(\mathbf{p}_n, \mathbf{y}_n) + \lambda_n \leq -n$ for all $n \in \mathbb{N}$. Now, let $\{\mathbf{x}_n\}$ be a bounded sequence such that $\psi_n(\mathbf{x}_n)$ is bounded. Then

$$\psi_n(\mathbf{x}_n) < c(\mathbf{x}_n, \mathbf{y}_n) - c(\mathbf{p}_n, \mathbf{y}_n) - n.$$

Since $\psi_n(\mathbf{x}_n), \mathbf{p}_n, \mathbf{x}_n$ are bounded, then $|\mathbf{y}_n| \to \infty$. For each n we choose the height $r_n \in [0, \infty]$ and the direction $\mathbf{z}_n$ of the largest cone with vertex $\mathbf{p}_n - \mathbf{y}_n$ such that

$$K\left(r_n, \tfrac{\pi}{1+r_n^{-1}}, \mathbf{z}_n, \mathbf{p}_n - \mathbf{y}_n\right) \subset \{\mathbf{x} : \ h(\mathbf{x}) \leq h(\mathbf{p}_n - \mathbf{y}_n)\}\,.$$

Since $\mathbf{z}_n \in \mathbb{S}_{d-1}$, then up to a sub-sequence, we can assume that $\mathbf{z}_n \to \mathbf{z} \in \mathbb{S}_{d-1}$. Also, since $|\mathbf{p}_n - \mathbf{y}_n| \to \infty$, condition (A2) implies that $r_n \to \infty$ (note that otherwise if $|r_n| < R$ then (A2) is no longer true for $r = R + 1$ and $\theta = \frac{\pi}{1+r_n^{-1}}$).

Now let $\{\mathbf{x}_n\}_{n\in\mathbb{N}} \subset\subset \{\mathbf{x} : \langle \mathbf{z}, \mathbf{x} - \mathbf{p}\rangle > 0\}$ be a bounded sequence. From the fact that $r_n \to \infty$ we see that

$$\cos\left(\tfrac{1}{2}\tfrac{\pi}{1+r_n^{-1}}\right) \to 0.$$

Therefore, for big enough n

$$|\mathbf{x}_n - \mathbf{p}_n| \cos\left(\tfrac{1}{2}\tfrac{\pi}{1+r_n^{-1}}\right) < \langle \mathbf{z}, \mathbf{x} - \mathbf{p}\rangle < r_n. \tag{2.46}$$

As a consequence $\mathbf{x}_n \in K(r_n, \frac{\pi}{1+r_n^{-1}}, \mathbf{z}, \mathbf{p}_n)$, which implies that

$$\mathbf{x}_n - \mathbf{y}_n \in K\left(r_n, \tfrac{\pi}{1+r_n^{-1}}, \mathbf{z}, \mathbf{p}_n - \mathbf{y}_n\right) \subset \{\mathbf{x} : \ h(\mathbf{x}) \leq h(\mathbf{p}_n - \mathbf{y}_n)\}\,.$$

From this we conclude that $c(\mathbf{x}_n, \mathbf{y}_n) \leq c(\mathbf{p}_n, \mathbf{y}_n)$, and turning back to (2.45), that

$$\psi_n(\mathbf{x}_n) \leq c(\mathbf{x}_n, \mathbf{y}_n) + \lambda_n \leq c(\mathbf{p}_n, \mathbf{y}_n) + \lambda_n \leq -n,$$

and the proof follows. □

PROOF OF LEMMA 2.3.3. We split the proof into the following steps:

Step 1 (Pointwise boundedness): Fix $\mathbf{x} \in \mathrm{Supp}(Q) \cap \mathrm{dom}(\psi)$. By Lemma 2.3.1 there exists a c-cyclically monotone set Γ such that, up to taking sub-sequences, $\partial^c\psi_n \to \Gamma$ in the sense of Painlevé-Kuratowski. Hence, there exists a sequence $(\mathbf{x}_{n_k}, \mathbf{y}_{n_k}) \in \partial^c\psi_{n_k}$ satisfying

$$(\mathbf{x}_{n_k}, \mathbf{y}_{n_k}) \to (\mathbf{x}, \mathbf{y}) \in \Gamma.$$

Assume $\{\psi_n(\mathbf{x}_{n_k})\}_{k\in\mathbb{N}}$ is not bounded. Then there exist a sub-sequence $\psi_{n_{k_m}}(\mathbf{x}_{n_{k_m}}) \to -\infty$ (the case $\psi_{n_{k_m}}(\mathbf{x}_{n_{k_m}}) \to \infty$ can be excluded arguing as at the beginning of the proof of Lemma 2.3.2. Now, we take $\mathbf{p}_0$ as in (2.12) and observe that,

$$0 \leq \psi_{n_{k_m}}(\mathbf{x}_{n_{k_m}}) + c(\mathbf{p}_0, \mathbf{y}_{n_{k_m}}) - c(\mathbf{x}_{k_m}, \mathbf{y}_{n_{k_m}}). \tag{2.47}$$

Taking limits as $m \to \infty$ in (2.47) leads to a contradiction. Hence, the sequence $\{\psi_{n_k}(\mathbf{x}_{n_k})\}_{k\in\mathbb{N}}$ must be bounded.

For ease of reading we will use the same notation for the subsequence $\{\psi_{n_k}\}_{k\in\mathbb{N}}$ and the main sequence $\{\psi_n\}_{n\in\mathbb{N}}$ in the subsequent steps 2 and 3.

Step 2 (For every compact $K \subset \text{Supp}(Q)$ there exists $M > 0$ such that $|\psi_n(K)| \leq M$ for large enough n): Assume, on the contrary, that for every $m \in \mathbb{N}$ there exists some $n_m \in \mathbb{N}$ such that $\mathbf{k}_{n_m} \in K$ and $|\psi_{n_m}(\mathbf{k}_{n_m})| > m$. Then $|\psi_{n_m}(\mathbf{k}_{n_m})| \to \infty$ as $m \to \infty$ and, by compactness, $\mathbf{k}_{n_m} \to \mathbf{k} \in K$ along a subsequence. By Lemma 2.3.2 we see that there exists $\mathbf{z} \in \mathbb{R}^d$ such that $\psi_n(\mathbf{x}_n)$ is not bounded, for every bounded sequence $\{\mathbf{x}_n\} \subset\subset \{\mathbf{x} : \langle \mathbf{z}, \mathbf{x} - \mathbf{k}\rangle > 0\}$. Now take $\mathbf{x}_0 \in \text{Supp}(Q) \cap \{\mathbf{x} : \langle \mathbf{z}, \mathbf{x} - \mathbf{k}\rangle > 0\}$. Since this last set is open, there exists $\varepsilon > 0$ such that $B(\mathbf{x}_0, \epsilon) \subset\subset \text{Supp}(Q) \cap \{\mathbf{x} : \langle \mathbf{z}, \mathbf{x} - \mathbf{k}\rangle > 0\}$, and this contradicts Step 1 applied to the point $\mathbf{x}_0$.

Step 3 (For every compact $K \subset \text{Supp}(Q)$ there exists $M > 0$ such that $\partial^c\psi_n(K) \subset B(\mathbf{0}, M)$ for large enough n): Assume this fails for a compact $K \subset \text{Supp}(Q)$. Since $\text{Supp}(Q)$ is open, there exists $\epsilon > 0$ such that

$$K^\epsilon := \{\mathbf{x} : d(\mathbf{x}, K) \leq \epsilon\} \subset\subset \text{Supp}(Q).$$

By Step 2 there exists $M > 0$ and $n_0 \in \mathbb{N}$ such that $|\psi_n(\mathbf{k})| \leq M$, for all $\mathbf{k} \in K^\epsilon$ and $n \geq n_0$. Now we can take $\{\mathbf{k}_n\}_{n\in\mathbb{N}} \subset K$ and $\mathbf{y}_n \in \partial^c\psi_n(\mathbf{k}_n)$ such that $|\mathbf{y}_n| \to \infty$, define $\mathbf{v}_n := \mathbf{k}_n - \mathbf{y}_n$ and observe that for n big enough $|\mathbf{v}_n| > 1$. Define $\xi_n := 1 - \epsilon\frac{1}{|\mathbf{v}_n|}$ and note that $\xi_n \to 1$. All $\mathbf{k}_n$ belong to the compact set K, hence define $\mathbf{z}_n := \mathbf{k}_n + (\xi_n - 1)\mathbf{v}_n = \mathbf{k}_n + \frac{\epsilon}{2|\mathbf{v}_n|}\mathbf{v}_n \in K^\epsilon$, for which we can ensure $\psi_n(\mathbf{z}_n) > -M$. By definition of superdifferentials we have

$$2M \geq \psi_n(\mathbf{k}_n) - \psi_n(\mathbf{z}_n) \geq h(\mathbf{v}_n) - h(\xi_n\mathbf{v}_n)$$

and by convexity of h there exists $\mathbf{s}_n \in \partial h(\xi_n\mathbf{v}_n)$, for which we have

$$2M \geq \langle (1 - \xi_n)\mathbf{v}_n, \mathbf{s}_n\rangle = \epsilon\left\langle \frac{\mathbf{v}_n}{|\mathbf{v}_n|}, \mathbf{s}_n\right\rangle. \tag{2.48}$$

Observe that we also have

$$h(\mathbf{0}) \geq h(\xi_n\mathbf{v}_n) + \langle \mathbf{0} - \xi_n\mathbf{v}_n, \mathbf{s}_n\rangle.$$

Now, since $\xi_n > 1 - \epsilon > 0$ and $|\mathbf{v}_n| \to \infty$ we have $|\xi_n\mathbf{v}_n| \to \infty$ and, consequently,

$$\liminf_{n\to\infty} \left\langle \frac{\mathbf{v}_n}{|\mathbf{v}_n|}, \mathbf{s}_n\right\rangle \geq \liminf_{n\to\infty} \frac{h(\xi_n\mathbf{v})}{|\xi_n\mathbf{v}_n|} \to \infty, \tag{2.49}$$

with the last limit following from the condition (A3). This contradicts (2.48).

□

PROOF OF LEMMA 2.4.1. We write X_1' for a random variable with law P and independent from $(X_1, \ldots, X_n)$. Denote by P_n' the empirical measure associated to $(X_1', X_2, \ldots, X_n)$,

$Z := \mathcal{T}_c(P_n, Q)$ and $Z' := \mathcal{T}_c(P_n', Q)$. Since $Q \ll \ell_d$, there exists an optimal transport map from Q to P_n', which we denote by T. We set

$$C_1' := \{\mathbf{y} \in \mathbb{R}^d : T(\mathbf{y}) = X_1'\}, \quad C_i' := \{\mathbf{y} \in \mathbb{R}^d : T(\mathbf{y}) = X_i\},\ i \geq 2,$$

and observe that $Q(C_i') = \frac{1}{n}$ and

$$Z' = \int c(\mathbf{x}, \mathbf{y}) d\pi'(\mathbf{x}, \mathbf{y}) = \int_{C_1'} c(X_1', \mathbf{y}) dQ(\mathbf{y}) + \sum_{i=2}^{n} \int_{C_i'} c(X_i, \mathbf{y}) dQ(\mathbf{y}),$$
$$Z \leq \int_{C_1'} c(X_1, \mathbf{y}) dQ(\mathbf{y}) + \sum_{i=2}^{n} \int_{C_i'} c(X_i, \mathbf{y}) dQ(\mathbf{y}).$$

From this we see that (recall that $h(X - \cdot)$ is convex and Q-a.s. differentiable)

$$\begin{aligned} Z - Z' &\leq \int_{C_1'} \big(c(X_1, \mathbf{y}) - c(X_1', \mathbf{y})\big)\, dQ(\mathbf{y}) \\ &\leq \int_{C_1'} \big\langle \nabla h(X_1 - \mathbf{y}), X_1 - X_1' \big\rangle\, dQ(\mathbf{y}) \\ &\leq |X_1 - X_1'| \int_{C_1'} |\nabla h(X_1 - \mathbf{y})| dQ(\mathbf{y}). \end{aligned}$$

Hence, by Hölder's inequality, for any pair $(q_1, q_2) \in \alpha$,

$$\begin{aligned} E(Z - Z')_+^2 &\leq E\Big\{|X_1 - X_1'|^2 \Big(\int_{C_1'} |\nabla h(X_1 - \mathbf{y})| dQ(\mathbf{y})\Big)^2\Big\} \\ &\leq \Big(E|X_1 - X_1'|^{2q_1}\Big)^{\frac{1}{q_1}} \Big(E\Big(\int_{C_1'} |\nabla h(X_1 - \mathbf{y})| dQ(\mathbf{y})\Big)^{2q_2}\Big)^{\frac{1}{q_2}}. \end{aligned} \tag{2.50}$$

Using again Hölder's inequality we get that

$$\begin{aligned} \int_{C_1'} |\nabla h(X_1 - \mathbf{y})| dQ(\mathbf{y}) &\leq \Big(\int_{C_1'} dQ(\mathbf{y})\Big)^{\frac{2q_2 - 1}{2q_2}} \Big(\int_{C_1'} |\nabla h(X_1 - \mathbf{y})|^{2q_2} dQ(\mathbf{y})\Big)^{\frac{1}{2q_2}} \\ &= \frac{1}{n^{\frac{2q_2 - 1}{2q_2}}} \Big(\int_{C_1'} |\nabla h(X_1 - \mathbf{y})|^{2q_2} dQ(\mathbf{y})\Big)^{\frac{1}{2q_2}}. \end{aligned}$$

Finally, by exchangeability,

$$\begin{aligned} E\Big\{\int_{C_1'} |\nabla h(X_1 - \mathbf{y})|^{2q_2} dQ(\mathbf{y})\Big\} &= \frac{1}{n} E\Big\{\sum_{i=1}^{n} \int_{C_i'} |\nabla h(X_1 - \mathbf{y})|^{2q_2} dQ(\mathbf{y})\Big\} \\ &= \frac{1}{n} E\Big\{\int_{\mathbb{R}^d} |\nabla h(X_1 - \mathbf{y})|^{2q_2} dQ(\mathbf{y})\Big\}, \end{aligned}$$

which implies that

$$E\Big(\int_{C_1'} |\nabla h(X_1 - \mathbf{y})| dQ(\mathbf{y})\Big)^{2q_2} \le \frac{1}{n^{2q_2}} E\Big(\int_{\mathbb{R}^d} |\nabla h(X_1 - \mathbf{y})|^{2q_2} dQ(\mathbf{y})\Big).$$

Combining the last estimates with (2.50) leads to

$$E(Z - Z')_+^2 \le \frac{1}{n^2}\Big(E|X_1 - X_1'|^{2q_1}\Big)^{\frac{1}{q_1}} \Big(E\Big(\int_{\mathbb{R}^d} |\nabla h(X_1 - \mathbf{y})|^{2q_2} dQ(\mathbf{y})\Big)\Big)^{\frac{1}{q_2}}.$$

□

Bibliography

Gabriel Acosta and Ricardo Duran. An optimal poincaré inequality in l1 for convex domains. *Proceedings of the American Mathematical Society*, 132:195–202, 01 2004. doi: 10.2307/1193994.

M. Ajtai, Janos Komlos, and G. Tusnády. On optimal matchings. *Combinatorica*, 4:259–264, 12 1984. doi: 10.1007/BF02579135.

Luigi Ambrosio, Federico Stra, and Dario Trevisan. A pde approach to a 2-dimensional matching problem. *Probability Theory and Related Fields*, 173, 02 2019. doi: 10.1007/s00440-018-0837-x.

François Bachoc, Fabrice Gamboa, Jean-Michel Loubes, and Nil Venet. A gaussian process regression model for distribution inputs. *IEEE Transactions on Information Theory*, 64 (10):6620–6637, 2018. doi: 10.1109/TIT.2017.2762322.

Eustasio Del Barrio, Evarist Giné, and Frederic Utzet. Asymptotics for L2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli*, 11(1):131 – 189, 2005. doi: 10.3150/bj/1110228245. URL https://doi.org/10.3150/bj/1110228245.

Philippe Berthet, Jean-Claude Fort, and Thierry Klein. A Central Limit Theorem for Wasserstein type distances between two distinct univariate distributions. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 56(2):954 – 982, 2020. doi: 10.1214/19-AIHP990. URL https://doi.org/10.1214/19-AIHP990.

Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. ISBN 0-471-19745-9. A Wiley-Interscience Publication.

Emily Black, Samuel Yeom, and Matt Fredrikson. Fliptest: Fairness testing via optimal transport. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, page 111–121, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450369367. doi: 10.1145/3351095.3372845. URL https://doi.org/10.1145/3351095.3372845.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 02 2013. ISBN 9780199535255. doi: 10.1093/acprof:oso/9780199535255.001.0001. URL https://doi.org/10.1093/acprof:oso/9780199535255.001.0001.

Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer New York, 2011. doi: 10.1007/978-0-387-70914-7. URL https://doi.org/10.10072F978-0-387-70914-7.

Nicolas Courty, Rémi Flamary, and Mélanie Ducoffe. Learning wasserstein embeddings. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=SJyEH91A-.

Juan Antonio Cuesta-Albertos, Carlos Matrán, and A. Tuero-Díaz. Optimal transportation plans and convergence in distribution. *Journal of Multivariate Analysis*, 60:72–83, 1997.

Eustasio del Barrio and Jean-Michel Loubes. Central limit theorems for empirical transportation cost in general dimension. *The Annals of Probability*, 47, 05 2017. doi: 10.1214/18-AOP1275.

Eustasio del Barrio, Evarist Giné, and Carlos Matrán. Central Limit Theorems for the Wasserstein Distance Between the Empirical and the True Distributions. *The Annals of Probability*, 27(2):1009 – 1071, 1999. doi: 10.1214/aop/1022677394. URL https://doi.org/10.1214/aop/1022677394.

Eustasio del Barrio, Paula Gordaliza, and Jean-Michel Loubes. A central limit theorem for lp transportation cost on the real line with application to fairness assessment in machine learning. *Information and Inference: A Journal of the IMA*, 8, 12 2019. doi: 10.1093/imaiai/iaz016.

Eustasio del Barrio, Alberto González-Sanz, and Jean-Michel Loubes. Central limit theorems for semi-discrete wasserstein distances. *Pre-Print*, 2021a.

Eustasio del Barrio, Alberto González-Sanz, and Jean-Michel Loubes. Central limit theorems for general transportation costs, 2021b. URL https://arxiv.org/abs/2102.06379.

Eustasio del Barrio, Alberto González-Sanz, Jean-Michel Loubes, and Jonathan Niles-Weed. An improved central limit theorem and fast convergence rates for entropic transportation costs. *ArXiv*, abs/2204.09105, 2022.

Nicolas Fournier and Arnaud Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162, 12 2013. doi: 10.1007/s00440-014-0583-7.

Wilfrid Gangbo and Robert J. McCann. The geometry of optimal transportation. *Acta Mathematica*, 177(2):113 – 161, 1996. doi: 10.1007/BF02392620. URL https://doi.org/10.1007/BF02392620.

Javier González-Delgado, Alberto González-Sanz, Juan Cortés, and Pierre Neuvial. Two-sample goodness-of-fit tests on the flat torus based on wasserstein distance and their relevance to structural biology. 2021.

Paula Gordaliza, Eustasio Del Barrio, Gamboa Fabrice, and Jean-Michel Loubes. Obtaining fairness using optimal transport theory. In *International Conference on Machine Learning*, pages 2357–2365. PMLR, 2019.

Shayan Hundrieser, Marcel Klatt, Thomas Staudt, and Axel Munk. A unifying approach to distributional limits for empirical optimal transport. *Pre-Print*, 02 2022a. URL https://arxiv.org/pdf/2202.12790.pdf.

Shayan Hundrieser, Thomas Staudt, and Axel Munk. Empirical optimal transport between different measures adapts to lower complexity. 2022b.

M. Ledoux. On optimal matching of gaussian samples. *Journal of Mathematical Sciences (United States)*, 238:495–522, 04 2019. doi: 10.1007/s10958-019-04253-6.

Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. Number 1 in Graduate studies in mathematics. Springer Berlin, Heidelberg, 1991.

Tudor Manole and Jonathan Niles-Weed. Sharp convergence rates for empirical optimal transport with smooth costs. *Pre-Print*, 2021.

Tudor Manole, Sivaraman Balakrishnan, Jonathan Niles-Weed, and Larry Wasserman. Plugin estimation of smooth optimal transport maps. *Pre-Print*, 07 2021. URL https://arxiv.org/pdf/2107.12364.pdf.

Gonzalo Mena and Jonathan Niles-Weed. Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/5acdc9ca5d99ae66afdfe1eea0e3b26b-Paper.pdf.

Gabriel Peyré and Marco Cuturi. Editorial ima iai - information and inference special issue on optimal transport in data sciences. *Information and Inference: A Journal of the IMA*, 8: 655–656, 2019.

Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019. ISSN 1935-8237. doi: 10.1561/2200000073. URL http://dx.doi.org/10.1561/2200000073.

Roger J. B. Wets R. Tyrrell Rockafellar. *Variational Analysis*. Springer, Berlin, Heidelberg, 1998. ISBN 978-3-642-02431-3.

Laurent Risser, Alberto González Sanz, Quentin Vincenot, and Jean-Michel Loubes. Tackling algorithmic bias in neural-network classifiers using wasserstein-2 regularization. *J. Math Imaging Vis*, 2021.

R. T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific Journal of Mathematics*, 17(3):497 – 510, 1966. doi: pjm/1102994514. URL https://doi.org/.

R. Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series. Princeton University Press, Princeton, N. J., 1970.

Ludger Rüschendorf. Optimal solutions of multivariate coupling problems. *Applicationes Mathematicae*, 23(3):325–338, 1995. URL http://eudml.org/doc/219135.

Ludger Rüschendorf. On c-optimal random variables. *Statistics and Probability Letters*, 27(3):267–270, 1996. ISSN 0167-7152. doi: https://doi.org/10.1016/0167-7152(95)00078-X. URL https://www.sciencedirect.com/science/article/pii/016771529500078X.

Filippo Santambrogio. Optimal transport for applied mathematicians. calculus of variations, pdes and modeling. 2015. URL https://www.math.u-psud.fr/~filippo/OTAM-cvgmt.pdf.

Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, Lia Lee, Jenny Chen, Justin Brumbaugh, Philippe Rigollet, Konrad Hochedlinger, Rudolf Jaenisch, Aviv Regev, and Eric S. Lander. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943.e22, 2019. ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2019.01.006. URL https://www.sciencedirect.com/science/article/pii/S009286741930039X.

Cyril Smith and Martin Knott. On hoeffding-fréchet bounds and cyclic monotone relations. *Journal of Multivariate Analysis*, 40(2):328–334, 1992. ISSN 0047-259X. doi: https://doi.org/10.1016/0047-259X(92)90029-F. URL https://www.sciencedirect.com/science/article/pii/0047259X9290029F.

Max Sommerfeld and Axel Munk. Inference for empirical wasserstein distances on finite spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):219–238, 2018.

Thomas Staudt, Shayan Hundrieser, and Axel Munk. On the uniqueness of kantorovich potentials. 2022.

M. Talagrand. The Transportation Cost from the Uniform Measure to the Empirical Measure in Dimension ≥ 3. *The Annals of Probability*, 22(2):919 – 959, 1994. doi: 10.1214/aop/1176988735. URL https://doi.org/10.1214/aop/1176988735.

M. Talagrand and J. E. Yukich. The Integrability of the Square Exponential Transportation Cost. *The Annals of Applied Probability*, 3(4):1100 – 1111, 1993. doi: 10.1214/aoap/1177005274. URL https://doi.org/10.1214/aoap/1177005274.

Michel Talagrand. Matching Random Samples in Many Dimensions. *The Annals of Applied Probability*, 2(4):846 – 856, 1992. doi: 10.1214/aoap/1177005578. URL https://doi.org/10.1214/aoap/1177005578.

Michel Talagrand. Scaling and non-standard matching theorems. *Comptes Rendus Mathematique*, 356(6):692–695, 2018. ISSN 1631-073X. doi: https://doi.org/10.1016/j.crma.2018.04.018. URL https://www.sciencedirect.com/science/article/pii/S1631073X18301328.

Carla Tameling, Max Sommerfeld, and Axel Munk. Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications. *Ann. Appl. Probab.*, 29 (5):2744–2781, 10 2019. doi: 10.1214/19-AAP1463. URL https://doi.org/10.1214/19-AAP1463.

Cédric Villani. *Optimal Transport: Old and New*. Number 338 in Grundlehren der mathematischen Wissenschaften. Springer, Berlin, 2008. ISBN 978-3-540-71049-3. OCLC: ocn244421231.

CHAPTER 3

Central Limit Theorems for Semidiscrete Wasserstein Distances

The content of this chapter is fully available in del Barrio et al. (2022).

Contents

We prove a Central Limit Theorem for the empirical optimal transport cost,

$$\sqrt{\frac{nm}{n+m}}\{\mathcal{T}_c(P_n, Q_m) - \mathcal{T}_c(P, Q)\},$$

in the semi discrete case, i.e when the distribution P is supported in N points, but without assumptions on Q. We show that the asymptotic distribution is the sup of a centered Gaussian process, which is Gaussian under some additional conditions on the probability Q and on the cost. Such results imply the central limit theorem for the p-Wassertein distance, for $p \geq 1$. This means that, for fixed N, the curse of dimensionality is avoided. To better understand the influence of such N, we provide bounds of $E|\mathcal{W}_p^p(P, Q_m) - \mathcal{W}_p^p(P, Q)|$ depending on m and N. Finally, the semi-discrete framework provides a control on the second derivative of the dual formulation, which yields the first central limit theorem for the optimal transport potentials and Laguerre cells. The results are supported by simulations that help to visualize the given limits and bounds. We analyse also the cases where classical bootstrap works.

3.1 Central Limit Theorems for semidiscrete distributions

3.1.1 Semidiscrete optimal transport reframed as optimization program

Consider general Polish spaces $\mathcal{X}, \mathcal{Y}$ and let $\mathcal{P}(\mathcal{Y})$ be the set of distributions on $\mathcal{Y}$. Consider also a generic finite set, $\mathbb{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subset \mathcal{X}$ be such that $\mathbf{x}_i \neq \mathbf{x}_j$, for $i \neq j$. In all this work, we consider $\mathcal{P}(\mathbb{X})$ the set of probabilities supported in this finite set. So any $P \in \mathcal{P}(\mathbb{X})$ can be written as

$$P := \sum_{k=1}^{N} p_k \delta_{\mathbf{x}_k}, \text{ where } p_i > 0, \text{ for all } i = 1, \ldots, N, \text{ and } \sum_{k=1}^{N} p_k = 1. \tag{3.1}$$

In consequence P is characterized by the vector $\mathbf{p} = (p_1, \ldots, p_N) \in \mathbb{R}^N$.

We focus on semi-discrete optimal transport cost which is defined as the optimal transport between a finite probability $P \in \mathcal{P}(\mathbb{X})$ and any probability $Q \in \mathcal{P}(\mathcal{Y})$.

The following result shows that the optimal transport problem in the semi-discrete case is equivalent to an optimization problem over a finite dimensional parameter space. Define the following function g_c, which depends on P and Q as

$$g_c(P, Q, \mathbf{z}) = \sum_{i=1}^{N} z_i p_i + \int \inf_{i=1,\ldots,N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\} dQ(\mathbf{y}), \quad \mathbf{z} \in \mathbb{R}^N \tag{3.2}$$

Lemma 3.1.1. *Let $P \in \mathcal{P}(\mathbb{X})$, $Q \in \mathcal{P}(\mathcal{Y})$ and c be a non-negative cost, then the optimal transport between P and Q for the cost c, $\mathcal{T}_c(P, Q)$, satisfies*

$$\mathcal{T}_c(P, Q) = \sup_{\mathbf{z} \in \mathbb{R}^N, \, |\mathbf{z}| \leq K^*} g_c(P, Q, \mathbf{z}). \tag{3.3}$$

for $K^* = \frac{1}{\inf_i p_i} \left(\sup_{i=1,\ldots,N} \int c(\mathbf{y}, \mathbf{x}_i) dQ(\mathbf{y}) \right)$. Moreover we can assume that $z_1 = 0$.

Remark 3.1.2. *Consider the dual expression for $\mathcal{T}_c(P, Q)$ and let φ denote an optimal transport potential from P to Q for the cost c, then*

$$\mathcal{T}_c(P, Q) = g_c(P, Q, (\varphi(\mathbf{x}_1), \ldots, \varphi(\mathbf{x}_N))).$$

Hence the optimal transport potentials and optimal values for (3.3) are linked through the expression $\mathbf{z} = (\varphi(\mathbf{x}_1), \ldots, \varphi(\mathbf{x}_N))$.

Note that $g_c(P, Q, \cdot)$ is a continuous function, a fact that follows the next lemma. Therefore, the sup in (3.3) is attained and the the class of optimal values

$$\mathrm{Opt}_c(P, Q) := \left\{\mathbf{z} \in \mathbb{R}^N : \mathcal{T}_c(P, Q) = g_c(P, Q, \mathbf{z})\right\} \tag{3.4}$$

and its restriction

$$\mathrm{Opt}_c^0(P, Q) := \left\{\mathbf{z} \in \mathbb{R}^N : \mathcal{T}_c(P, Q) = g_c(P, Q, \mathbf{z}), \; z_1 = 0\right\}. \tag{3.5}$$

are both non-empty.

Lemma 3.1.3. *If* $f(\mathbf{y}) = \inf_{i=1,\ldots,N}\{c(\mathbf{x}_i,\mathbf{y}) - z_i\}$ *and* $g(\mathbf{y}) = \inf_{i=1,\ldots,N}\{c(\mathbf{x}_i,\mathbf{y}) - s_i\}$, *then*

$$|f(\mathbf{y}) - g(\mathbf{y})| \leq \sup_{i=1,\ldots,N}\{|z_i - s_i|\} \leq |\mathbf{z} - \mathbf{s}|. \tag{3.6}$$

3.1.2 Main results : Central Limit Theorems for semi-discrete optimal transport cost

Our aim is to study the empirical semi-discrete optimal transport cost. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ and $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$ be two independent sequences of i.i.d. random variables with laws P and Q respectively, since $\mathbf{X}_k \in \mathbb{X}$ for all $k = 1, \ldots, n$, the empirical measure $P_n := \frac{1}{n}\sum_{k=1}^n \delta_{\mathbf{X}_k}$ belongs also to $\mathcal{P}(\mathbb{X})$. In consequence it can be written as $P_n := \sum_{k=1}^N p_k^n \delta_{\mathbf{x}_k}$, where $p_1^n, \ldots, p_N^n$ are real random variables such that $p_i^n \geq 0$, for all $i = 1, \ldots, N$, and $\sum_{k=1}^n p_k^n = 1$. We want to study the weak limit of the following sequences corresponding to all possible asymptotics

$$\left\{\sqrt{n}\left(\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q)\right)\right\}_{n\in\mathbb{N}}, \quad \left\{\sqrt{n}\left(\mathcal{T}_c(P, Q_m) - \mathcal{T}_c(P, Q)\right)\right\}_{m\in\mathbb{N}},$$

and the two sample case

$$\left\{\sqrt{\tfrac{nm}{n+m}}\left(\mathcal{T}_c(P_n, Q_m) - \mathcal{T}_c(P, Q)\right)\right\}_{m,n\in\mathbb{N}},$$

under the assumption $\frac{m}{n+m} \to \lambda \in (0,1)$.

To state the asymptotic behaviour we introduce first a centered Gaussian vector, $(\mathbf{X}_1, \ldots, \mathbf{X}_N)$ with covariance matrix $\Sigma(\mathbf{p})$ with entries

$$\Sigma(\mathbf{p})_{i,j} = -p_i p_j, \quad i \neq j \quad \text{and} \quad \Sigma(\mathbf{p})_{i,i} = p_i(1 - p_i). \tag{3.7}$$

We also define a centered Gaussian process $\mathbb{G}_Q^c$ in $\mathbb{R}^N$ with covariance function

$$\begin{aligned}\Xi_Q^c(\mathbf{z},\mathbf{s}) := &\int \inf_{i=1,\ldots,N}\{c(\mathbf{x}_i,\mathbf{y}) - z_i\} \inf_{i=1,\ldots,N}\{c(\mathbf{x}_i,\mathbf{y}) - s_i\} dQ(\mathbf{y}) \\ &- \int \inf_{i=1,\ldots,N}\{c(\mathbf{x}_i,\mathbf{y}) - z_i\} dQ(\mathbf{y}) \int \inf_{i=1,\ldots,N}\{c(\mathbf{x}_i,\mathbf{y}) - s_i\} dQ(\mathbf{y}).\end{aligned} \tag{3.8}$$

We can now state our main theorem.

Theorem 3.1.4. *Let* $P \in \mathcal{P}(\mathbb{X})$, $Q \in \mathcal{P}(\mathcal{Y})$, *c be non-negative and*

$$\int c(\mathbf{x}_i, \mathbf{y}) dQ(\mathbf{y}) < \infty, \ \textit{for all } i = 1, \ldots, m, \tag{3.9}$$

then the following limits hold.

- ***(One sample case for empirical discrete distribution*** P_n*)*

$$\sqrt{n}\,(\mathcal{T}_c(P_n,Q)-\mathcal{T}_c(P,Q)) \xrightarrow{w} \sup_{\mathbf{z}\in \mathrm{Opt}_c^0(P,Q)} \sum_{i=1}^{N} z_i \mathbb{U}_i.$$

Suppose that

$$\int c(\mathbf{x}_i,\mathbf{y})^2 dQ(\mathbf{y}) < \infty, \quad \text{for all } i = 1,\dots,N. \tag{3.10}$$

- ***(One sample case for empirical distribution*** Q_m*)*

$$\sqrt{m}\,(\mathcal{T}_c(P,Q_m)-\mathcal{T}_c(P,Q)) \xrightarrow{w} \sup_{\mathbf{z}\in \mathrm{Opt}_c^0(P,Q)} \mathbb{G}_Q^c(\mathbf{z}).$$

- ***(Two sample case)*** *if* $n,m\to\infty$*, with* $\frac{m}{n+m}\to\lambda\in(0,1)$*, then*

$$\sqrt{\frac{nm}{n+m}}\,(\mathcal{T}_c(P_n,Q_m)-\mathcal{T}_c(P,Q)) \xrightarrow{w} \sup_{\mathbf{z}\in \mathrm{Opt}_c^0(P,Q)} \Big(\sqrt{\lambda}\sum_{i=1}^{N} z_i\mathbb{U}_i + (\sqrt{1-\lambda})\mathbb{G}_Q^c(\mathbf{z})\Big).$$

Here $(\mathbb{U}_1,\dots,\mathbb{U}_N)\sim\mathcal{N}(\mathbf{0},\Sigma(\mathbf{p}))$*, with* $\Sigma(\mathbf{p})$ *as in* (3.7)*, and* $\mathbb{G}_Q^c$ *is a centered Gaussian process with covariance function* $\Xi^c(Q)$ *defined in* (3.8)*. Moreover* $\mathbb{G}_Q^c$ *and* $(\mathbb{U}_1,\dots,\mathbb{U}_N)$ *are independent.*

When $\mathcal{X}$ and $\mathcal{Y}$ are contained in the same Polish space $(\mathcal{Z},d)$, a particular cost that satisfies the assumptions of Theorem 3.1.4 is the metric d^p for all $p\geq 1$. Then applying Theorem 3.1.4 to the empirical estimations of $\mathcal{T}_{d^p}(P,Q)$ and a delta-method, enable to prove the asymptotic behaviour of the p-Wasserstein distance $\mathcal{W}_p^p(P,Q)=\mathcal{T}_{d^p}(P,Q)$ as given in the following corollary. The case $P=Q$ is a discrete optimal transport, so that its asymptotic behaviour has been previously studied in Sommerfeld and Munk (2018). Therefore, the following result assumes $\mathcal{W}_p(P,Q)\neq 0$.

Corollary 3.1.5. *Let* $P\in\mathcal{P}(\mathbb{X})$ *and* $Q\in\mathcal{P}(\mathcal{Z})$ *be such that* $\mathcal{W}_p(P,Q)\neq 0$ *and* $\int d(\mathbf{x}_0,\mathbf{y})^p dQ(\mathbf{y})<\infty$*, for some* $\mathbf{x}_0\in\mathbb{X}$*. Then, for any* $p\geq 1$*, we have*

- ***(One sample case for*** P*)*

$$\sqrt{n}\,(\mathcal{W}_p(P_n,Q)-\mathcal{W}_p(P,Q)) \xrightarrow{w} \frac{1}{p\,(\mathcal{W}_p(P,Q))^{p-1}} \sup_{\mathbf{z}\in \mathrm{Opt}_{d^p}(P,Q)} \sum_{i=1}^{N} z_i\mathbb{U}_i,$$

If we further assume that $\int d(\mathbf{x}_0,\mathbf{y})^{2p} dQ(\mathbf{y})<\infty$*, for some* $\mathbf{x}_0\in\mathbb{X}$*, then*

- ***(One sample case for Q)***

$$\sqrt{m}\left(\mathcal{W}_p(P,Q_m)-\mathcal{W}_p(P,Q)\right)\xrightarrow{w}\frac{1}{p\left(\mathcal{W}_p(P,Q)\right)^{p-1}}\sup_{\mathbf{z}\in\mathrm{Opt}_c^0(P,Q)}\mathbb{G}_Q^{d^p}(\mathbf{z}),$$

(Two sample case) *if $n,m\to\infty$ with $\frac{m}{n+m}\to\lambda\in(0,1)$, then*

$$\sqrt{\tfrac{nm}{n+m}}\left(\mathcal{W}_p(P_n,Q_m)-\mathcal{W}_p(P,Q)\right)\xrightarrow{w}\frac{\sup_{\mathbf{z}\in\mathrm{Opt}_c^0(P,Q)}\left(\sqrt{\lambda}\sum_{i=1}^N z_i\mathbb{U}_i+(\sqrt{1-\lambda})\mathbb{G}_Q^{d^p}(\mathbf{z})\right)}{p\left(\mathcal{W}_p(P,Q)\right)^{p-1}},$$

where $(\mathbb{U}_1,\dots,\mathbb{U}_N)$ follow a Gaussian distribution $\mathcal{N}(\mathbf{0},\Sigma(\mathbf{p}))$, with $\Sigma(\mathbf{p})$ defined in (3.7), and $\mathbb{G}_Q^{d^p}$ is a centered Gaussian process with covariance function $\Xi^{d^p}(Q)$, defined in (3.8). Moreover, $\mathbb{G}_Q^{d^p}$ and $(\mathbb{U}_1,\dots,\mathbb{U}_N)$ are independent.

We prove Theorem 3.1.4 in the two sample case. The same proof verbatim applies also for the CLT for the one sample case for Q. The one sample case for P can be proven under weaker moment assumptions on Q and will be commented separately.

3.1.3 An upper-bound on the expectation for the Wasserstein distance

Theorem 3.1.4 states the central limit theorem, when one of both probabilities is supported on a finite set. Now, we investigate the influence of the number of points of the discrete measure on the convergence bounds. In order to better understand the influence of the number of points, we will restrict our analysis to the Euclidean cost.

Theorem 3.1.6. *Let P be supported on N points in $\mathbb{X}$, $Q\in\mathcal{P}(\mathcal{Y})$ be a distribution with finite second order moment and Q_m its corresponding empirical version, then*

$$E\left|\mathcal{W}_1(P,Q_m)-\mathcal{W}_1(P,Q)\right|\le\tfrac{8\sqrt{2N}}{\sqrt{m}}K(\mathrm{diam}(\mathbb{X}),Q)$$

where

$$K(\mathrm{diam}(\mathbb{X}),Q)=\left(4\,\mathrm{diam}(\mathbb{X})+2\sqrt{\textstyle\int|\mathbf{y}|^2dQ(\mathbf{y})}+2\,\mathrm{diam}(\mathbb{X})\right)\left(\log(2)+\sqrt{2\,\mathrm{diam}(\mathbb{X})+1}\right)$$

and $\mathcal{W}_1$ is the 1-Wasserstein distance for the Euclidean distance. Moreover, if $\mathrm{diam}(\mathcal{Y})<\infty$, then

$$E\left|\mathcal{W}_p^p(P,Q_m)-\mathcal{W}_p^p(P,Q)\right|\le\tfrac{4\sqrt{2N}}{\sqrt{m}}K(4\,\mathrm{diam}(\mathbb{X})\,\mathrm{diam}(\mathcal{Y})^{p-1},Q).$$

The theorem provides a control on the consistency of the empirical bias for the Wasserstein distance. The rate becomes slower when N the number of points defining the support of the discrete measures P increases. If P models an approximation of a continuous probability on $\mathbb{R}^d$, hence the number N required to obtain a proper approximation grows exponentially larger when the dimension d increases. Hence the influence with respect to N stands for the curse of dimension.

The previous bound has a practical consequence in the following approximation problem. Assume that Q and P are probability distributions supported on a compact set $\Omega \subset \mathbb{R}^d$. Assume further that Q is unknown but observed through the empirical distribution Q_m. We approximate the (known) probability P by the N-points discretization P^N. If we aim at approximating the true 1-Wasserstein distance $\mathcal{W}_1(P,Q)$ from the empirical semi-discrete distance $\mathcal{W}_1(P^N, Q_m)$ (which is what can be indeed computed), Theorem 3.1.6 and the triangle inequality give the following upper bound

$$E\big|\mathcal{W}_1(P^N, Q_m) - \mathcal{W}_1(P,Q)\big| \leq \tfrac{8\sqrt{2N}}{\sqrt{m}} K(\Omega, Q) + \big|\mathcal{W}_1(P^N, Q) - \mathcal{W}_1(P,Q)\big|.$$

We can see that there is a trade-off between the size of the sample and the size of the discretization: the first term requires N/m to be small while the second term is only driven by N the discretization, being smaller when the number of points is larger.

3.2 Asymptotic Gaussian distribution optimal transport cost

Theorem 3.1.4 is valid for generic Polish spaces. When $\mathcal{X}, \mathcal{Y}$ are subsets of $\mathbb{R}^d$, the limit distribution in the CLT can be specified. Under the following regularity assumptions, we prove in this section that the limit distribution is Gaussian.

Let $Q \in \mathcal{P}(\mathbb{R}^d)$ be a probability measure absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}^d$. Assume that $c(\mathbf{x}, \mathbf{y}) = h(\mathbf{x} - \mathbf{y})$ where $h : \mathbb{R}^d \to [0, \infty)$ is a non negative function satisfying:

(A1): h is strictly convex on $\mathbb{R}^d$.

(A2): Given a radius $r \in \mathbb{R}^+$ and an angle $\theta \in (0, \pi)$, there exists some $M := M(r, \theta) > 0$ such that for all $|\mathbf{p}| > M$, one can find a cone

$$K(r, \theta, \mathbf{z}, \mathbf{p}) := \Big\{ \mathbf{x} \in \mathbb{R}^d : |\mathbf{x} - \mathbf{p}||\mathbf{z}| \cos(\theta/2) \leq \langle \mathbf{z}, \mathbf{x} - \mathbf{p} \rangle \leq r|\mathbf{z}| \Big\}, \qquad (3.11)$$

with vertex at $\mathbf{p}$ on which $h(\mathbf{x})$ attains its maximum at $\mathbf{p}$.

(A3): $\lim_{|\mathbf{x}| \to 0} \frac{h(\mathbf{x})}{|\mathbf{x}|} = \infty$.

Under such assumptions, Gangbo and McCann (1996) shows the existence of an optimal transport map T solving

$$\mathcal{T}_c(P, Q) := \inf_T \int c(\mathbf{y}, T(\mathbf{y})) dQ(\mathbf{y}), \quad \text{and} \quad T_\# Q = P, \qquad (3.12)$$

where $T_\# Q$ denotes the *push-forward* measure, defined for each measurable set A by $T_\# Q(A) := Q(T^{-1}(A))$. The minimizer in (3.12) is an *optimal transport map* from P to Q. Moreover, it is defined as the unique Borel function satisfying

$$T(\mathbf{x}) = \mathbf{x} - \nabla h^*(\nabla \varphi(\mathbf{x})), \quad \text{where } \varphi \text{ solves (3.3)}. \qquad (3.13)$$

Here h^* denotes the convex conjugate of h, see Rockafellar (1970). Such uniqueness enabled del Barrio et al. (2021) to deduce the uniqueness, under additive constants, of the solutions of (3.3) in φ. They assumed (A1)-(A3) to show that if two solutions of (3.3) have the same gradient almost everywhere for ℓ_d in a connected open set, then both are equal, up to an additive constant. In consequence, assuming that h is differentiable, the interior of the support of Q is connected and with Lebesgue negligible boundary, that is, $\ell_d\left(\partial\, \mathrm{supp}(Q)\right) = 0$, the uniqueness, up to additive constants, of the solutions of (3.3) holds. The proof of the main theorem in this section is a direct consequence of Lemma 3.2.1, which proves that there exists a unique, up to an additive constant, $\mathbf{z} \in \mathrm{Opt}(P, Q)$. We use within this section the notation $\mathbf{1} := (1, \dots, 1)$.

Lemma 3.2.1. *Let $P \in \mathcal{P}(\mathbb{X})$ and $Q \in \mathcal{P}(\mathbb{R}^d)$ be such that $Q \ll \ell_d$ and its support is connected with Lebesgue negligible boundary. If the cost c satisfies (A1)-(A3), is differentiable and*

$$\int c(\mathbf{x}_i, \mathbf{y})dQ(\mathbf{y}) < \infty, \quad \textit{for all } i = 1, \dots, N.$$

Then the set $\mathrm{Opt}_c^0(P, Q)$ is a singleton.

The following theorem states, under the previous assumptions, that the limit distribution described in Theorem 3.1.4 is the centered Gaussian variable $\left(\sqrt{\lambda}\sum_{i=1}^N z_i \mathbb{U}_i + (\sqrt{1-\lambda})\mathbb{G}_Q^c(\mathbf{z})\right)$ where $\{\mathbf{z}\} = \mathrm{Opt}_c^0(P, Q)$. Note that $\sum_{i=1}^N z_i \mathbb{U}_i$ is Gaussian and centered, with variance

$$\sigma^2(P, \mathbf{z}) = \mathrm{Var}(\sum_{i=1}^N z_i \mathbb{U}_i) \quad \text{and} \quad (\mathbb{U}_1, \dots, \mathbb{U}_N) \sim \mathcal{N}(\mathbf{0}, \Sigma(\mathbf{p})), \tag{3.14}$$

where $\Sigma(\mathbf{p})$ is defined in (3.7). On the other side $\mathbb{G}_Q^c(\mathbf{z})$ follows the distribution $\mathcal{N}(0, \sigma_c^2(Q, \mathbf{z}))$, where

$$\sigma_c^2(Q, \mathbf{z}) = \int \inf_{i=1,\dots,N}\{c(\mathbf{x}_i, \mathbf{y}) - z_i\}^2 dQ(\mathbf{y}) - \left(\int \inf_{i=1,\dots,N}\{c(\mathbf{x}_i, \mathbf{y}) - z_i\} dQ(\mathbf{y})\right)^2. \tag{3.15}$$

Since, for every $\lambda \in \mathbb{R}$, we have that $\sigma^2(P, \mathbf{z}) = \sigma^2(P, \mathbf{z} + \lambda\mathbf{1})$, then the asymptotic variance obtained in the following theorem is well defined.

Theorem 3.2.2. *Let $P \in \mathcal{P}(\mathbb{X})$ and $Q \in \mathcal{P}(\mathbb{R}^d)$ be such that $Q \ll \ell_d$ and its support is connected with Lebesgue negligible boundary. If the cost c satisfies (A1)-(A3), is differentiable and $\int c(\mathbf{x}_i, \mathbf{y})dQ(\mathbf{y}) < \infty$, for all $i = 1, \dots, N$, then*

- ***(One sample case for** P**)***

$$\sqrt{n}\left(\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q)\right) \xrightarrow{w} X \sim \mathcal{N}(0, \sigma^2(P, \mathbf{z})).$$

If, additionally, $\int c(\mathbf{x}_i, \mathbf{y})^2 dQ(\mathbf{y}) < \infty$, for all $i = 1, \dots, N$, then

- ***(One sample case for** Q)*

$$\sqrt{m}\,(\mathcal{T}_c(P,Q_m)-\mathcal{T}_c(P,Q)) \xrightarrow{w} Y \sim \mathcal{N}(0,\sigma_c^2(Q,\mathbf{z})).$$

- ***(Two sample case)** if $n,m\to\infty$, with $\frac{m}{n+m}\to\lambda\in(0,1)$, then*

$$\sqrt{\tfrac{nm}{n+m}}\,(\mathcal{T}_c(P_n,Q_m)-\mathcal{T}_c(P,Q)) \xrightarrow{w} \sqrt{\lambda}X+(\sqrt{1-\lambda})Y.$$

Here, $\sigma^2(P,\mathbf{z})$ and $\sigma_c^2(Q,\mathbf{z})$ are defined in (3.14) and (3.15) and, moreover, X and Y are independent.

As in the previous section, we provide an application to the CLT for Wasserstein distances. The potential costs $c_p=|\cdot|^p$, for $p>0$, satisfy (A1)-(A3), then the following result follows immediately from Theorem 3.2.2 and the Delta-Method for the function $t\mapsto|t|^{\frac{1}{p}}$. Recall that, in the potential cost cases, $\mathcal{T}_p(P,Q)$ denotes the optimal transport cost and $\mathcal{W}_p(P,Q)=(\mathcal{T}_p(P,Q))^{\frac{1}{p}}$ the p-Wasserstein distance.

Corollary 3.2.3. *Let $P\in\mathcal{P}(\mathbb{X})$ be as in (3.1) and $Q\in\mathcal{P}(\mathbb{R}^d)$ be such that $Q\ll\ell_d$, has finite moments of order p and its support is connected with Lebesgue negligible boundary. Then, for every $p>1$, we have that*

- ***(One sample case for** P)*

$$\sqrt{n}\,(\mathcal{T}_p(P_n,Q)-\mathcal{T}_p(P,Q)) \xrightarrow{w} \mathcal{N}(0,\sigma^2(P,\mathbf{z})),$$

and

$$\sqrt{n}\,(\mathcal{W}_p(P_n,Q)-\mathcal{W}_p(P,Q)) \xrightarrow{w} \mathcal{N}\left(0,\frac{\sigma^2(P,\mathbf{z})}{p^2\mathcal{W}_p(P,Q)^{2p-2}}\right),$$

Suppose that Q has finite moments of order $2p$, then

- ***(One sample case for** Q)*

$$\sqrt{m}\,(\mathcal{T}_p(P,Q_m)-\mathcal{T}_p(P,Q)) \xrightarrow{w} \mathcal{N}(0,\sigma_p^2(Q,\mathbf{z})),$$

and

$$\sqrt{m}\,(\mathcal{W}_p(P,Q_m)-\mathcal{W}_p(P,Q)) \xrightarrow{w} \mathcal{N}\left(0,\frac{\sigma_p^2(Q,\mathbf{z})}{p^2\mathcal{W}_p(P,Q)^{2p-2}}\right),$$

- ***(Two sample case)** if $n,m\to\infty$, with $\frac{m}{n+m}\to\lambda\in(0,1)$, then*

$$\sqrt{\tfrac{nm}{n+m}}\,(\mathcal{T}_p(P_n,Q_m)-\mathcal{T}_p(P,Q)) \xrightarrow{w} \mathcal{N}(0,\lambda\sigma^2(P,\mathbf{z})+(1-\lambda)\sigma_p^2(Q,\mathbf{z})),$$

and

$$\sqrt{\tfrac{nm}{n+m}}\,(\mathcal{W}_p(P_n,Q_m)-\mathcal{W}_p(P,Q)) \xrightarrow{w} \mathcal{N}\left(0,\frac{\lambda\sigma^2(P,\mathbf{z})+(1-\lambda)\sigma_p^2(Q,\mathbf{z})}{p^2\mathcal{W}_p(P,Q)^{2p-2}}\right).$$

Here, $\sigma^2(P,\mathbf{z})$ and $\sigma_p^2(Q,\mathbf{z})$ are defined in (3.14) and (3.15) for $\mathbf{z} \in \mathrm{Opt}_{|\cdot|^p}(P,Q)$ and the cost $|\cdot|^p$.

We observe that, since P is discrete and Q is continuous, $\mathcal{W}_p(P,Q) > 0$ and the limit distribution of Corollary 3.2.3 is always well defined. We note also that Corollary 3.2.3 is a particular case of Corollary 3.1.5 in the cases where the optimal transport potential is unique—the hypotheses of Theorem 3.2.2 hold—which is the reason why the case $p = 1$ can not be considered. Concerning other potential costs, $p > 1$, it is straightforward to see that the hypotheses (A1)-(A3) hold, see for instance del Barrio et al. (2021) or Gangbo and McCann (1996).

Remark 3.2.4. *Note that in this Gaussian limit case the variance of the limit can be consistently estimated. Let $\hat{\mathbf{z}}^{n,m}$ be a solution of $\mathrm{Opt}_c^0(P_n,Q_m)$, then del Barrio et al. (2021) proves that*

$$\sigma_{n,m}^2(P,Q) = \tfrac{n\,m}{n+m}\Big(\int (z_i^{n,m})^2 P_n(\mathbf{x_i}) - \Big(\int z_i^{n,m} P_n(\mathbf{x_i})\Big)^2 +$$
$$\int \big(\inf_{i=1,\dots,N}\{c(\mathbf{x}_i,\mathbf{y}) - z_i^{n,m}\}\big)^2 dQ_m(\mathbf{y}) - \Big(\int \inf_{i=1,\dots,N}\{c(\mathbf{x}_i,\mathbf{y}) - z_i^{n,m}\} dQ_m(\mathbf{y})\Big)^2\Big)$$

is a consistent estimator of $\lambda\sigma^2(P,\mathbf{z}) + (1-\lambda)\sigma_p^2(Q,\mathbf{z})$, in the two sample case. The same holds for the one sample cases. We underline that the value $\inf_{i=1,\dots,N}\{c(\mathbf{x}_i,\mathbf{y}) - z_i^{n,m}\}$ should not be computed in the two sample case—it is the solution of the (discrete-discrete) empirical dual problem.

3.3 Central Limit theorems for the potentials and Laguerre cells

3.3.1 A central Limit theorem for the potentials

The aim of this section is to provide a CLT for the empirical potentials, defined as the solutions to the empirical version of the dual formulation of the Monge-Kantorovich problem (3.3). In the semidiscrete case the potentials are pairs formed by $\mathbf{z} = (z_1,\dots,z_N) \in \mathrm{Opt}_c(P,Q)$ and $\varphi(\mathbf{y}) := \inf_{i=1,\dots,N}\{c(\mathbf{x}_i,\mathbf{y}) - z_i\}$. Note that potentials are defined up to a constant in the sense that if (ψ,φ) solves (3.3) then $(\psi + C, \varphi - C)$ also solves (3.3), for any constant C. Hence we will study the properties of the following functional, defined in $\langle \mathbf{1}\rangle^\perp$ which denotes the orthogonal complement of the vector space generated by $\mathbf{1} = (1,\dots,1)$

$$\mathcal{M}_\mathbf{p} : \langle \mathbf{1}\rangle^\perp \longrightarrow \mathbb{R}$$
$$\mathbf{z} \mapsto g_c(P,Q,\mathbf{z}),$$

where $g_c(P,Q,\mathbf{z})$ is defined as in (3.2). The first one holds by Theorem 3.4 in del Barrio et al. (2021) while the second by Corollary 6 in Pollard (2006).

In this section we will use some framework developed in Kitagawa et al. (2019). Hence, we make some slight changes in the notation, yet maintaining as much coherence as possible with the previous one. First we will assume that $\mathcal{Y}$ is an open domain of a d-dimensional Riemannian manifold $\mathcal{R}$ endowed with the volume measure $\mathcal{V}_g^d$ and metric g. We consider $\mathcal{C}(\mathcal{Y})$, $\mathcal{C}^1(\mathcal{Y})$ and $\mathcal{C}^{1,1}(\mathcal{Y})$ the spaces of real valued continuous functions, real valued continuously differentiable functions and the space of real valued continuously differentiable functions with Lipschitz derivatives, respectively.

Following the approach in Kitagawa et al. (2019), we assume that the cost satisfies the following assumptions

$$c(\mathbf{x}_i, \cdot) \in \mathcal{C}^{1,1}(\mathcal{Y}), \text{ for all } i = 1, \dots, N, \tag{Reg}$$

$$D_{\mathbf{y}} c(\mathbf{x}_i, \mathbf{y}) : \mathcal{Y} \to T^*_{\mathbf{y}}(\Omega) \text{ is injective as a function of } \mathbf{y}, \text{ for all } i = 1, \dots, N, \tag{Twist}$$

where $D_{\mathbf{y}} c$ denotes the partial derivative of c w.r.t. the second variable and $T^*_{\mathbf{y}}(\mathcal{Y})$ the tangent space. For every $i \in \{1, \dots, N\}$ there exists $\mathcal{Y}_i \subset \mathbb{R}^d$ open and convex set, and a $\mathcal{C}^{1,1}$ diffeomorphism $\exp_i^c : \mathcal{Y}_i \to \mathcal{Y}$ such that the functions

$$\mathcal{Y}_i \ni \mathbf{p} \mapsto f_{i,j}(\mathbf{p}) := c(\mathbf{x}_i, \exp_i^c \mathbf{p}) - c(\mathbf{x}_j, \exp_i^c \mathbf{p}) \text{ are quasi-convex for all } j = 1, \dots, N. \tag{QC}$$

Here quasi-convex, according to Kitagawa et al. (2019), means that for every $\lambda \in \mathbb{R}$ the sets $f_{i,j}^{-1}([-\infty, \lambda])$ are convex.
Besides the assumptions on the cost, we assume that the probability is supported in a *c-convex set* $\mathcal{Y}$, which means that $(\exp_i^c)^{-1}(\mathcal{Y})$ is convex, for every $i = 1, \dots, N$. Formally, let $\mathcal{Y} \subset \mathcal{R}$ be a compact c-convex set, $P \in \mathcal{P}(\mathbb{X})$ be as in (3.1) and suppose that

$$Q \in \mathcal{P}(\mathcal{Y}) \text{ satisfies } Q \ll \mathcal{V}_g^d \text{ with density } q \in \mathcal{C}(\mathcal{Y}). \tag{Cont}$$

The last required assumption in Kitagawa et al. (2019) is that Q satisfies a *Poincaré-Wirtinger inequality with constant* C_{PW}: a probability measure Q supported in a compact set $\mathcal{Y} \subset \mathcal{R}$ satisfies a Poincaré-Wirtinger inequality with constant C_{PW} if for every $f \in \mathcal{C}^1(\mathcal{Y})$ we have that for $Y \sim Q$

$$E(|f(Y) - E(f(Y))|) \leq C_{PW} E(|\nabla f(Y)|). \tag{PW}$$

In order to clarify the feasibility of such assumptions, we will provide some insights on them at the end of the section. Kitagawa et al. (2019) proved the following assertions.

1. Under assumptions (Reg) and (Twist) the function $\mathcal{M}_{\mathbf{p}}(\mathbf{z})$ is concave and differentiable with derivative

$$\nabla_{\mathbf{z}} \mathcal{M}_{\mathbf{p}}(\mathbf{z}) = (-Q(\mathrm{Lag}_1(\mathbf{z})) + p_1, \dots, -Q(\mathrm{Lag}_N(\mathbf{z})) + p_N), \tag{3.16}$$

where

$$\mathrm{Lag}_k(\mathbf{z}) := \{\mathbf{y} \in \mathbb{R}^d : \ c(\mathbf{x}_k, \mathbf{y}) - z_k < c(\mathbf{x}_i, \mathbf{y}) - z_i, \ \text{for all } i \neq k\}. \tag{3.17}$$

2. Under assumptions (Reg), (Twist) and (QC), the function $\mathcal{M}_{\mathbf{p}}$ is twice continuously differentiable with Hessian matrix $D^2_{\mathbf{z}}\mathcal{M}_{\mathbf{p}}(\mathbf{z}) = \left(\frac{\partial^2 \mathcal{M}_{\mathbf{p}}}{\partial z_i \partial z_j}(\mathbf{z})\right)_{i,j=1,\dots,N}$ and partial derivatives

$$\frac{\partial^2}{\partial z_i \partial z_j}\mathcal{M}_{\mathbf{p}}(\mathbf{z}) = \int_{\mathrm{Lag}_k(\mathbf{z})\cap \mathrm{Lag}_k(\mathbf{z})} \frac{q(\mathbf{y})}{|\nabla_{\mathbf{y}} c(\mathbf{x}_i,\mathbf{y}) - \nabla_{\mathbf{y}} c(\mathbf{x}_j,\mathbf{y})|} d\mathcal{V}_g^{d-1}(\mathbf{y}), \quad \text{if } i \neq j,$$
$$\frac{\partial^2}{\partial^2 z_i}\mathcal{M}_{\mathbf{p}}(\mathbf{z}) = -\sum_{j\neq i} \frac{\partial^2}{\partial z_i \partial z_j}\mathcal{M}_{\mathbf{p}}(\mathbf{z}). \tag{3.18}$$

3. Under assumptions (Reg),(Twist) and (QC), and if Q satisfies (PW), then there exists a positive constant C such that

$$\langle D^2_{\mathbf{z}}\mathcal{M}_{\mathbf{p}}(\mathbf{z})\mathbf{v},\mathbf{v}\rangle \leq -C\epsilon^3|\mathbf{v}|^2, \text{ for all } \mathbf{z}\in\mathcal{K}^\epsilon \text{ and } \mathbf{v}\in\langle\mathbf{1}\rangle^\perp, \tag{3.19}$$

where
$$\mathcal{K}^\epsilon := \{\mathbf{z}\in\mathbb{R}^d : \ Q(\mathrm{Lag}_i(\mathbf{z})) > \epsilon, \quad \text{for all } i = 1,\dots,N.\}.$$

Under this assumptions we can state the main result of the section: a CLT for the O.T. potentials.

Theorem 3.3.1. *Let $\mathcal{Y}\subset\mathcal{R}$ be a compact c-convex set, $P\in\mathcal{P}(\mathbb{X})$ and $Q\in\mathcal{P}(\mathcal{Y})$. Under Assumptions (Reg), (Twist) and (QC) on the cost c, and (PW) and (Cont) on Q, then the following limits hold.*

- ***(One sample case for** P**)***

$$\sqrt{n}(\hat{\mathbf{z}}^n - \mathbf{z}^*) \xrightarrow{w} -(D^2\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*))^{-1}((\mathbb{U}_1,\dots,\mathbb{U}_N)),$$

- ***(One sample case for** Q**)***

$$\sqrt{m}(\hat{\mathbf{z}}^m - \mathbf{z}^*) \xrightarrow{w} -(D^2\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*))^{-1}((\mathbb{U}_1,\dots,\mathbb{U}_N))$$

- ***(Two sample case)** if $n,m\to\infty$, with $\frac{m}{n+m}\to\lambda\in(0,1)$, then*

$$\sqrt{\frac{n\,m}{n+m}}(\hat{\mathbf{z}}^{n,m} - \mathbf{z}^*) \xrightarrow{w} -(D^2\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*))^{-1}((\mathbb{U}_1,\dots,\mathbb{U}_N)).$$

Here $\mathbf{z}^*\in\langle\mathbf{1}\rangle^\perp\cap\mathrm{Opt}_c(P,Q)$, $\hat{\mathbf{z}}^n\in\langle\mathbf{1}\rangle^\perp\cap\mathrm{Opt}_c(P_n,Q)$, $\hat{\mathbf{z}}^m\in\langle\mathbf{1}\rangle^\perp\cap\mathrm{Opt}_c(P,Q_m)$ $\hat{\mathbf{z}}^{n,m}\in\langle\mathbf{1}\rangle^\perp\cap\mathrm{Opt}_c(P_n,Q_m)$ and $(\mathbb{U}_1,\dots,\mathbb{U}_N)\sim\mathcal{N}(\mathbf{0},\Sigma(\mathbf{p}))$, for $\Sigma(\mathbf{p})$ defined in (3.7).

For $\mathbf{z}^*$ defined as in Theorem 3.3.1, set

$$\varphi(\mathbf{y}) := \inf_{i=1,\dots,N}\{c(\mathbf{x}_i,\mathbf{y}) - z_i^*\} \tag{3.20}$$

and note that it is an optimal transport map from Q to P, set also the value $i(\mathbf{y}) \in \{1, \ldots, N\}$ where the infumum of (3.20) is attained. As before, we can define their empirical counterparts

$$\varphi_{n,m}(\mathbf{y}) := \inf_{i=1,\ldots,N}\{c(\mathbf{x}_i, \mathbf{y}) - \hat{z}_i^{n,m}\},\ \varphi_n(\mathbf{y}) := \inf_{i=1,\ldots,N}\{c(\mathbf{x}_i, \mathbf{y}) - \hat{z}_i^{n}\},$$
$$\text{and} \quad \varphi_m(\mathbf{y}) := \inf_{i=1,\ldots,N}\{c(\mathbf{x}_i, \mathbf{y}) - \hat{z}_i^{m}\}, \tag{3.21}$$

which are an optimal transport map from Q to P_n and $i_{n,m}(\mathbf{y})$ the index where the infimum of (3.21) is attained for the two sample case. Then we have that

$$\sqrt{\tfrac{n\,m}{n+m}}(\hat{z}^{n,m}_{i_{n,m}(\mathbf{y})} - z^*_{i_{n,m}(\mathbf{y})}) \le \sqrt{\tfrac{n\,m}{n+m}}(\varphi_{n,m}(\mathbf{y}) - \varphi(\mathbf{y})) \le \sqrt{\tfrac{n\,m}{n+m}}(\hat{z}^{n,m}_{i(\mathbf{y})} - z^*_{i(\mathbf{y})}). \tag{3.22}$$

We can take supremum over $\mathbf{y}$ in both sides of (3.22) and derive that

$$\sqrt{\tfrac{n\,m}{n+m}} \sup_{i=1,\ldots,N}(\hat{z}_i^{n,m} - z_i^*) = \sqrt{\tfrac{n\,m}{n+m}} \sup_{\mathbf{y}\in\mathcal{Y}}(\varphi_{n,m}(\mathbf{y}) - \varphi(\mathbf{y})).$$

By symmetry we have that

$$\sqrt{\tfrac{n\,m}{n+m}}|\hat{z}_i^{n,m} - z_i^*| = \sqrt{\tfrac{n\,m}{n+m}} \sup_{\mathbf{y}\in\mathcal{Y}}|\varphi_{n,m}(\mathbf{y}) - \varphi(\mathbf{y})|,$$

which implies the following corollary.

Corollary 3.3.2. *Under the hypotheses and notation of Theorem 3.3.1, for φ, φ_n, φ_m and $\varphi_{n,m}$ defined in (3.20) and (3.21), we have the following limits.*

- ***(One sample case for** P**)***

$$\sqrt{n} \sup_{\mathbf{y}\in\mathcal{Y}}|\varphi_n(\mathbf{y}) - \varphi(\mathbf{y})| \xrightarrow{w} \left|(D^2\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*))^{-1}((\mathbb{U}_1, \ldots, \mathbb{U}_N))\right|_\infty$$

- ***(One sample case for** Q**)***

$$\sqrt{m} \sup_{\mathbf{y}\in\mathcal{Y}}|\varphi_m(\mathbf{y}) - \varphi(\mathbf{y})| \xrightarrow{w} \left|(D^2\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*))^{-1}((\mathbb{U}_1, \ldots, \mathbb{U}_N))\right|_\infty$$

- ***(Two sample case)** if $n, m \to \infty$, with $\frac{m}{n+m} \to \lambda \in (0, 1)$, then*

$$\sqrt{\tfrac{n\,m}{n+m}} \sup_{\mathbf{y}\in\mathcal{Y}}|\varphi_{n,m}(\mathbf{y}) - \varphi(\mathbf{y})| \xrightarrow{w} \left|(D^2\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*))^{-1}((\mathbb{U}_1, \ldots, \mathbb{U}_N))\right|_\infty.$$

We will conclude by some comments on the assumptions made in this section.

1. Under the hypotheses of Theorem 3.3.1, the optimal potential is unique once we fix its value at a given point. Then Corollary 3.3.2 provides a uniform confidence band for this optimal potential, namely,

$$\left[\varphi_n(\mathbf{y}) \pm \tfrac{\Delta_\alpha}{\sqrt{n}}\right]_{\mathbf{y}\in\mathcal{Y}},$$

where Δ_α is the $1-\alpha$ quantile of the limit distribution.

2. Note that if we consider $\mathcal{R} = \mathbb{R}^d$ and the quadratic cost, then (Reg), (Twist) and (QC) are obviously satisfied, by taking the function $\exp_j$ as the identity. Actually the map $\mathbf{y} \mapsto |\mathbf{x}_j - \mathbf{y}|^2$ is $\mathcal{C}^\infty(\mathbb{R}^d)$ and $\mathbf{y} - \mathbf{x}_j$ is its derivative w.r.t. $\mathbf{y}$. Finally note that the function

$$\mathbb{R}^d \ni \mathbf{p} \mapsto |\mathbf{x}_i - \mathbf{p}|^2 - |\mathbf{x}_j - \mathbf{p}|^2 = |\mathbf{x}_i|^2 - |\mathbf{x}_j|^2 + \langle \mathbf{x}_j - \mathbf{x}_i, \mathbf{p}\rangle$$

is linear in $\mathbf{p}$ and consequently quasi-convex.

3. Assumption (PW) on the probability Q has been widely studied in the literature for its implications in PDEs, see Acosta and Durán (2004). They proved that (PW) holds for a uniform distribution on a convex set $\mathcal{Y}$. In Rathmair (2019), Lemma 1 claims that (PW) is equivalent to the bound of $\inf_{t\in\mathbb{R}} E(|f(Y) - t|)$, for every $f \in \mathcal{C}^1(\mathcal{Y})$. Let $Y \sim Q$ be such that there exists a $\mathcal{C}^1(\mathcal{Y})$ map T satisfying the relation $T(U) = Y$, where U follows a uniform distribution on a compact convex set A. Since $f \circ T \in \mathcal{C}^1(A)$, by the powerful result of Acosta and Durán (2004), there exists $C_A > 0$ such that

$$\begin{aligned}\inf_{t\in\mathbb{R}} E(|f(Y) - t|) = \inf_{t\in\mathbb{R}} E(|f(T(U)) - t|) &\leq C_A E(|\nabla f(T(U))| \cdot ||T'(U)||_2)\\ &\leq C_A \sup_{\mathbf{u}\in A} ||T'(\mathbf{u})||_2 E(|\nabla f(T(U))|),\end{aligned}$$

where $||T'(U)||_2$ denotes the matrix operator norm. We conclude that, in such cases, (PW) holds. Note that the existence of this map relies on the well known existence of continuously differentiable optimal transport maps, which is treated by Caffarelli's theory. This is the case, for instance of log-concave probability measures (see Caffarelli (1996)). As pointed out by an anonymous reviewer, we can arrive directly to this conclusion using Milman (2007). We refer to the most recent work Cordero-Erausquin and Figalli (2019) and references therein. However, as pointed out in Kitagawa et al. (2019), more general probabilities can satisfy that assumption such as radial functions on $\mathbb{R}^d$ with density

$$\frac{p(|\mathbf{x}|)}{|\mathbf{x}|^{d-1}}, \quad \text{for } |\mathbf{x}| \leq R, \text{ with } p = 0 \text{ in } [0, r] \text{ and concave in } [r, R].$$

Moreover the spherical uniform $\mathbb{U}_d$, used in Hallin et al. (2021) to generalize the distribution function to higher dimension, also satisfies (PW). This can be proved by using previous argument with the function $T(\mathbf{x}) = \mathbf{x}|\mathbf{x}|^{d-1}$, which is continuously differentiable. But note that this probability measure does not satisfy (Cont). We conjecture that Theorem 3.3.1 still holds in this case, but some additional, that we leave for future work is needed. Similarly, the regularity of the transport can be obtained in the continuous case by a careful treatment of the Monge-Ampére equation, see del Barrio et al. (2020).

4. The limit distribution described in 3.3.1 is not easy to derive, even knowing the exact probabilities P and Q. But note that the limits are consequence of its transformation as a Z-estimation problem (eg. chapter 3.3 in Vaart and Wellner (1996)), as the limit is a N-dimensional Gaussian, using example 3.9.35 in Vaart and Wellner (1996) we obtain the consistency of the parametric bootstrap. Hence a bootstrap approximation can be used to approximate the limit distribution. The approximation will be consistent as in Fang and Santos (2018). In Figure 3.3.1 we compute such an approximation by using bootstrap where P is supported on three points in $\mathbb{R}^6$ and Q is the uniform on $(0,1)^6$.

3.3.2 A central Limit theorem for the Laguerre cells and square-Euclidean cost

A natural question for semi-discrete problems is the convergence of the Laguerre tessellations. Actually, the semi-discrete framework is mainly applied to quantization, sampling or resource allocation problems. The Laguerre cells represent the optimal cluster (quantization or sampling) or the population choosing certain product in a resource allocation problems. Therefore, our objective will be to infer the population Laguerre cell from the empirical one. Note that, although we are working with probabilities supported on a compact set, the cells may not be bounded. Hence for $0 < R < +\infty$, define $R\,\mathbb{B}_d$ the ball with radius R. We will consider in the following, the restricted version of Laguerre cell

$$\mathrm{Lag}_i^R(\mathbf{z}) = \overline{\mathrm{Lag}_i(\mathbf{z}) \cap R\,\mathbb{B}_d},$$

which are compact, which implies that the distances between the empirical and the population can be measured by means of the support functions. Recall (cf. p. 317 in Rockafellar and Wets (1998) eg.) that the support function of a set $A \subset \mathbb{R}^d$ is defined as the functional in the unit sphere

$$\mathbb{S}^{d-1} \ni \mathbf{v} \to h_A(\mathbf{v}) := \sup_{\mathbf{y}\in\mathbf{A}} \langle \mathbf{v}, \mathbf{x} \rangle.$$

Due to purely geometrical reasons, we are forced to restrict ourselves to the Euclidean case with quadratic cost—otherwise the following argument cannot be applied (see e.g. Remark 5.1 in Bansil and Kitagawa (2020)).

To give a characterization of the convergence of sets usually the L^p metrics, with $p \in [1, \infty]$, are used. Recall that, for $p \in [1, \infty)$ the L^p metric is defined for two sets A, B as

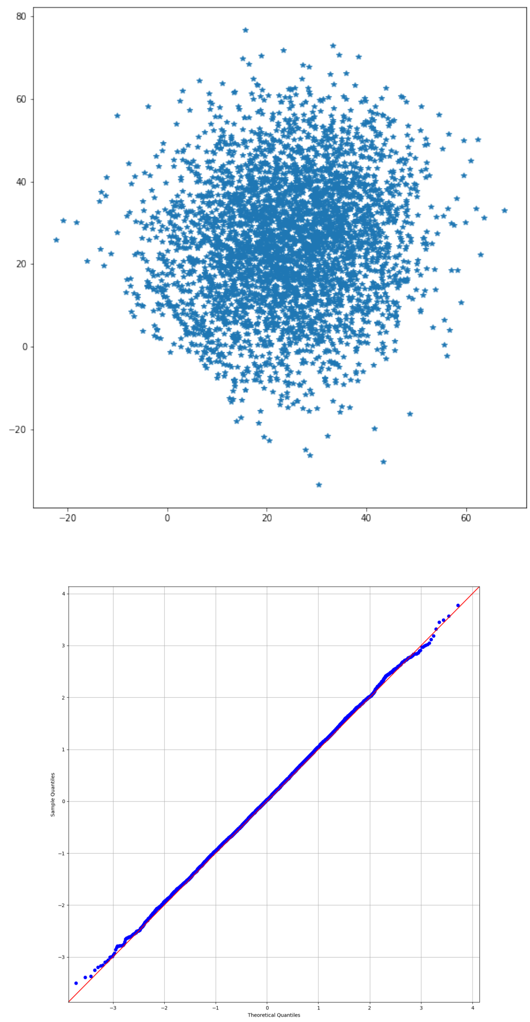

Figure 3.3.1: Bootstrap approximation of $N(\mathbf{0}, \Sigma(\mathbf{z}^*))$. Here P is supported in tree points $P = \frac{1}{3}(\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3)$, and Q is uniform on $(0,1)^6$. We assume that P_n is deterministic, we compute the empirical potentials $\hat{\mathbf{z}}$ for a sample of $10,000$ points and the Bootstrap potentials $\hat{\mathbf{z}}^s$, for $s = 1, \dots, 10,000$. Both—the empirical and the bootstrap—are projected to the space $\langle \mathbf{1} \rangle^{\perp}$. Since the space $\langle \mathbf{1} \rangle^{\perp}$ is, in this case, $2-$dimensional, we can plot the $2D$ distribution of $(\hat{\mathbf{z}} - \hat{\mathbf{z}}^s)\sqrt{10000}$ (left). The qq-plot of the projection to the second coordinate is in the right hand side.

$d_p(A, B) := (\int |h_A - h_B|^p d\mathcal{H}^{d-1})^{\frac{1}{p}}$, where $\mathcal{H}^{d-1}$ is the Hausdorff measure in $\mathbb{S}^{d-1}$. The case $p = \infty$ is $d_\infty(A, B) = \sup_{v \in \mathbb{S}^{d-1}} |h_A - h_B|$, which corresponds with the Hausdorff distance, i.e.

$$\inf\{\epsilon > 0 : \ A \subset B + \epsilon\mathbb{B} \text{ and } B \subset A + \epsilon\mathbb{B}\}.$$

Note that all these norms are equivalent for compact convex sets (see Vitale (1985)), which is our case.
For any interior point $\mathbf{y}^0$ of $\mathrm{Lag}_k^R(\mathbf{z}^*)$, we set the notation

$$\mathrm{Sol}(\mathbf{z}^*, \mathbf{v}, \mathbf{y}^0) := \arg\min_{t_j>0} \Big\{ \sum_{j\neq k} t_j\psi_j + R|\mathbf{v} - \sum_{j\neq k} t_j(\mathbf{x}_k - \mathbf{x}_j)| - \langle \mathbf{v} - \sum_{j\neq k} t_j(\mathbf{x}_k - \mathbf{x}_j), \mathbf{y}^0\rangle \Big\},$$

where $\psi_j(\mathbf{z}^*) = (|\mathbf{x}_k|^2 - |\mathbf{x}_j|^2 - z_k^* + z_j^*) - \langle \mathbf{x}_k - \mathbf{x}_j, \mathbf{y}^0\rangle$ and $\mathbf{z}^* \in \langle \mathbf{1}\rangle^\perp \cap \mathrm{Opt}_{|\cdot|^2}(P, Q)$, for the quadratic cost $|\cdot|^2$. With this notation, the following result provides the weak limit, with parametric rate, of the L^p distance, for $p \in [0, \infty)$, between the empirical and population Laguerre cells.

Theorem 3.3.3. *Let $\mathcal{Y} \subset \mathbb{R}^d$ be a compact convex set such that $\mathcal{Y} \subset R\,\mathbb{B}_d$, $P \in \mathcal{P}(\mathbb{X})$ and $Q \in \mathcal{P}(\mathcal{Y})$. Under Assumptions (PW) and (Cont) on Q and considering the quadratic cost $|\cdot|^2$, we have the following limits, for $p \in (1, \infty)$.*

- ***(One sample case for P)***

$$\sqrt{n} d_p(\mathrm{Lag}_k^R(\hat{\mathbf{z}}^n), \mathrm{Lag}_k^R(\mathbf{z}^*)) \xrightarrow{w} \Big(\int \Big| \inf_{\mathbf{t}\in \mathrm{Sol}(\mathbf{z}^*, \mathbf{v}, \mathbf{y}^0)} \Big\{ \sum_{j\neq k} t_j(\mathbb{M}_k - \mathbb{M}_j) \Big\} \Big|^p d\mathcal{H}^{d-1} \Big)^{\frac{1}{p}}.$$

- ***(One sample case for Q)***

$$\sqrt{m} d_p(\mathrm{Lag}_k^R(\hat{\mathbf{z}}^m), \mathrm{Lag}_k^R(\mathbf{z}^*)) \xrightarrow{w} \Big(\int \Big| \inf_{\mathbf{t}\in \mathrm{Sol}(\mathbf{z}^*, \mathbf{v}, \mathbf{y}^0)} \Big\{ \sum_{j\neq k} t_j(\mathbb{M}_k - \mathbb{M}_j) \Big\} \Big|^p d\mathcal{H}^{d-1} \Big)^{\frac{1}{p}}.$$

- ***(Two sample case)*** *if $n, m \to \infty$, with $\frac{m}{n+m} \to \lambda \in (0, 1)$, then*

$$\sqrt{\tfrac{nm}{n+m}} d_p(\mathrm{Lag}_k^R(\hat{\mathbf{z}}^{n,m}), \mathrm{Lag}_k^R(\mathbf{z}^*)) \xrightarrow{w} \Big(\int \Big| \inf_{\mathbf{t}\in \mathrm{Sol}(\mathbf{z}^*, \mathbf{v}, \mathbf{y}^0)} \Big\{ \sum_{j\neq k} t_j(\mathbb{M}_k - \mathbb{M}_j) \Big\} \Big|^p d\mathcal{H}^{d-1} \Big)^{\frac{1}{p}}.$$

Here $\mathbf{z}^* \in \langle \mathbf{1}\rangle^\perp \cap \mathrm{Opt}_{|\cdot|^2}(P, Q)$, for the quadratic cost $|\cdot|^2$ and

$$(\mathbb{M}_1, \ldots, \mathbb{M}_N) = \left(D^2\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*)\right)^{-1} ((\mathbb{U}_1, \ldots, \mathbb{U}_N)),$$

where $(\mathbb{U}_1, \ldots, \mathbb{U}_N) \sim \mathcal{N}(\mathbf{0}, \Sigma(\mathbf{p}))$, for $\Sigma(\mathbf{p})$, is defined in (3.7).

The proof is based on the following description of the support functions of Laguerre's cells

$$h_{\mathrm{Lag}_k^R(\mathbf{z})}(\mathbf{v}) = \min_{t_j>0} \Big\{ \sum_{j\neq k} t_j(|\mathbf{x}_k|^2 - |\mathbf{x}_j|^2 - z_k + z_j) + R|\mathbf{v} - \sum_{j\neq k} t_j(\mathbf{x}_k - \mathbf{x}_j)| \Big\}, \quad (3.23)$$

which allows to apply standard arguments to handle Hadamard derivative of the infimum. However, the compactness of the solution-set of (3.23) is not guaranteed. Not that the same problem happens in Bansil and Kitagawa (2020).Therefore we adopt the same reparation strategy (cf. Remark 5.1 in Bansil and Kitagawa (2020)). This consists in setting an interior point $\mathbf{y}^0$ and observing that Remark 3.3.5 below implies that

$$h_{\mathrm{Lag}_k^R(\hat{\mathbf{z}}^{n,m})}(\mathbf{v}) - h_{\mathrm{Lag}_k^R(\mathbf{z}^*)}(\mathbf{v}) = h_{\mathrm{Lag}_k^R(\hat{\mathbf{z}}^{n,m})+\{-\mathbf{y}^0\}}(\mathbf{v}) - h_{\mathrm{Lag}_k^R(\mathbf{z}^*)+\{-\mathbf{y}^0\}}(\mathbf{v}).$$

In such a case

$$h_{\mathrm{Lag}_k^R(\mathbf{z}^*)+\{-\mathbf{y}^0\}}(\mathbf{v}) = \min_{t_j>0}\Big\{\sum_{j\neq k} t_j\psi_j(\mathbf{z}^*)+R|\mathbf{v}-\sum_{j\neq k} t_j(\mathbf{x}_k-\mathbf{x}_j)|-\langle \mathbf{v}-\sum_{j\neq k} t_j(\mathbf{x}_k-\mathbf{x}_j), \mathbf{y}^0\rangle\Big\}, \tag{3.24}$$

where, for all $j \neq k$, $\psi_j(\mathbf{z}^*) = (|\mathbf{x}_k|^2 - |\mathbf{x}_j|^2 - z_k^* + z_j^*) - \langle \mathbf{x}_k - \mathbf{x}_j, \mathbf{y}^0\rangle \geq a > 0$, for some $a > 0$. This uniform bound implies the compactness of $\mathrm{Sol}(\mathbf{z}^*, \mathbf{v}, \mathbf{y}^0)$ and gives, as an intermediate step, the point-wise limit of the support functions.

Lemma 3.3.4. *Let $\mathcal{Y} \subset \mathbb{R}^d$ be a compact convex set such that $\mathcal{Y} \subset R\,\mathbb{B}_d$, $P \in \mathcal{P}(\mathbb{X})$ and $Q \in \mathcal{P}(\mathcal{Y})$. Under Assumptions (PW) and (Cont) on Q, we have the following limits.*

- ***(One sample case for*** P***)***

$$\sqrt{n}(h_{\mathrm{Lag}_k^R(\hat{\mathbf{z}}^n)}(\mathbf{v}) - h_{\mathrm{Lag}_k^R(\mathbf{z}^*)}(\mathbf{v})) \xrightarrow{w} \inf_{\mathbf{t}\in \mathrm{Sol}(\mathbf{z}^*,\mathbf{v},\mathbf{y}^0)} \Big\{\sum_{j\neq k} t_j(\mathbb{M}_k - \mathbb{M}_j)\Big\}.$$

- ***(One sample case for*** Q***)***

$$\sqrt{m}(h_{\mathrm{Lag}_k^R(\hat{\mathbf{z}}^m)}(\mathbf{v}) - h_{\mathrm{Lag}_k^R(\mathbf{z}^*)}(\mathbf{v})) \xrightarrow{w} \inf_{\mathbf{t}\in \mathrm{Sol}(\mathbf{z}^*,\mathbf{v},\mathbf{y}^0)} \Big\{\sum_{j\neq k} t_j(\mathbb{M}_k - \mathbb{M}_j)\Big\}.$$

- ***(Two sample case)*** *if $n, m \to \infty$, with $\frac{m}{n+m} \to \lambda \in (0, 1)$, then*

$$\sqrt{\tfrac{n\,m}{n+m}}(h_{\mathrm{Lag}_k^R(\hat{\mathbf{z}}^{n,m})}(\mathbf{v}) - h_{\mathrm{Lag}_k^R(\mathbf{z}^*)}(\mathbf{v})) \xrightarrow{w} \inf_{\mathbf{t}\in \mathrm{Sol}(\mathbf{z}^*,\mathbf{v},\mathbf{y}^0)} \Big\{\sum_{j\neq k} t_j(\mathbb{M}_k - \mathbb{M}_j)\Big\}.$$

Here $(\mathbb{M}_1, \ldots, \mathbb{M}_N) = \left(D^2\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*)\right)^{-1}((\mathbb{U}_1, \ldots, \mathbb{U}_N))$, where $(\mathbb{U}_1, \ldots, \mathbb{U}_N) \sim \mathcal{N}(\mathbf{0}, \Sigma(\mathbf{p}))$, for $\Sigma(\mathbf{p})$, is defined in (3.7), and

Remark 3.3.5. *Let us recall some basic properties of the support function (cf. Corollary 11.24 in Rockafellar and Wets (1998) eg.). Let $A, B \subset \mathbb{R}^d$ be non empty sets and $\lambda > 0$, then:*

- $h_{\lambda A} = \lambda h_A$,
- $h_{A+B} = h_A + h_B$,
- $h_{A\cup B} = \max(h_A, h_B)$,

- *and, if A and B are convex $A \cap B \neq \emptyset$, then* $h_{A\cap B}(\mathbf{v}) = \inf_{\mathbf{u}+\mathbf{w}=\mathbf{v}}(h_A(\mathbf{u}) + h_B(\mathbf{w}))$.

Remark 3.3.5 is important for the proof of Theorem 3.3.3. It also has been extracted here because of the interpretation it gives of the limits. Note that thanks to it we can obtain, equivalently, the following limit:

$$d_p\left(\sqrt{\tfrac{n\,m}{n+m}}\,\mathrm{Lag}_k^R(\hat{\mathbf{z}}^{n,m}), \sqrt{\tfrac{n\,m}{n+m}}\,\mathrm{Lag}_k^R(\mathbf{z}^*)\right) \xrightarrow{w} \left(\int \Big| \inf_{\mathbf{t}\in \mathrm{Sol}(\mathbf{z}^*,\mathbf{v},\mathbf{y}^0)} \Big\{ \sum_{j\neq k} t_j(\mathbb{M}_k - \mathbb{M}_j) \Big\} \Big|^p d\mathcal{H}^{d-1}\right)^{\frac{1}{p}}. \quad (3.25)$$

To go from point-wise convergence to L^p convergence, the proof of Theorem 3.3.3 uses the following result, which is direct consequence of Remark 5.1 in Bansil and Kitagawa (2020) and Theorem 3.9 in Segers (2022) (see also Theorem 3.4 in del Barrio et al. (2021)).

Lemma 3.3.6. *Under the hypotheses and notation of Theorem 3.3.3, we have*

$$t_n d_\infty(\mathrm{Lag}_k^R(\mathbf{z}^* + t_n\mathbf{h}_n), \mathrm{Lag}_k^R(\mathbf{z}^*)) \leq \tfrac{4R}{\delta_n}|\mathbf{h}_n|_\infty,$$

for any $t_n \searrow 0$ and $\mathbf{d}_n \to \mathbf{d}$, with

$$\delta_n = \sup_{\mathbf{y}\in \mathrm{Lag}_k^R(\mathbf{z}^*+t_n\mathbf{h}_n)\cap \mathrm{Lag}_k^R(\mathbf{z}^*)} \min\left(d_\infty(\{\mathbf{y}\}, \partial\,\mathrm{Lag}_k^R(\mathbf{z}^*)), d_\infty(\{\mathbf{y}\}, \partial\,\mathrm{Lag}_k^R(\mathbf{z}^* + t_n\mathbf{h}_n))\right).$$

Moreover, if $\mathbf{u} + 2a\mathbb{B}_d \subset \mathrm{Lag}_k^R(\mathbf{z}^*)$, then $\liminf_n \delta_n \geq a$.

Lemma (3.3.6) gives confidence intervals for the Hausdorff distance between the cells– uncovered in Theorem 3.3.3. Note that Lemma (3.3.6) implies that

$$\mathbb{P}\left(\mathrm{Lag}_k^R(\mathbf{z}^*) \subset \mathrm{Lag}_k^R(\hat{\mathbf{z}}^m) + \tfrac{4R\Psi^{-1}(\alpha)}{a\sqrt{m}}\mathbb{B}_d\right) \geq \mathbb{P}\left(d_\infty(\mathrm{Lag}_k^R(\mathbf{z}^*), \mathrm{Lag}_k^R(\hat{\mathbf{z}}^m)) \leq \tfrac{4R\Psi^{-1}(\alpha)}{a\sqrt{m}}\right)$$
$$\geq \mathbb{P}\left(\tfrac{4R|(\mathbb{M}_1,\ldots,\mathbb{M}_N)|_\infty}{\delta_m} \leq \tfrac{4R\Psi^{-1}(\alpha)}{a\sqrt{m}}\right) = \mathbb{P}\left(\tfrac{a\sqrt{m}|(\mathbb{M}_1,\ldots,\mathbb{M}_N)|_\infty}{\delta_m} \leq \Psi^{-1}(\alpha)\right),$$

where, taking inferior limits, we obtain the following result.

Remark 3.3.7. *Set* $(\mathbb{M}_1, \ldots, \mathbb{M}_N) = \left(D^2\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*)\right)^{-1}((\mathbb{U}_1, \ldots, \mathbb{U}_N))$, *where* $(\mathbb{U}_1, \ldots, \mathbb{U}_N) \sim \mathcal{N}(\mathbf{0}, \Sigma(\mathbf{p}))$, *for $\Sigma(\mathbf{p})$, is defined in (3.7). Let Ψ^{-1} be the quantile function of* $|(\mathbb{M}_1, \ldots, \mathbb{M}_N)|_\infty$. *Under the assumptions of Theorem 3.3.3, if $\mathbf{u} + 2a\mathbb{B}_d \subset \mathrm{Lag}_k^R(\mathbf{z}^*)$, then*

$$\liminf_m \mathbb{P}\left(\mathrm{Lag}_k^R(\mathbf{z}^*) \subset \mathrm{Lag}_k^R(\hat{\mathbf{z}}^m) + \tfrac{4R\Psi^{-1}(\alpha)}{a\sqrt{m}}\mathbb{B}_d\right) \geq \alpha.$$

We notice that we need to approximate the distribution Ψ. In the previous section we justified the consistency of the parametric bootstrap to approximate the distribution of $(\mathbb{M}_1, \ldots, \mathbb{M}_N) = \left(D^2\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*)\right)^{-1}((\mathbb{U}_1, \ldots, \mathbb{U}_N))$. More precisely, let $\mathbf{Y}_1^B, \ldots, \mathbf{Y}_m^B$ be a bootstrap sample of i.i.d. (conditionally given $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$) with common law Q_m. We

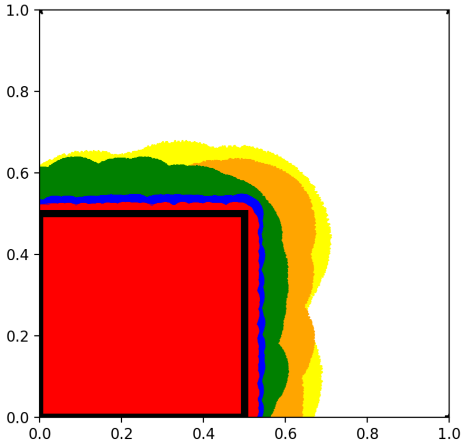

Figure 3.3.2: Estimated upper confidence intervals for the set $\mathrm{Lag}_1^R(\mathbf{z}^*)$, where $P = \frac{1}{4}(\delta_{(0,0)} + \delta_{(0,1)} + \delta_{(1,0)} + \delta_{(1,1)})$ is deterministic and Q_m is the empirical measure of $Q \sim \mathcal{U}_{(0,1)^2}$ for different values of m. Represented in yellow for $m = 100$; in orange for $m = 500$; in green for $m = 1000$, in blue for $m = 5000$ and in red for $m = 10000$. The black square is the border of the population cell $\mathrm{Lag}_1^R(\mathbf{z}^*)$.

assume that the empirical process $\sqrt{m}(Q_m^B - Q_m)$ of $\mathbf{Y}_1^B, \dots, \mathbf{Y}_m^B$ converges conditionally given $\mathbf{Y}_1, \dots, \mathbf{Y}_m$ in distribution to a tight random element $\mathbb{G}$. Lemma 3.9.34 and Theorem 3.9.11 in Vaart and Wellner (1996) give

$$\sup_{f \in BL(\mathbb{R}^N)} |\mathbb{E}(f(\sqrt{m}((\hat{\mathbf{z}}^m)^B - \mathbf{z}^*))|\mathbf{Y}_1, \dots, \mathbf{Y}_m) - f((\mathbb{M}_1, \dots, \mathbb{M}_N))| \xrightarrow{P} 0,$$

where the set $BL(\mathbb{R}^N)$ is the set of Bounded Lipschitz functions (see eg. p.73 in Vaart and Wellner (1996)) and $(\hat{\mathbf{z}}^m)^B$ is the solution of (3.3) for Q_m^B and P. Any function of the form $g \circ |\cdot|_\infty$, with $g \in BL(\mathbb{R})$, belongs to $BL(\mathbb{R}^N)$, so that

$$\sup_{g \in BL(\mathbb{R})} |\mathbb{E}(f(|\sqrt{m}((\hat{\mathbf{z}}^m)^B - \mathbf{z}^*|_\infty))|\mathbf{Y}_1, \dots, \mathbf{Y}_m) - f(|(\mathbb{M}_1, \dots, \mathbb{M}_N|\infty)| \xrightarrow{P} 0.$$

We can thus estimate the distribution of $|(\mathbb{M}_1, \dots, \mathbb{M}_N|\infty$ by means of the bootstrap sample.

We consider now a synthetic example to illustrate the use of Remark 3.3.7. Let Q be the uniform law on the unit square and $P = \frac{1}{4}(\delta_{(0,0)} + \delta_{(0,1)} + \delta_{(1,0)} + \delta_{(1,1)})$. We easily see that population cells are

$$\mathrm{Lag}_1(\mathbf{z}^*) = [0, \tfrac{1}{2}]^2,\ \mathrm{Lag}_2(\mathbf{z}^*) = [\tfrac{1}{2}, 1] \times [0, \tfrac{1}{2}],\ \mathrm{Lag}_3(\mathbf{z}^*) = [\tfrac{1}{2}, 1]^2,\ \mathrm{Lag}_4(\mathbf{z}^*) = [0, \tfrac{1}{2}] \times [\tfrac{1}{2}, 1].$$

We analyze the behavior of the first cell. In figure 3.3.2 we plot, for sample sizes $m = 100, 500, 1000, 5000, 10000$, the values of $\mathrm{Lag}_1^R(\hat{\mathbf{z}}^m) + \frac{4R(\Psi^{-1})^B(0.05)}{a\sqrt{m}}\mathbb{B}_d$. Here $(\Psi^{-1})^B$ denotes the quantile of the bootstrap approximation.

Remark 3.3.8. *The same technique can be applied to the Voronoi cells, i.e.*

$$\mathrm{Vor}_k(\mathbf{z}) := \{\mathbf{y} \in \mathbb{R}^d : \ |\mathbf{c}_k - \mathbf{y}|^2 < |\mathbf{c}_i - \mathbf{y}|^2, \ \textit{for all } i \neq k\}, \quad k = 1, \dots, N,$$

where $\{\mathbf{c}_i\}_{i=1}^N \in \arg\min_{\{\mathbf{a}_i\}_{i=1}^N \subset \mathbb{R}^N} \int \min_{0 \le i \le N} |\mathbf{a}_i - \mathbf{y}|^2 dQ(\mathbf{y})$ *is the solution of the* N*-means clustering. Indeed, under certain uniqueness conditions of the minimum, the weak limit of* $\sqrt{n}(\mathbf{c}_i^n - \mathbf{c}_i)$ *exists (see Pollard (1982)) and the rewriting of the cells as in (3.23) can be made, which yields the differentiability.*

3.4 Applications to Hotelling's location model

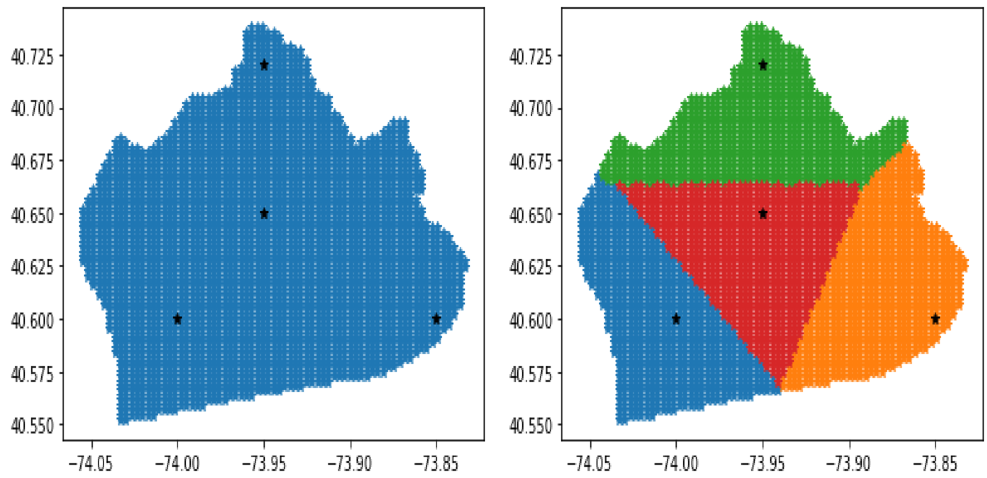

Figure 3.4.1: In the left original data of the sample $m = 3,129$ of Brooklyn's population distribution. On the Right, the empirical Laguerre cells $\mathrm{Lag}_i^R(\mathbf{z}^*)$ for the fountains located at $(-74.0, 40.6)$ (blue), at $(-73.85, 40.6)$ (orange), at $(-73.95, 40.72)$ (green) and at $(-73.95, 40.65)$ (red). Black points represent the fountains.

As mentioned in the introduction and driven by the application described in Galichon (2016), the equilibrium of the Hotelling's location model becomes

$$\sup_{\mathbf{z}} \sum_{i=1}^N z_i p_i + \int \inf_{i=1,\dots,N} \{|\mathbf{x}_i - \mathbf{y}|^2 - z_i\} dQ(\mathbf{y}), \tag{3.26}$$

where the distribution Q models the location of the population, the distribution P the fountains and z_i the prize of the fountain i. Denote $\mathbf{z}^*$ the solution of (3.26). The set of population that prefers to consume from the fountain i—called demand set—is $\mathrm{Lag}_i(\mathbf{z})$. In this section we apply the methodology developed in previous section to provide asymptotic confidence intervals for prices and for the demand sets. That means that assuming the observation of a sample $(\mathbf{Y}_1, \ldots, \mathbf{Y}_m)$ of the population Q, and computing the discrete-discrete optimal transport problem between the empirical measure Q_m and P_N, we obtain empirical prizes $\mathbf{z}^m$ and demand sets $\{\mathrm{Lag}_i(\mathbf{z}^m)\}_{i=1}^N$. Note that we obtain also the empirical solution of (3.26), which does not play an important in this problem. We assume that the

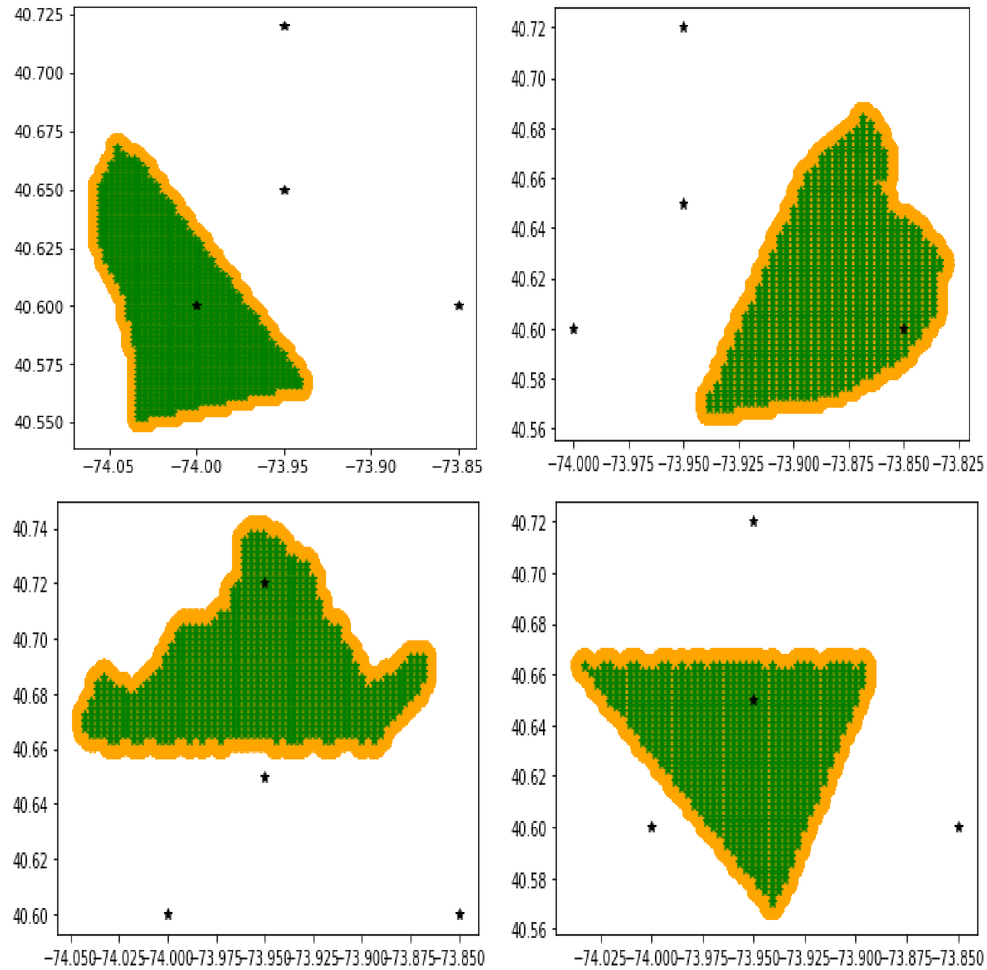

Figure 3.4.2: Estimated upper confidence intervals for the set $\mathrm{Lag}_i^R(\mathbf{z}^*)$. For $i = 1$, (upper left) the fountain is located at $(-74.0, 40.6)$; for $i = 2$, (upper right) at $(-73.85, 40.6)$; for $i = 3$, (lower left) at $(-73.95, 40.72)$ and, for $i = 3$, (lower left) at $(-73.9540.65)$. The empirical region (computed by solving the discrete-discrete optimal transport problem) is represented by blue points whereas the asymptotic confidence intervals for the Hausdorff distance are in yellow. Black points represent the fountains.

population Q satisfies the assumptions of Theorem 3.3.3, which asserts that, denoting Ψ^{-1} the quantile function of $|(\mathbb{M}_1, \ldots, \mathbb{M}_N)|_\infty$, then

$$\mathbb{P}\left(|\mathbf{z}^* - \mathbf{z}^m|_\infty > \frac{\Psi^{-1}(\alpha)}{\sqrt{m}}\right) \longrightarrow 1 - \alpha, \quad \alpha \in (0,1) \tag{3.27}$$

On the other hand, in view of Lemma 3.3.6, we have

$$\liminf_m \mathbb{P}\left(\mathrm{Lag}_k^R(\mathbf{z}^*) \subset \mathrm{Lag}_k^R(\hat{\mathbf{z}}^m) + \frac{4R\Psi^{-1}(\alpha)}{\delta_m\sqrt{m}}\mathbb{B}_d\right) \geq \alpha, \quad \alpha \in (0,1) \tag{3.28}$$

with

$$\delta_m = \sup_{\mathbf{y} \in \mathrm{Lag}_k^R(\mathbf{z}^m) \cap \mathrm{Lag}_k^R(\mathbf{z}^*)} \min\left(d_\infty(\{\mathbf{y}\}, \partial\,\mathrm{Lag}_k^R(\mathbf{z}^*)), d_\infty(\{\mathbf{y}\}, \partial\,\mathrm{Lag}_k^R(\mathbf{z}^m))\right).$$

Let us apply it to an artificial example based on real data. The population Q will be the demographic distribution of Brooklyn (NYC), which was $2,592,149$ at 2014[1] and can be modeled as a continuous probability. However, the data-set with spatial data around Brooklyn we found on internet is the *"New York City Census Data"*, which comes from the *American Community Survey 2015* and fully available on-line in *https://www.kaggle.com/datasets/muonneutrino/new-york-city-census-data*. Once cleaned, the data-set contains, a sample of size $m = 3,129$ of Brooklyn's population distribution. We suppose the existence of four different fountains, located at $(-74.0, 40.6)$, $(-73.85, 40.6)$, $(-73.95, 40.72)$ $(-73.95, 40.65])$ with same amount of stock. The data is displayed in Figure 3.4.1.

On the one hand, we compute the asymptotic confidence intervals for the norm infinity of the differences $|\mathbf{z}^* - \mathbf{z}^m|_\infty$ and for the individual variation of $z_i^* - \mathbf{z}_i^m$, $i = 1, 2, 3, 4$, of the prices. We obtain the following results $\mathbf{z}^* = \mathbf{z}^m + [0, 5.217 \cdot 10^{-04}]^4$, $z_1^* = z_1^m + 10^{-04}[-4.302, 4.242]$, $z_2^* = z_2^m + 10^{-04}[-5.260, 5.486]$, $z_3^* = z_3^m + 10^{-04}[-5.260, 5.486]$, $z_4^* = z_4^m + 10^{-04}[-1.814, 1.416]$.

where z_i^* represents, for $i = 1$ the prices of the fountain located at $(-74.0, 40.6)$; for $i = 2$, at $(-73.85, 40.6)$; for $i = 3$, at $(-73.95, 40.72)$ and, for $i = 3$, at $(-73.95, 40.65)$. The same notation is shared by z_i^m.

Previous result enable to obtain, in Figure 3.4.2, the asymptotic confidence intervals for the Hausdorff distance for each one of the demand sets by using the proposed approach. The value R is assumed to be 0.19 and δ_n is approximated by taking the maximum distance between points in the empirical cell.

[1] According to the United States Census Bureau *http://www.census.gov/quickfacts/*

Appendix

Contents

3..1 Simulations

First, we illustrate the precision of the upper bound in Theorem 3.1.6 with the following simulation. Consider the uniform measure on the unit interval $U(0,1)$ and draw a sample of size $m = 2000$ to obtain the empirical U_m. Then from a uniform discretization of size N of the unit interval, we obtain the discrete measure P^N. We compute, using Monte-Carlo simulations, the empirical error $E|\mathcal{W}_1(P^N, U_m) - \mathcal{W}_1(P^N, U)|$ for different choices for N. The results are presented in Figure 3..3. We observe, in the left figure, that, for regular values of N, the growth of $E|\mathcal{W}_1(P^N, U_m) - \mathcal{W}_1(P^N, U)|$ is exactly of order $\sqrt{N}$, following the bound. Yet for larger values of N (right side) we observe that the order is no longer $\sqrt{N}$. This is because $\sqrt{N}$ is only an upper bound for $E|\mathcal{W}_1(P^N, U_m) - \mathcal{W}_1(P^N, U)|$ and the true rate becomes smaller.

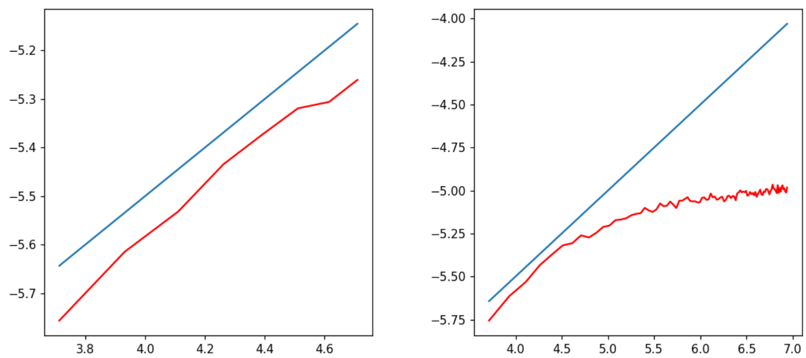

Figure 3..3: Plot, in double logarithmic scale, of $E|\mathcal{W}_1(P^N, U_m) - \mathcal{W}_1(P^N, U)|$ (y axis) with respect to N (x axis).

The next part of section is devoted to illustrate empirically Theorems 3.1.4 and 3.2.2. The limit distribution depends on the true optimal transport cost between the distributions. Hence to simulate the central limit theorems, the difficulty lies in proving the consistency of its bootstrap approximation. Actually the non fully Hadamard differentiability of the functional implies that the limit in Theorem 3.1.4 is the supremum of Gaussian processes. In consequence, as pointed out in Fang and Santos (2018), the bootstrap will not be consistent. However, in the framework of Theorem 3.2.2, the dual problem has a unique solution. In consequence, the mapping is fully Hadamard differentiable (Corollary 2.4 in Cárcamo et al. (2020)) which implies that the bootstrap procedure is consistent (Fang and Santos (2018)). This enables us to approximate the variance as shown in the following simulations.

Here we implement one favorable case for bootstrap approximation. In particular we choose the quadratic cost $|\cdot|^2$ and the discrete probability $P = \frac{1}{7}\sum_{i=1}^{7} \delta_{\mathbf{x}_i}$, where

$$\mathbb{X} = \{\mathbf{x}_i\}_{i=1}^{7} = \{(1,0,0),(0,1,0),(0,0,1),(-1,0,0),(0,-1,0),(0,0,-1),(0,0,0)\}.$$

The continuous probability $Q \in \mathcal{P}(\mathbb{R}^3)$ is the direct product $\mathcal{U}(-1,1)\times\mathcal{N}(0,1)\times\mathcal{N}(0,1)$. Note that its support is connected with Lebesgue negligible boundary—we can visualize the data in Figure 3..4—and satisfies the assumptions of Theorem 3.2.2. As commented before, we can use the bootstrap procedure. In this example, it is assumed that the discrete P is known and the sample, of size $m = 5000$, comes from the continuous Q. Figure 3..5 shows the result of the bootstrap procedure for a re-sampling size of 10000. The simulations follow the asymptotic theory we provide.

Now we illustrate a case where the assumptions of Theorem 3.2.2 are no longer fulfilled. More precisely, we consider Q as the continuous probability with density $\frac{1}{0.008\cdot 7}\sum_{i=1}^{7} \mathbb{1}_{\mathbf{x}_i+(-0.1,0.1)^3}$, —this is a mixture model of uniform probabilities on small cubes centered in the points of $\mathbb{X}$—we can see a 3d plot in Figure 3..4. To approximate the limit distribution we need first to estimate the value $\mathcal{W}_2^2(P,Q)$. We make it by an independent sample of size 10000 and computing the mean by Monte Carlo 100 times. Then we compute the histogram of $\sqrt{m}\frac{\left(\mathcal{W}_2^2(P,Q_m)-\mathcal{W}_2^2(P,Q)\right)}{\sigma_2(Q_m,\mathbf{z}^m)}$ with the original sample. The results are shown in Figure 3..6, we can see, clearly, that the limit is not Gaussian. Similar examples with non-Gaussian limits can be found in Figure 1 in Sommerfeld and Munk (2018). But Figure 3..6 is quite different from their experimentation since one of the probabilities is continuous and Sommerfeld and Munk (2018) studies only the optimal transport problem between discrete probabilities.

3..2 Proofs

Proof of Theorem 3.1.4. The strategy of the proof is the following, first we start by proving the central limit theorem for bounded potentials. That means the study of the asymptotic behaviour of the sequence

$$\sqrt{\frac{nm}{n+m}}\left(\sup_{|\mathbf{z}|\leq K} g_c(P_n,Q_m,\mathbf{z}) - \sup_{|\mathbf{z}|\leq K} g_c(P,Q,\mathbf{z})\right)_{n,m}.$$

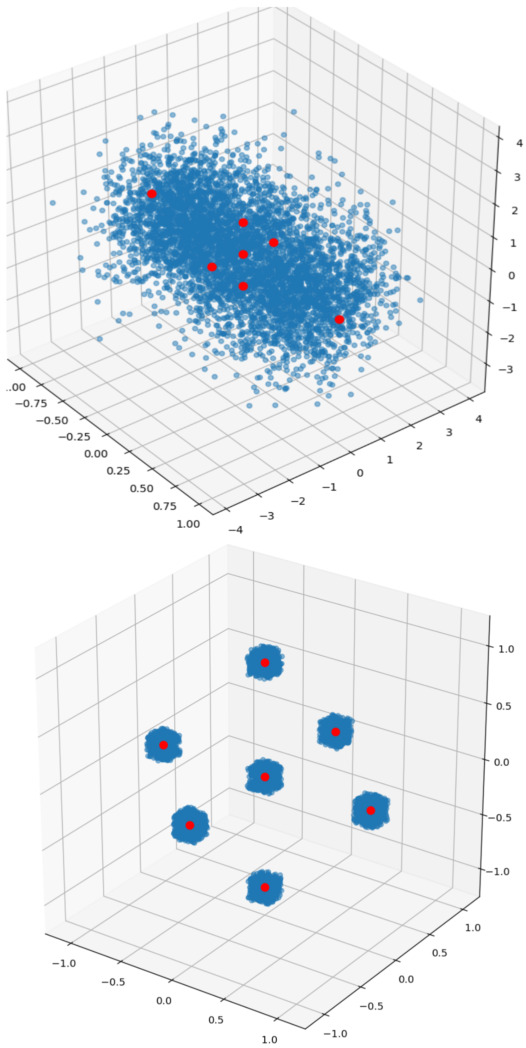

Figure 3..4: 3D visualization of the data set. In blue the continuous distribution Q and in red the discrete one P. Left: Q has a density with connected support and satisfies the assumptions of Theorem 3.2.2. Right: Q has a density but not a connected support.

The weak limit depends on the set of restricted optimal points.

$$\mathrm{Opt}_c^K(P,Q) := \left\{ \mathbf{z} : \sup_{|\mathbf{s}|\leq K} g_c(P,Q,\mathbf{s}) = g_c(P,Q,\mathbf{z}), \;\; z_1 = 0 \right\}.$$

Lemma 3..1. *Set $K > 0$, under the assumptions of Theorem 3.1.4, we have the limit*

$$\sqrt{\frac{nm}{n+m}} \left(\sup_{|\mathbf{z}|\leq K} g_c(P_n, Q_m, \mathbf{z}) - \sup_{|\mathbf{z}|\leq K} g_c(P,Q,\mathbf{z}) \right) \xrightarrow{w} \sup_{\mathbf{z}\in \mathrm{Opt}_c^K(P,Q)} \left(\sqrt{\lambda} \sum_{i=1}^{N} z_i \mathbb{U}_i + (\sqrt{1-\lambda}) \mathbb{G}_Q^c(\mathbf{z}) \right),$$

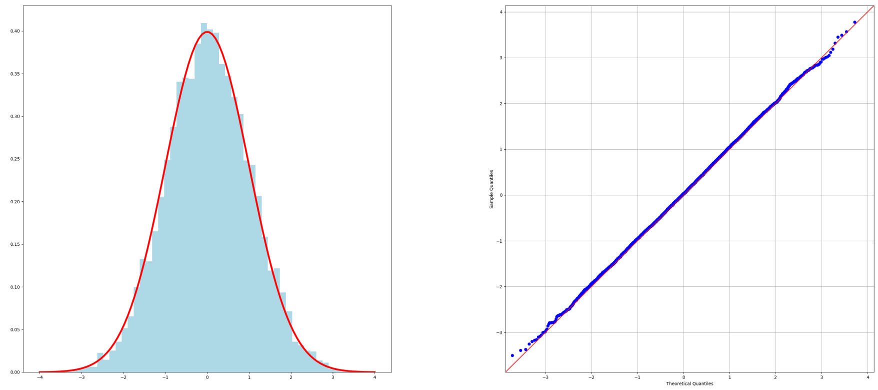

Figure 3..5: Illustration of Theorem 3.2.2 using bootstrap procedures. Histograms (left) and Q–Q plot (right) of the bootstrap estimation of $\sqrt{m}\frac{\left(\mathcal{W}_2^2(P,Q_m)-\mathcal{W}_2^2(P,Q)\right)}{\sigma_2(Q_m,\mathbf{z}^m)}$.

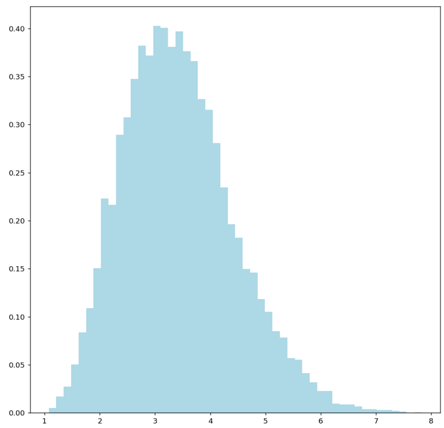

Figure 3..6: Illustration of Theorem 3.1.4, by using Monte Carlo's method, for Q with disconnected support.

with $(\mathbb{U}_1,\ldots,\mathbb{U}_N)$ *and* $\mathbb{G}_Q^c$ *as in Theorem 3.1.4.*

Proof of Lemma 3..1. For each $K>0$ we define the restricted set

$$\mathcal{F}_c^K=\left\{\mathbf{y}\mapsto\inf_{i=1,\ldots,N}\{c(\mathbf{x}_i,\mathbf{y})-z_i\},\ \ \mathbf{x}_i\in\mathbb{X},\ \ \text{and}\ \ z_1=0,\ |\mathbf{z}|\leq K\right\},$$

Lemma 3..7 proves that such a class is Q-Donsker, see Theorem 1.5.7 in Vaart and Wellner

(1996), in the sense that

$$\sqrt{m}(Q_m - Q) \xrightarrow{w} \mathbb{G}_Q \text{ in } \ell^\infty(\mathcal{F}_c^K),$$

where $\mathbb{G}_Q$ is the Brownian bridge in $\mathcal{F}_c^K$. This is a centered Gaussian process with covariance function

$$(f,g) \mapsto \int f(\mathbf{y})g(\mathbf{y})dQ(\mathbf{y}) - \int f(\mathbf{y})dQ(\mathbf{y}) \int g(\mathbf{y})dQ(\mathbf{y}).$$

Let $\bar{\mathbb{B}}_K(\mathbf{0})$ be the closure of the centered ball of radius K in $\mathbb{R}^N$. Note that the functional

$$\begin{aligned} C : \ell^\infty(\mathcal{F}_c^K) &\longrightarrow \ell^\infty(\bar{\mathbb{B}}_K(\mathbf{0})) \\ f &\mapsto \left(\mathbf{z} \mapsto f\left(\inf_{i=1,\dots,N}\{c(\mathbf{x}_i,\mathbf{y}) - z_i\}\right)\right) \end{aligned}$$

is actually continuous, hence for any $f, g \in \ell^\infty(\mathcal{F}_c^K)$, we have

$$\sup_{\mathbf{z}\in\mathbb{B}_K(\mathbf{0})} \left| f\left(\inf_{i=1,\dots,N}\{c(\mathbf{x}_i,\mathbf{y}) - z_i\}\right) - g\left(\inf_{i=1,\dots,N}\{c(\mathbf{x}_i,\mathbf{y}) - z_i\}\right)\right| = \sup_{\phi\in\ell^\infty(\mathcal{F}_c^K)} |f(\phi) - g(\phi)|.$$

Moreover, the multivariate CLT implies $\sqrt{n}\,(\mathbf{p}_n - \mathbf{p}) \xrightarrow{w} (\mathbb{U}_1,\dots,\mathbb{U}_N) \sim N\left(\mathbf{0},\Sigma(\mathbf{p})\right)$, where $\Sigma(\mathbf{p})$ is defined in (3.7). Since the sequences $\sqrt{n}(\mathbf{p}_n - \mathbf{p})$ and $\sqrt{m}(Q_m - Q)$ are independent we derive the following result.

Lemma 3..2. *Under the assumptions of Theorem 3.1.4, we have the limit*

$$\sqrt{\frac{nm}{n+m}}\,(g_c(P_n,Q_m,\cdot) - g_c(P,Q,\cdot)) \xrightarrow{w} \sqrt{\lambda}\langle\mathbf{X},\cdot\rangle + \sqrt{1-\lambda}C(\mathbb{G}_Q) \;\; \textit{in} \;\; \ell^\infty(\bar{\mathbb{B}}_K(\mathbf{0})), \tag{3.29}$$

with $(\mathbb{U}_1,\dots,\mathbb{U}_N) = \mathbf{X}$.

Let $(\mathcal{B}, d)$ be a compact metric space, Corollary 2.3 in Cárcamo et al. (2020), provides the directional Hadamard derivative of the functional

$$\begin{aligned} \delta : \ell^\infty(\mathcal{B}) &\longrightarrow \mathbb{R} \\ F &\mapsto \delta(F) = \sup_{\mathbf{z}\in\mathcal{B}} F(\mathbf{z}), \end{aligned}$$

tangentially to $\mathcal{C}(\mathcal{B})$ (the space of continuous functions from $\mathcal{B}$ to $\mathbb{R}$) with respect to F in a direction $G \in \mathcal{C}(\mathcal{B})$. Recall that a function $f : \Theta \to \mathbb{R}$, defined in a Banach space, Θ, is said to be *Hadamard directionally differentiable* at $\theta \in \Theta$ tangentially to $\Theta_0 \subset \Theta$ if there exists a function $f'_\theta : \Theta_0 \to \mathbb{R}$ such that

$$\frac{f(\theta + t_n h_n) - f(\theta)}{t_n} \xrightarrow[n\to\infty]{} f'_\theta(h), \quad \text{for all sequences } t_n \searrow 0 \text{ and } h_n \to h, \text{ for all } h \in \Theta_0.$$

If $F \in \mathcal{C}(\mathcal{B})$ is not identically 0, the precise formula for the derivative, provided by Corollary 2.3 in Cárcamo et al. (2020), is

$$\delta'_F(G) = \sup_{\{\mathbf{z}:\ F(\mathbf{z})=\delta(F)\}} G(\mathbf{z}), \quad \text{for } G \in \mathcal{C}(\mathcal{B}). \tag{3.30}$$

In our case the compact metric space is the ball $\bar{\mathbb{B}}_K(\mathbf{0})$, the functional F correspond with $g_c(P, Q, \cdot)$ and the set of optimal points is $\mathrm{Opt}_c^K(P, Q)$. The following result rewrites (3.30) in our setting.

Lemma 3..3. *Set $K > 0$, under the assumptions of Theorem 3.1.4, the map δ is Hadamard directionally differentiable at $g_c(P, Q, \cdot)$, tangentially to the set $\mathcal{C}(\bar{\mathbb{B}}_K(\mathbf{0}))$ with derivative, for $G \in \mathcal{C}(\bar{\mathbb{B}}_K(\mathbf{0}))$,*

$$\delta'_{g_c(P,Q,\cdot)}(G) = \sup_{\mathbf{z}\in \mathrm{Opt}_c^K(P,Q)} G(\mathbf{z}).$$

The last step is the application of the delta-method. Let Θ be a Banach space, $\theta \in \Theta$ and $\{Z_n\}_{n\in\mathbb{N}}$ be a sequence of random variables such that $Z_n : \Omega_n \to \Theta$ and $r_n(Z_n - \theta) \xrightarrow{w} Z$ for some sequence $r_n \to +\infty$ and some random element Z that takes values in $\Theta_0 \subset \Theta$. If $f : \Theta \to \mathbb{R}$ is Hadamard differentiable at θ tangentially to $\Theta_0 \subset \Theta$, with derivative $f'_\theta(\cdot) : \Omega_0 \to \mathbb{R}$, then Theorem 1 in Römisch (2014), so-called delta-method, states that $r_n(f(Z_n) - f(\theta)) \xrightarrow{w} f'_\theta(Z)$.

Now, it only remains to prove that the limit in (3.29) belongs to $\mathcal{C}(\bar{\mathbb{B}}_K(\mathbf{0}))$. Such a limit is a mixture of two independent processes. The first one $\langle \mathbf{X}, \cdot\rangle$ has clearly continuous sample paths with respect to the euclidean norm $|\cdot|$ in $\mathbb{R}^N$. On the other side, $\mathbb{G}_Q$ has continuous sample paths in $\mathcal{F}_c^K$ with respect to the semi-metric

$$\rho_Q(f) = \int f(\mathbf{y})^2 dQ(\mathbf{y}) - \left(\int f(\mathbf{y}) dQ(\mathbf{y})\right)^2,$$

in the sense that, see pag 89 in Vaart and Wellner (1996), there exists some sequence $\delta_n \searrow 0$ such that

$$\sup_{f,g\in\mathcal{F}_c^K,\ \rho_Q(f,g)<\delta_n} |\mathbb{G}_Q(f) - \mathbb{G}_Q(g)| \xrightarrow{a.s.} 0. \tag{3.31}$$

We want now to analyse the value $\sup_{|\mathbf{z}-\mathbf{s}|<\delta_n} |C(\mathbb{G}_Q)(\mathbf{z}) - C(\mathbb{G}_Q)(\mathbf{s})|$ Note that for every $f \in \mathcal{F}_c^K$ there exists some $\mathbf{z}^f \in \bar{\mathbb{B}}_K(\mathbf{0})$ such that $f(\mathbf{y}) = \inf_{i=1,\dots,m}\{c(\mathbf{x}_i, \mathbf{y}) - z_i^f\}$. Lemma 3.1.3 states that

$$\{f, g \in \mathcal{F}_c^K : |\mathbf{z}^f - \mathbf{z}^g| < \delta\} \subset \{f, g \in \mathcal{F}_c^K : ||f - g||_\infty < \delta\} \subset \{f, g \in \mathcal{F}_c^K : \rho_Q(f, g) < \delta\}. \tag{3.32}$$

Since $|C(\mathbb{G}_Q)(\mathbf{z}^f) - C(\mathbb{G}_Q)(\mathbf{z}^g)| = |\mathbb{G}_Q(f) - \mathbb{G}_Q(g)|$, then we have

$$\sup_{|\mathbf{z}^f-\mathbf{z}^g|<\delta_n} |C(\mathbb{G}_Q)(\mathbf{z}^f) - C(\mathbb{G}_Q)(\mathbf{z}^g)| = \sup_{|\mathbf{z}^f-\mathbf{z}^g|<\delta_n} |\mathbb{G}_Q(f) - \mathbb{G}_Q(g)|,$$

and, consequently, using (3.32) and (3.31), we obtain

$$\sup_{|\mathbf{z}^f-\mathbf{z}^g|<\delta_n} |C(\mathbb{G}_Q)(\mathbf{z}^f)-C(\mathbb{G}_Q)(\mathbf{z}^g)| \leq \sup_{\rho_Q(f,g)<\delta_n} |\mathbb{G}_Q(f)-\mathbb{G}_Q(g)| \xrightarrow{a.s.} 0.$$

Finally, Lemma 3..2 implies that $\sqrt{\frac{nm}{n+m}}\,(g_c(P_n,Q_m,\cdot)-g_c(P,Q,\cdot))$ has a weak limit Z in $\ell^\infty(\bar{\mathbb{B}}_K(\mathbf{0}))$ having a version in $\mathcal{C}(\bar{\mathbb{B}}_K(\mathbf{0}))$. Applying the so-called delta-method to the function δ and Lemma 3..3 we derive the limit

$$\sqrt{\frac{nm}{n+m}}\left(\sup_{|\mathbf{z}|\leq K} g_c(P_n,Q_m,\mathbf{z}) - \sup_{|\mathbf{z}|\leq K} g_c(P,Q,\mathbf{z})\right) \xrightarrow{w} \sup_{\mathbf{z}\in \mathrm{Opt}_c^K(P,Q)} Z(\mathbf{z}).$$

Note, that the process $\mathbf{z}\mapsto \sqrt{1-\lambda}\,\mathbb{G}_Q\left(\inf_{i=1,\dots,N}\{c(\mathbf{x}_i,\mathbf{y})-z_i\}\right)$ is Gaussian in $\mathbb{R}^N$ with covariance function Ξ_Q^c. Moreover, it is independent from $\mathbf{X}$, then the law of the process Z is the same of the process $\sqrt{\lambda}\langle \mathbf{X},\cdot\rangle + (\sqrt{1-\lambda})\mathbb{G}_Q^c$ and the theorem holds. □

Unfortunately, the optimal solutions need not be universally bounded. In order to go from the bounded to the unbounded, we observe that Lemma 3.1.1 implies

$$\mathcal{T}_c(P_n,Q_m) = \sup_{\mathbf{z}\in\mathbb{R}^N} g_c(P_n,Q_m,\mathbf{z}) = \sup_{|\mathbf{z}|\leq K_{n,m}} g_c(P_n,Q_m,\mathbf{z}),$$

for $K_{n,m} = \frac{1}{\inf_i p_i^n}\left(\sup_{i=1,\dots,N}\int c(\mathbf{y},\mathbf{x}_i)dQ_m(\mathbf{y})\right)$. Let K^* be the constant provided in Lemma 3.1.1 for P and Q (that means $K^* = \frac{1}{\inf_i p_i}\left(\sup_{i=1,\dots,N}\int c(\mathbf{y},\mathbf{x}_i)dQ(\mathbf{y})\right)$). The strong law of large numbers implies the a.s. convergence of $\inf_i p_i^n$ to $\inf_i p_i$, and, assuming (3.10), we have that the sequence $\sqrt{m}\,(K_{n,m}-K^*)$ is stochastically bounded. Finally, the difference $\sqrt{\frac{nm}{n+m}}\,(\mathcal{T}_c(P_n,Q_m)-\mathcal{T}_c(P,Q))$ is equal to

$$\sqrt{\frac{nm}{n+m}}\left(\mathcal{T}_c(P_n,Q_m) - \sup_{|\mathbf{z}|\leq K^*+1} g_c(P_n,Q_m,\mathbf{z})\right) + \sqrt{\frac{nm}{n+m}}\left(\sup_{|\mathbf{z}|\leq K^*+1} g_c(P_n,Q_m,\mathbf{z}) - \mathcal{T}_c(P,Q)\right), \tag{3.33}$$

and Lemma 3..1 implies the weak convergence of the second term to

$$\sup_{\mathbf{z}\in \mathrm{Opt}_c^{K^*+1}(P,Q)}\left(\sqrt{\lambda}\sum_{i=1}^N z_i\mathbb{U}_i + (\sqrt{1-\lambda})\mathbb{G}_Q^c(\mathbf{z})\right) = \sup_{\mathbf{z}\in \mathrm{Opt}_c^{0}(P,Q)}\left(\sqrt{\lambda}\sum_{i=1}^N z_i\mathbb{U}_i + (\sqrt{1-\lambda})\mathbb{G}_Q^c(\mathbf{z})\right),$$

where the equality is a direct consequence of Lemma 3.1.1. It only remains to prove that the first term of (3.33) tends to 0 in probability. Note that we have two cases.

- The first one is $K_{n,m}\leq K^*+1$, which implies that

  $$\mathcal{T}_c(P_n,Q_m) = \sup_{|\mathbf{z}|\leq K_{n,m}} g_c(P_n,Q_m,\mathbf{z}) = \sup_{|\mathbf{z}|\leq K^*+1} g_c(P_n,Q_m,\mathbf{z}) \leq \mathcal{T}_c(P_n,Q_m),$$

  and makes 0 the first term of (3.33).

- The second one is $K_{n,m} \geq K^* + 1$, which implies the bound

$$0 \leq \mathcal{T}_c(P_n, Q_m) - \sup_{|\mathbf{z}| \leq K^*+1} g_c(P_n, Q_m, \mathbf{z}) \leq \sup_{|\mathbf{z}| \leq K_{n,m}} g_c(P_n, Q_m, \mathbf{z}) - \sup_{|\mathbf{z}| \leq K^*+1} g_c(P_n, Q_m, \mathbf{z}). \tag{3.34}$$

Note that the right side of the inequality (3.34) can be rewritten as

$$\sup_{|\mathbf{z}| \leq K_{n,m}} g_c(P_n, Q_m, \mathbf{z}) - \sup_{|\mathbf{z}| \leq K^*+1} g_c(P_n, Q_m, \mathbf{z}) = \sup_{|\mathbf{z}| \leq K_{n,m}} \inf_{|\mathbf{z}'| \leq K^*+1} g_c(P_n, Q_m, \mathbf{z}) - g_c(P_n, Q_m, \mathbf{z}')$$

and upper bounded by

$$\sup_{|\mathbf{z}| \leq K_{n,m}} \inf_{|\mathbf{z}'| \leq K^*+1} |g_c(P_n, Q_m, \mathbf{z}) - g_c(P_n, Q_m, \mathbf{z}')|.$$

Since

$$|g_c(P_n, Q_m, \mathbf{z}) - g_c(P_n, Q_m, \mathbf{z}')| \leq \sum_{i=1}^{N} |z_i - z_i'| p_i^n + \int | \inf_{i=1,\dots,N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\} - \inf_{i=1,\dots,N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i'\} | dQ_m(\mathbf{y}),$$

we can conclude from (3.6) that

$$0 \leq \mathcal{T}_c(P_n, Q_m) - \sup_{|\mathbf{z}| \leq K^*+1} g_c(P_n, Q_m, \mathbf{z}) \leq 2 \sup_{|\mathbf{z}| \leq K_{n,m}} \inf_{|\mathbf{z}| \leq K^*+1} |\mathbf{z} - \mathbf{z}'| \leq |K_{n,m} - K^* - 1|.$$

Both cases together yield the inequality

$$0 \leq \mathcal{T}_c(P_n, Q_m) - \sup_{|\mathbf{z}| \leq K^*+1} g_c(P_n, Q_m, \mathbf{z}) \leq |K_{n,m} - K^* - 1| \mathbb{1}_{(K_{n,m} \geq K^*+1)}. \tag{3.35}$$

To see that the $\sqrt{\frac{nm}{n+m}} |K_{n,m} - K^* - 1| \mathbb{1}_{(K_{n,m} \geq K^*+1)}$ tends to 0 in probability, we write $|K_{n,m} - K^* - 1| \mathbb{1}_{(K_{n,m} \geq K^*+1)} = \max(0, K_{n,m} - K^* - 1)$. Note that

$$\sqrt{\frac{nm}{n+m}} \max(0, K_{n,m} - K^* - 1) = \max(0, \sqrt{\frac{nm}{n+m}} (K_{n,m} - K^* - 1)).$$

Since $\sqrt{\frac{nm}{n+m}} (K_{n,m} - K^*)$ is stochastically bounded implies that $\sqrt{\frac{nm}{n+m}} (K_{n,m} - K^* - 1)$ converges to $-\infty$ in probability and

$$\sqrt{\frac{nm}{n+m}} |K_{n,m} - K^* - 1| \mathbb{1}_{(K_{n,m} \geq K^*+1)} = \max(0, \sqrt{\frac{nm}{n+m}} (K_{n,m} - K^* - 1)) \xrightarrow{P} 0.$$

That proves Theorem 3.1.4.

Remark 3..4. *When dealing with the case where the asymptotics depend only on the empirical distribution P_n, note that Assumption (3.9), which depends only on Q*

$$\int c(\mathbf{x}_i, \mathbf{y}) dQ(\mathbf{y}) < \infty, \ \textit{for all } i = 1, \dots, m,$$

is enough to prove the CLT. Actually, the multidimensional CLT yields that

$$\sqrt{n}\,(g_c(P_n, Q, \cdot) - g_c(P, Q, \cdot)) \xrightarrow{w} \langle \mathbf{X}, \cdot \rangle \ \textit{in} \ \ell^\infty(\bar{\mathbb{B}}_K(\mathbf{0})),$$

with $(\mathbb{U}_1, \dots, \mathbb{U}_N) = \mathbf{X}$. Therefore, all of the arguments above can be now repeated verbatim.

□

Proof of Theorem 3.1.6. Let $\mathbf{Y}_1, \dots, \mathbf{Y}_m$ be i.i.d with law Q. Recall that, when the cost c is the euclidean distance $|\cdot|$, then the optimal transport potentials are 1-Lipschitz functions. This yields trivially

$$\mathcal{W}_1(P, Q) = \sup_{\mathbf{z} \in \mathbb{R}^N} g_1(P, Q, \mathbf{z}) = \sup_{|\mathbf{z}| \le \operatorname{diam}(\mathbb{X})} g_1(P, Q, \mathbf{z}), \tag{3.36}$$

where $g_1(P, Q, \mathbf{z}) = \sum_{i=1}^N z_i p_i + \int \inf_{i=1,\dots,N}\{|\mathbf{x}_i - \mathbf{y}| - z_i\} dQ(\mathbf{y})$. We want to bound the quantity $E\,|\mathcal{W}_1(P, Q_m) - \mathcal{W}_1(P, Q)|$, which can be rewritten, by (3.36), as

$$E\left| \sup_{|\mathbf{z}| \le \operatorname{diam}(\mathbb{X})} g_1(P, Q, \mathbf{z}) - \sup_{|\mathbf{z}| \le \operatorname{diam}(\mathbb{X})} g_1(P, Q_m, \mathbf{z}) \right|,$$

and upper bounded by $E\left|\sup_{f \in \mathcal{F}_1} \int f(\mathbf{y})(dQ_m(\mathbf{y}) - dQ(\mathbf{y}))\right|$, where

$$\mathcal{F}_1 = \left\{ \mathbf{y} \mapsto \inf_{i=1,\dots,N} \{|\mathbf{x}_i - \mathbf{y}| - z_i\}, \ \ \mathbf{x}_i \in \mathbb{X}, \ \text{ and } \ |\mathbf{z}| \le \operatorname{diam}(\mathbb{X}) \right\}.$$

We set $D = \operatorname{diam}(\mathbb{X})$ in order to simplify the following formulas. Denote by $N(\epsilon, \mathcal{F}_1, ||\cdot||_{L^2(Q_m)})$ the he covering number with respect to the metric $L^2(Q_m)$. Lemma 4.14 in Massart and Picard (2007) and Lemma 3.1.3 imply that

$$\sqrt{\log\left(2N(\epsilon, \mathcal{F}_1, ||\cdot||_{L^2(Q_m)})\right)} \le \sqrt{N \log\left(\frac{2D}{\epsilon} + 1\right) + \log(2)}.$$

Let denote as $a_{N,m}$ the (random) quantity $a_{N,m} = 2\sqrt{\int \sup_{i=1,\dots,N} |\mathbf{x}_i - \mathbf{y}|^2 dQ_m(\mathbf{y})}$, then

$$\int_0^{2D + a_{N,m}} \sqrt{\log\left(2N(\epsilon, \mathcal{F}_1, ||\cdot||_{L^2(Q_m)})\right)} d\epsilon \le \sqrt{N} \int_0^{2D + a_{N,m}} \sqrt{\log\left(\frac{2D}{\epsilon} + 1\right)} d\epsilon + 2D + a_{N,m} \log(2). \tag{3.37}$$

Now recall by Theorem 3.5.1 in Giné and Nickl (2015) that

$$\begin{aligned}
&\sqrt{m}E\left|\sup_{f\in\mathcal{F}_1}\int f(\mathbf{y})(dQ_m(\mathbf{y})-dQ(\mathbf{y}))\right| \\
&\leq 8\sqrt{2}E\left|\int_0^{2D+a_{N,m}}\sqrt{\log\left(2N(\epsilon,\mathcal{F}_1,||\cdot||_{L^2(Q_m)})\right)}d\epsilon\right| \\
&\leq 8\sqrt{2}\sqrt{N}E\int_0^{2D+a_{N,m}}\sqrt{\log\left(\frac{2D}{\epsilon}+1\right)}d\epsilon+8\sqrt{2}E(2D+a_{N,m})\log(2),
\end{aligned}$$

where the last inequality is consequence of (3.37). Since, using Jensen's inequality,

$$\begin{aligned}
&\int_0^{2D+a_{N,m}}\sqrt{\log\left(\frac{2D}{\epsilon}+1\right)}d\epsilon \\
&\leq (2D+a_{N,m})\sqrt{\frac{2D\log(2D+a_{N,m})+(4D+a_{N,m})\log\left(\frac{4D+a_{N,m}}{2D+a_{N,m}}\right)}{2D+a_{N,m}}} \\
&\leq\sqrt{2D(2D+a_{N,m})\log(2D+a_{N,m})+(4D+a_{N,m})(2D+a_{N,m})\log\left(\frac{4D+a_{N,m}}{2D+a_{N,m}}\right)}.
\end{aligned}$$

The mean value theorem yields $\log\left(\frac{4D+a_{N,m}}{2D+a_{N,m}}\right)\leq\frac{2D}{2D}=1$, and in consequence the following inequality

$$\int_0^{2D+a_{N,m}}\sqrt{\log\left(\frac{2D}{\epsilon}+1\right)}d\epsilon\leq(4D+a_{N,m})\sqrt{2D+1}.$$

Finally we can derive the bound

$$\sqrt{m}E\left|\sup_{f\in\mathcal{F}_c^D}\int f(\mathbf{y})(dQ_m(\mathbf{y})-dQ(\mathbf{y}))\right|\leq 8\sqrt{2}\sqrt{N}E\{(4D+a_{N,m})\left(\log(2)+\sqrt{2D+1}\right)\}.$$

Since, using triangle inequality, we have

$$Ea_{N,m}=E\left(2\sqrt{\int\sup_{i=1,\dots,N}|\mathbf{y}-\mathbf{x}_i|^2dQ_m(\mathbf{y})}\right)\leq 2\sqrt{\int|\mathbf{y}|^2dQ(\mathbf{y})}+2D,$$

which proves the result.

For the generalization to further potential costs, we observe that equation (2.5) in del Barrio et al. (2021) yields

$$|z_i^*|\leq|\mathbf{x}_0-\mathbf{x}_i|4\,\mathrm{diam}(\mathcal{Y})^{p-1}$$

and therefore, repeating the previous argumentation, we obtain the result.

□

Proof of Theorem 3.3.1.
We note that the population potential is described through (3.16), i.e. $\nabla_{\mathbf{z}}\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*)=0$, and the empirical by $\nabla_{\mathbf{z}}\mathcal{M}_{\mathbf{p}}^{n,m}(\mathbf{z}_{n,m})=0$, with

$$\nabla_{\mathbf{z}}\mathcal{M}_{\mathbf{p}}^{n,m}(\mathbf{z})=(-Q_m(\mathrm{Lag}_1(\mathbf{z}))+p_1^n,\dots,-Q_m(\mathrm{Lag}_N(\mathbf{z}))+p_N^n), \tag{3.38}$$

where Lag_k, for $k=1,\dots,N$, are defined in (3.17). This is a Z-estimation problem (eg. chapter 3.3 in Vaart and Wellner (1996)). The strategy is the following; we show that the population optimal transport potential is well-separated—meaning that $\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*)-\mathcal{M}_{\mathbf{p}}(\hat{\mathbf{z}}^{n,m})\geq c|\hat{\mathbf{z}}^{n,m}-\mathbf{z}^*|^2$—by deriving the second order derivative of $\mathcal{M}_{\mathbf{p}}$, then we show that the map $\mathcal{M}_{\mathbf{p}}-\mathcal{M}_{\mathbf{p}}^{n,m}$ is $\sqrt{\frac{n+m}{nm}}$–Lipschitz. As a consequence we obtain the tightness of $\sqrt{\frac{nm}{n+m}}|\hat{\mathbf{z}}^{n,m}-\mathbf{z}^*|$ and, by using second derivative of $\mathcal{M}_{\mathbf{p}}$, also its limit.

Lemma 3..5. *Let $\mathcal{Y}\subset\mathcal{R}$ be a compact c-convex set, $P\in\mathcal{P}(\mathbb{X})$ and $Q\in\mathcal{P}(\mathcal{Y})$. Under assumptions (Reg), (Twist) and (QC) on the cost c and (PW) and (Cont) on Q, we have that the function $\mathcal{M}_{\mathbf{p}}$ is strictly concave and twice continuously differentiable, with*

$$\nabla\mathcal{M}_{\mathbf{p}}(\mathbf{z})=\nabla_{\mathbf{z}}\mathcal{M}_{\mathbf{p}}(\mathbf{z})|_{\langle\mathbf{1}\rangle^\perp},$$
$$D^2\mathcal{M}_{\mathbf{p}}(\mathbf{z})=D^2_{\mathbf{z}^*}\mathcal{M}_{\mathbf{p}}(\mathbf{z})|_{\langle\mathbf{1}\rangle^\perp}.$$

Moreover if $\mathbf{z}^*\in\langle\mathbf{1}\rangle^\perp\cap\mathrm{Opt}_c(P,Q)$, then there exists a positive constant C such that

$$\langle D^2\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*)\mathbf{v},\mathbf{v}\rangle\leq -C\inf_{i=1,\dots,N}|p_i|^3|\mathbf{v}|^2,\ \textit{for all}\ \mathbf{v}\in\langle\mathbf{1}\rangle^\perp. \tag{3.39}$$

Proof. Note that it only remains to prove that (3.39) holds. But this is a direct consequence of (3.19). In fact, since $\mathbf{z}^*$ is the unique $\mathbf{z}\in\mathrm{Opt}(P,Q)$, then $Q(\mathrm{Lag}_k(\mathbf{z}^*))=p_k$ for $k=1,\dots,N$ and we can conclude. □

We want now to prove the tightness of $\sqrt{\frac{nm}{n+m}}|\hat{\mathbf{z}}^{n,m}-\mathbf{z}^*|$. Note that Lemma 3..5 implies the well-separability property

$$\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*)-\mathcal{M}_{\mathbf{p}}(\hat{\mathbf{z}}^{n,m})\geq c|\hat{\mathbf{z}}^{n,m}-\mathbf{z}^*|^2,\quad \text{for some } c>0. \tag{3.40}$$

The relations

$$\begin{aligned}&|\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*)-\mathcal{M}_{\mathbf{p}}(\hat{\mathbf{z}}^{n,m})-\mathcal{M}_{\mathbf{p}}^{n,m}(\mathbf{z}^*)+\mathcal{M}_{\mathbf{p}}^{n,m}(\hat{\mathbf{z}}^{n,m})|\\&=\left|\sum_{k=2}^N(p_k-p_k^n)(z_k-\hat{z}_k^n)+\int(\inf_{i=2,\dots,N}\{c(\mathbf{x}_i,\mathbf{y})-z_i\}-\inf_{i=2,\dots,N}\{c(\mathbf{x}_i,\mathbf{y})-\hat{z}_i^n\})(dQ-dQ_m)(\mathbf{y})\right|,\\&\leq|\mathbf{p}^n-\mathbf{p}|_2|\mathbf{z}^*-\hat{\mathbf{z}}^{n,m}|_2+\|\inf_{i=2,\dots,N}\{c(\mathbf{x}_i,\cdot)-z_i\}-\inf_{i=2,\dots,N}\{c(\mathbf{x}_i,\cdot)-\hat{z}_i^n\}\|_\infty\|Q-Q_m\|_{\overline{\mathrm{spam}(\mathcal{F}_c)}},\end{aligned}$$

where $\overline{\mathrm{spam}(\mathcal{F}_c)}$ is the topological closure in $\mathcal{C}(\mathcal{Y})$ of the set of finite linear combinations of $\mathcal{F}_c$, which is still Donsker (see Theorems 2.10.2 and 2.10.6 in Vaart and Wellner (1996)), give the Lipschitz property

$$|\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*)-\mathcal{M}_{\mathbf{p}}(\hat{\mathbf{z}}^{n,m})-\mathcal{M}_{\mathbf{p}}^{n,m}(\mathbf{z}^*)-\mathcal{M}_{\mathbf{p}}^{n,m}(\hat{\mathbf{z}}^{n,m})| \leq C\sqrt{\frac{n+m}{n\,m}}|\mathbf{z}^*-\hat{\mathbf{z}}^{n,m}|, \quad \text{for some } C>0. \tag{3.41}$$

Therefore, (3.40) and (3.41) give

$$\begin{aligned} C\sqrt{\frac{n+m}{n\,m}}|\mathbf{z}^*-\hat{\mathbf{z}}^{n,m}| &\geq \mathcal{M}_{\mathbf{p}}(\mathbf{z}^*)-\mathcal{M}_{\mathbf{p}}(\hat{\mathbf{z}}^{n,m})-\mathcal{M}_{\mathbf{p}}^{n,m}(\mathbf{z}^*)+\mathcal{M}_{\mathbf{p}}^{n,m}(\hat{\mathbf{z}}^{n,m}) \\ &\geq \mathcal{M}_{\mathbf{p}}(\mathbf{z}^*)-\mathcal{M}_{\mathbf{p}}(\hat{\mathbf{z}}^{n,m}) \geq c|\hat{\mathbf{z}}^{n,m}-\mathbf{z}^*|^2, \end{aligned}$$

from where we deduce the tightness of $\sqrt{\frac{n\,m}{n+m}}|\hat{\mathbf{z}}^{n,m}-\mathbf{z}^*|$.

Lemma 3..5 gives the relation

$$\nabla_{\mathbf{z}}\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*)-\nabla_{\mathbf{z}}\mathcal{M}_{\mathbf{p}}(\mathbf{z}) = D^2\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*)(\mathbf{z}-\mathbf{z}^*)+O(|\mathbf{z}-\mathbf{z}|^2).$$

Therefore, we obtain the following equation

$$\nabla_{\mathbf{z}}\mathcal{M}_{\mathbf{p}}(\hat{\mathbf{z}}^{n,m})-\nabla_{\mathbf{z}}\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*) = D^2\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*)(\hat{\mathbf{z}}^{n,m}-\mathbf{z}^*)+O_P(|\hat{\mathbf{z}}^{n,m}-\mathbf{z}^*|^2),$$

and, since $0=\nabla_{\mathbf{z}}\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*)=\nabla_{\mathbf{z}}\mathcal{M}_{\mathbf{p}}^{n,m}(\mathbf{z}_{n,m})$, also

$$\nabla_{\mathbf{z}}\mathcal{M}_{\mathbf{p}}(\hat{\mathbf{z}}^{n,m})-\nabla_{\mathbf{z}}\mathcal{M}_{\mathbf{p}}^{n,m}(\hat{\mathbf{z}}^{n,m}) = D^2\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*)(\hat{\mathbf{z}}^{n,m}-\mathbf{z}^*)+O_P(|\hat{\mathbf{z}}^{n,m}-\mathbf{z}^*|^2). \tag{3.42}$$

The left hand side of (3.42) can be written as

$$\begin{aligned} &(Q_m(\mathrm{Lag}_1(\hat{\mathbf{z}}^{n,m}))-Q(\mathrm{Lag}_1(\hat{\mathbf{z}}^{n,m}))+p_1-p_1^n,\dots, \\ &\qquad Q_m(\mathrm{Lag}_N(\hat{\mathbf{z}}^{n,m}))-Q(\mathrm{Lag}_N(\hat{\mathbf{z}}^{n,m}))+p_N-p_N^n), \end{aligned}$$

The following result provides its weak limit.

Lemma 3..6. *Suppose that the assumptions of Lemma 3..5 hold. If $m=m(n)$ is such that $n\to\infty$ and $\frac{m}{n+m}\to\lambda\in(0,1)$, we have*

$$\sqrt{\frac{n\,m}{n+m}}\left(\nabla_{\mathbf{z}}\mathcal{M}_{\mathbf{p}}(\hat{\mathbf{z}}^{n,m})-\nabla_{\mathbf{z}}\mathcal{M}_{\mathbf{p}}^{n,m}(\hat{\mathbf{z}}^{n,m})\right) \xrightarrow{w} (\mathbb{U}_1,\dots,\mathbb{U}_N).$$

Proof. From the relation

$$\begin{aligned} &\nabla_{\mathbf{z}}\mathcal{M}_{\mathbf{p}}(\hat{\mathbf{z}}^{n,m})-\nabla_{\mathbf{z}}\mathcal{M}_{\mathbf{p}}^{n,m}(\hat{\mathbf{z}}^{n,m}) \\ &\qquad = (Q_m(\mathrm{Lag}_1(\hat{\mathbf{z}}^{n,m}))-Q(\mathrm{Lag}_1(\hat{\mathbf{z}}^{n,m}))+p_1-p_1^n, \\ &\qquad\qquad \dots, Q_m(\mathrm{Lag}_N(\hat{\mathbf{z}}^{n,m}))-Q(\mathrm{Lag}_N(\hat{\mathbf{z}}^{n,m}))+p_N-p_N^n) \end{aligned}$$

and the multivariate central limit theorem, we observe that it is enough to show the convergence

$$\sqrt{m}(Q_m(\mathrm{Lag}_1(\hat{\mathbf{z}}^{n,m})) - Q(\mathrm{Lag}_1(\hat{\mathbf{z}}^{n,m})), \dots, Q_m(\mathrm{Lag}_N(\hat{\mathbf{z}}^{n,m})) - Q(\mathrm{Lag}_N(\hat{\mathbf{z}}^{n,m}))) \xrightarrow{w} (\mathbb{U}_1, \dots, \mathbb{U}_N).$$

To prove it we set k and observe that

$$\mathrm{Lag}_k(\mathbf{z}) = \bigcap_{i=1,\dots,N} \{\mathbf{y} \in \mathbb{R}^d : \ c(\mathbf{x}_k, \mathbf{y}) - c(\mathbf{x}_i, \mathbf{y}) < z_k - z_i\}, \tag{3.43}$$

which means that the class of all possible cells $\{\mathrm{Lag}_k(\mathbf{z}) : \ \mathbf{z} \in \mathbb{R}^N\} \subset 2^{\mathcal{Y}}$ is contained in

$$\left\{ \bigcap_{i=1}^{N} \{\mathbf{y} \in \mathbb{R}^d : \ c(\mathbf{x}_k, \mathbf{y}) - c(\mathbf{x}_i, \mathbf{y}) < s_k, \ s_k \in \mathbb{R}\} \right\}.$$

We note that if the Vapnik-Chervonenkis (VC) dimension of

$$\left\{ \{\mathbf{y} \in \mathbb{R}^d : \ c(\mathbf{x}_k, \mathbf{y}) - c(\mathbf{x}_i, \mathbf{y}) - s < 0\} : \ s \in \mathbb{R} \right\}$$

is D, the one of $\{\mathrm{Lag}_k(\mathbf{z}) : \ \mathbf{z} \in \mathbb{R}^N\}$ is $2ND\log(3N)$ (see Lemma 3.2.3. in Linial et al. (1991)). It is trivial to show that the VC dimension of the space of functions $\{\mathbf{y} \mapsto c(\mathbf{x}_k, \mathbf{y}) - c(\mathbf{x}_i, \mathbf{y}) - s : \ s \in \mathbb{R}\}$ is 2. Then, in view of Theorem 2.6.4 in Vaart and Wellner (1996), we have

$$\mathcal{N}(\epsilon, \{\mathbb{1}_{\mathrm{Lag}_k(\mathbf{z})} : \ \mathbf{z} \in \mathbb{R}^N\}, L_2(Q)) \le C \frac{1}{\epsilon^{8N\log(3N)}},$$

which means that the class $\{\mathbb{1}_{\mathrm{Lag}_k(\mathbf{z})} : \ \mathbf{z} \in \mathbb{R}^N\}$ is $Q-$Donsker. This means that, for every $\epsilon > 0$

$$\lim_{\delta \to 0} \limsup_{m \to \infty} \mathbb{P}\left(\sup_{Q(\mathrm{Lag}_k(\mathbf{u}) \Delta \mathrm{Lag}_k(\mathbf{v})) < \delta} \left| \sqrt{m} \int \mathbb{1}_{\mathrm{Lag}_k(\mathbf{u})} - \mathbb{1}_{\mathrm{Lag}_k(\mathbf{v})} d(Q_m - Q) \right| > \epsilon \right) = 0, \tag{3.44}$$

where

$$Q(\mathrm{Lag}_k(\mathbf{u}) \Delta \mathrm{Lag}_k(\mathbf{v})) = Q((\mathrm{Lag}_k(\mathbf{u}) \setminus \mathrm{Lag}_k(\mathbf{v})) \cup (\mathrm{Lag}_k(\mathbf{v}) \setminus \mathrm{Lag}_k(\mathbf{u}))) = \int (\mathbb{1}_{\mathrm{Lag}_k(\mathbf{u})} - \mathbb{1}_{\mathrm{Lag}_k(\mathbf{v})})^2 dQ.$$

First we bound

$$Q(\mathrm{Lag}_k(\hat{\mathbf{z}}^{n,m}) \Delta \mathrm{Lag}_k(\mathbf{z}^*)) \le Q(\mathrm{Lag}_k(\hat{\mathbf{z}}^{n,m}) \backslash \mathrm{Lag}_k(\mathbf{z}^*)) + Q(\mathrm{Lag}_k(\mathbf{z}^*) \backslash \mathrm{Lag}_k(\hat{\mathbf{z}}^{n,m})) \tag{3.45}$$

The we show that $Q(\mathrm{Lag}_k(\hat{\mathbf{z}}^{n,m}) \setminus \mathrm{Lag}_k(\mathbf{z}^*)) \xrightarrow{P} 0$, the same holds for $Q(\mathrm{Lag}_k(\mathbf{z}^*) \setminus \mathrm{Lag}_k(\hat{\mathbf{z}}^{n,m}))$, yielding the limit in probability of (3.45) towards 0. Using (3.43), we have

$$\mathrm{Lag}_k(\hat{\mathbf{z}}^{n,m}) \setminus \mathrm{Lag}_k(\mathbf{z}^*) = \bigcup_{i=1,\dots,N} \mathrm{Lag}_k(\hat{\mathbf{z}}^{n,m}) \setminus \{\mathbf{y} \in \mathbb{R}^d : \; c(\mathbf{x}_k, \mathbf{y}) - c(\mathbf{x}_i, \mathbf{y}), < z_k^* - z_i^*\}.$$

which, in view of the union bound and (3.43), gives

$$\begin{aligned} Q & \left(\mathrm{Lag}_k(\hat{\mathbf{z}}^{n,m}) \setminus \mathrm{Lag}_k(\mathbf{z}^*)\right) \\ & \le \sum_{i=1}^N Q(\mathrm{Lag}_k(\hat{\mathbf{z}}^{n,m}) \setminus \{\mathbf{y} \in \mathbb{R}^d : \; c(\mathbf{x}_k, \mathbf{y}) - c(\mathbf{x}_i, \mathbf{y}) < z_k^* - z_i^*\}) \\ & \qquad \le \sum_{i=1}^N Q(\{\mathbf{y} \in \mathbb{R}^d : \; z_k^{n,m} - z_i^{n,m} < c(\mathbf{x}_k, \mathbf{y}) - c(\mathbf{x}_i, \mathbf{y}) < z_k^* - z_i^*\}). \end{aligned}$$

Since $\sqrt{\frac{n\,m}{n+m}}|\hat{\mathbf{z}}^{n,m} - \mathbf{z}^*|$ is tight, there exists a sub-sequence of $|\hat{\mathbf{z}}^{n,m} - \mathbf{z}^*|$ converging a.s. to 0. We keep the same notation for the sub-sequence. Set

$$\mathbf{y} \in \bigcap_{l \in \mathbb{N}} \bigcup_{n,m > l} \{\mathbf{y} \in \mathbb{R}^d : \; z_k^{n,m} - z_i^{n,m} < c(\mathbf{x}_k, \mathbf{y}) - c(\mathbf{x}_i, \mathbf{y}) < z_k^* - z_i^*\}.$$

Then it satisfies that for all $l \in \mathbb{N}$, there exist $n, m > l$ such that

$$z_k^{n,m} - z_i^{n,m} < c(\mathbf{x}_k, \mathbf{y}) - c(\mathbf{x}_i, \mathbf{y}) < z_k^* - z_i^*.$$

The a.s. limit $z_k^{n,m} - z_i^{n,m} \to z_k^* - z_i^*$ implies that $c(\mathbf{x}_k, \mathbf{y}) - c(\mathbf{x}_i, \mathbf{y}) = z_k^* - z_i^*$, which, under the assumption (Twist), is negligible for Q—which satisfies (Cont). Therefore, $Q(\mathrm{Lag}_k(\hat{\mathbf{z}}^{n,m}) \setminus \mathrm{Lag}_k(\mathbf{z}^*)) \xrightarrow{P} 0$ and, by symmetry,

$$Q(\mathrm{Lag}_k(\hat{\mathbf{z}}^{n,m}) \Delta \, \mathrm{Lag}_k(\mathbf{z}^*)) \xrightarrow{P} 0 \tag{3.46}$$

Moreover, by repeating the same argument, we obtain that if $\mathbf{u} \to \mathbf{v}$, then

$$Q(\mathrm{Lag}_k(\mathbf{u}) \Delta \, \mathrm{Lag}_k(\mathbf{u})) \to 0 \tag{3.47}$$

Thanks to (3.46), (3.47) and (3.44), we are under the hypotheses of Lemma 3.3.5 in Vaart and Wellner (1996); as a consequence we have

$$\begin{aligned} \sqrt{m} & |Q_m(\mathrm{Lag}_k(\hat{\mathbf{z}}^{n,m})) - Q(\mathrm{Lag}_k(\hat{\mathbf{z}}^{n,m})) - (Q_m(\mathrm{Lag}_k(\mathbf{z}^*)) - Q(\mathrm{Lag}_k(\mathbf{z}^*)))| \\ & = \sqrt{m} \left| \int \mathbb{1}_{\mathrm{Lag}_k(\hat{\mathbf{z}}^{n,m})} - \mathbb{1}_{\mathrm{Lag}_k(\mathbf{z}^*)} d(Q_m - Q) \right| \xrightarrow{P} 0. \end{aligned}$$

This implies

$$\sqrt{m}(Q_m(\mathrm{Lag}_1(\hat{\mathbf{z}}^{n,m})) - Q(\mathrm{Lag}_1(\hat{\mathbf{z}}^{n,m})) = \sqrt{m}(Q_m(\mathrm{Lag}_k(\mathbf{z}^*)) - Q(\mathrm{Lag}_k(\mathbf{z}^*)) + o_P(1).$$

As k was chosen arbitrarily, we obtain also

$$\sqrt{m}(Q_m(\mathrm{Lag}_1(\hat{\mathbf{z}}^{n,m})) - Q(\mathrm{Lag}_1(\hat{\mathbf{z}}^{n,m})), \dots, Q_m(\mathrm{Lag}_N(\hat{\mathbf{z}}^{n,m})) - Q(\mathrm{Lag}_N(\hat{\mathbf{z}}^{n,m})))$$
$$\sqrt{m}(Q_m(\mathrm{Lag}_1(\hat{\mathbf{z}}^*)) - Q(\mathrm{Lag}_1(\hat{\mathbf{z}}^*)), \dots, Q_m(\mathrm{Lag}_N(\hat{\mathbf{z}}^*)) - Q(\mathrm{Lag}_N(\hat{\mathbf{z}}^*))) + o_P(1).$$

The multi-variate central limit theorem yields the limit of

$$\sqrt{m}(Q_m(\mathrm{Lag}_1(\hat{\mathbf{z}}^*)) - Q(\mathrm{Lag}_1(\hat{\mathbf{z}}^*)), \dots, Q_m(\mathrm{Lag}_N(\hat{\mathbf{z}}^*)) - Q(\mathrm{Lag}_N(\hat{\mathbf{z}}^*)))$$

as a centered Gaussian r.v. $(X_1 \dots, X_N)$ with covariance

$$E(X_i X_j) = Q(\mathrm{Lag}_i(\hat{\mathbf{z}}^*) \cap \mathrm{Lag}_j(\hat{\mathbf{z}}^*)) - Q(\mathrm{Lag}_i(\hat{\mathbf{z}}^*))Q(\mathrm{Lag}_j(\hat{\mathbf{z}}^*)) = \mathbb{1}_{i=j} p_i^2 - p_i p_j.$$

Since $(p_1 - p_1^n, \dots, p_N - p_N^n)$ converges weakly to a independent copy of $(X_1 \dots, X_N)$, we conclude. □

Lemma 3..6, (3.42) and the continuous mapping theorem conclude the proof. □

Proof of Theorem 3.3.3.
The first step is to show that (3.24) holds. Set $k \in \{1, \dots, N\}$. Since

$$\mathrm{Lag}_k^R(\mathbf{z}) = R\,\mathbb{B}_d \bigcap_{i=1,\dots,N} \{\mathbf{y} \in \mathbb{R}^d : \; 2\langle \mathbf{x}_k - \mathbf{x}_i, \mathbf{y}\rangle \geq |\mathbf{x}_k|^2 - |\mathbf{x}_i|^2 - z_k + z_i\} \cap R\,\mathbb{B}_d,$$

the support functions of the sets

$$\{\mathbf{y} \in \mathbb{R}^d : \; 2\langle \mathbf{x}_k - \mathbf{x}_i, \mathbf{y}\rangle \geq |\mathbf{x}_k|^2 - |\mathbf{x}_i|^2 - z_k + z_i\}$$

and $R\,\mathbb{B}_d$ are respectively

$$h_i(\mathbf{v}) = \begin{cases} t\,(-|\mathbf{x}_k|^2 + |\mathbf{x}_i|^2 + z_k - z_i) & \text{if } \mathbf{v} = t(\mathbf{x}_i - \mathbf{x}_k),\; t \geq 0 \\ +\infty & \text{otherwise} \end{cases}$$

and $h_{R\,\mathbb{B}_d}(\mathbf{v}) = R\,|\mathbf{v}|$. Then, in view of Remark 3.3.5, we have

$$h_{\mathrm{Lag}_k^R(\mathbf{z})}(\mathbf{v}) = \inf\left\{ R|\mathbf{v}_k| + \sum_{i\neq k} h_i(\mathbf{v}_i) : \; \sum_{i=1}^N \mathbf{v}_i = \mathbf{v} \right\}.$$

We parameterize $\mathbf{v}_i = t_i(\mathbf{x}_i - \mathbf{x}_k)$, for $t_i \geq 0$, so that $h_i(\mathbf{v}_i) \in \mathbb{R}$ and thus $\mathbf{v}_k = \mathbf{v} - \sum_{i\neq k} t_i(\mathbf{x}_i - \mathbf{x}_k)$. This gives

$$h_{\mathrm{Lag}_k^R(\mathbf{z})}(\mathbf{v}) = \inf\left\{ R|\mathbf{v} - \sum_{i\neq k} t_i(\mathbf{x}_i - \mathbf{x}_k)| + \sum_{i\neq k} t_i\,(-|\mathbf{x}_k|^2 + |\mathbf{x}_i|^2 + z_k - z_i) : \; t_i \geq 0 \right\}$$

and (3.23) holds. The support function of the singleton $\{-\mathbf{y}^0\}$ is given by

$$h_{\{-\mathbf{y}^0\}}(\mathbf{v}) = -\langle \mathbf{y}^0, \mathbf{v}\rangle = -\langle \mathbf{v} - \sum_{j\neq k} t_j(\mathbf{x}_k - \mathbf{x}_j), \mathbf{y}^0\rangle - \langle \sum_{j\neq k} t_j(\mathbf{x}_k - \mathbf{x}_j), \mathbf{y}^0\rangle,$$

so that, in view of Remark 3.3.5, (3.24) holds.

The next step is to show the Hadamard differentiability of $h_{\mathrm{Lag}_k^R(\mathbf{z})+\{-\mathbf{y}^0\}}(\mathbf{v})$ with respect to $\mathbf{z}$ in a neighborhood of $\mathbf{z}^*$. Note that the objective function

$$(\mathbf{z}, \mathbf{t}) \mapsto F(\mathbf{z}, \mathbf{t}) = \sum_{j\neq k} t_j \psi_j(\mathbf{z}) + R|\mathbf{v} - \sum_{j\neq k} t_j(\mathbf{x}_k - \mathbf{x}_j)| - \langle \mathbf{v} - \sum_{j\neq k} t_j(\mathbf{x}_k - \mathbf{x}_j), \mathbf{y}^0\rangle)$$

is $2\,N$-Lipschitz w.r.t. $\mathbf{z}$—it is immediate to check out that $|F(\mathbf{z}, \mathbf{t}) - F(\mathbf{z}', \mathbf{t})| \leq 2\,N\,|\mathbf{z} - \mathbf{z}'|_\infty$—and by the relation

$$|\inf_{\mathbf{t}} F(\mathbf{z}, \mathbf{t}) - \inf_{\mathbf{t}} F(\mathbf{z}', \mathbf{t})| \leq \sup_{\mathbf{t}} |F(\mathbf{z}, \mathbf{t}) - F(\mathbf{z}', \mathbf{t})| \leq 2\,N\,|\mathbf{z} - \mathbf{z}'|_\infty,$$

the same holds for $h_{\mathrm{Lag}_k^R(\mathbf{z})+\{-\mathbf{y}^0\}}(\mathbf{v})$. Then we only need to prove the Gateaux differentiablity. Sufficient conditions conditions are given by Proposition 4.13 in Bonnans and Shapiro (2000), which are as follows:

(i) The function $F(\cdot, \mathbf{t})$ is Gateaux differentiable in $\mathbb{R}^N$ with derivative $D_{\mathbf{z}} F(\mathbf{z}, \mathbf{t})$,

(ii) the function F is continuous in $\mathbb{R}^N \times \mathbb{R}^{N-1}$, and

(iii) there exists $\alpha \in \mathbb{R}$ and a compact set $C \subset \mathbb{R}^N$ such that for every $\mathbf{z}$ near $\mathbf{z}^*$, the level set

$$\mathrm{Lev}_\alpha\, F(\mathbf{z}, \cdot) = \{\mathbf{t} \in \mathbb{R}^{N-1} :\ t_i \geq 0,\ F(\mathbf{z}, \mathbf{t}) \leq \alpha\}$$

is non empty and contained in C.

Since the objective function is linear in $\mathbf{z}$ and continuous in $\mathbf{t}$, (i) and (ii) hold. To prove the compactness of the level sets we observe that, for all $j \neq k$, $\psi_j(\mathbf{z}^*) = (|\mathbf{x}_k|^2 - |\mathbf{x}_j|^2 - z_k^* + z_j^*) - \langle \mathbf{x}_k - \mathbf{x}_j, \mathbf{y}^0\rangle \geq a > 0$, for some $a > 0$. Moreover, the uniform continuity of the function $\mathbf{z} \to \psi_j(\mathbf{z})$, implies the existence of a neighborhood $\mathcal{U}^*$ of $\mathbf{z}^*$ such that $\psi_j(\mathbf{z}) > \frac{a}{2}$, for all $\mathbf{z} \in \mathcal{U}^*$ and $j \neq k$. Evaluating at $\mathbf{t} = \mathbf{0}$, we have $F(\mathbf{z}^*, \cdot) = R - \langle \mathbf{v}, \mathbf{y}^0\rangle \leq 2R$. As a consequence,

$$\frac{a}{2} \sum_{j\neq k} t_j \leq \sum_{j\neq k} t_j \psi_j(\mathbf{z}) \leq h_{\mathrm{Lag}_k^R(\mathbf{z})+\{-\mathbf{y}^0\}}(\mathbf{v}) \leq 2R,$$

for all $\mathbf{z} \in \mathcal{U}^*$. The set of $\mathbf{t} \in \mathbb{R}^{N-1}$ such that $t_j \geq 0$ and $\sum_{j\neq k} t_j \leq \frac{4\,R}{a}$ plays the role of C in (iii), which is thus proven. The Hadamard derivative in a direction $\mathbf{d}$ of $h_{\mathrm{Lag}_k^R(\mathbf{z})+\{-\mathbf{y}^0\}}(\mathbf{v})$ in $\mathbf{z}^*$ is thus given by

$$\inf_{\mathbf{t}\in \mathrm{Sol}(\mathbf{z}^*, \mathbf{v}, \mathbf{y}^0)} \left\{ \sum_{j\neq k} t_j(d_k - d_j) \right\}.$$

Set $\mathbf{d} \in \mathbb{R}^N$. Now we claim that

$$\frac{\left(\int \left|h_{\mathrm{Lag}_k^R(\mathbf{z}^*+t_n\mathbf{d}_n)}(\mathbf{v}) - h_{\mathrm{Lag}_k^R(\mathbf{z}^*)}(\mathbf{v})\right|^p d\mathcal{H}^{d-1}(\mathbf{v})\right)^{\frac{1}{p}}}{t_n} \longrightarrow \left(\int \left|\inf_{\mathbf{t}\in\mathrm{Sol}(\mathbf{z}^*,\mathbf{v},\mathbf{y}^0)} \left\{\sum_{j\neq k} t_j(d_k - d_j)\right\}\right|^p\right)^{\frac{1}{p}}, \quad (3.48)$$

for any $t_n \searrow 0$ and $\mathbf{d}_n \to \mathbf{d}$. Lemma 3.3.6 states the existence of some $M > 0$ such that that

$$\limsup_n t_n d_\infty(\mathrm{Lag}_k^R(\mathbf{z}^* + t_n\mathbf{h}_n), \mathrm{Lag}_k^R(\mathbf{z}^*)) \leq M.$$

The quantity $(t_n d_\infty(\mathrm{Lag}_k^R(\mathbf{z}^* + t_n\mathbf{h}_n))^p$ is constant w.r.t. $\mathbf{v}$ and dominates

$$\frac{\left|h_{\mathrm{Lag}_k^R(\mathbf{z}^*+t_n\mathbf{d}_n)}(\mathbf{v}) - h_{\mathrm{Lag}_k^R(\mathbf{z}^*)}(\mathbf{v})\right|^p}{t_n},$$

for all $\mathbf{v} \in \mathbb{S}^{d-1}$. In view of the dominated convergence theorem and the point-wise Hadamard differentiablitiy (previous step), (3.48) holds. Finally, following verbatim the proof of Theorem 1 in Römisch (2014), we conclude. □

3..3 Proofs of Lemmas

Proof of Lemma 3.1.1.
First, strong duality (3.3) yields that

$$\mathcal{T}_c(P,Q) = \sup_{(f,g)\in\Phi_c(P,Q)} \int f(\mathbf{x})dP(\mathbf{x}) + \int g(\mathbf{y})dQ(\mathbf{y}) = \sup_{(f,g)\in\Phi_c(P,Q)} \sum_{i=1}^N f(\mathbf{x_i})p_i + \int g(\mathbf{y})dQ(\mathbf{y}).$$

Set $(z_1, \ldots, z_N) = (f(\mathbf{x}_i), \ldots, f(\mathbf{x}_N))$, then $\mathcal{T}_c(P,Q) = \sup_{(\mathbf{z},g)} \sum_{i=1}^N z_i p_i + \int g(\mathbf{y})dQ(\mathbf{y})$, where the sup is taken on the set $(\mathbf{z}, g)$ such that $z_i + g(\mathbf{y}) \leq c(\mathbf{x}_i, \mathbf{y})$ for all $i = 1, \ldots, N$. Then $g(\mathbf{y}) \leq \inf_{i=1,\ldots,N}\{c(\mathbf{x}_i,\mathbf{y}) - z_i\}$ and $\mathcal{T}_c(P,Q) = \sup_{\mathbf{z}\in\mathbb{R}^N} g_c(P,Q,\mathbf{z})$.

Let $\mathbf{z}^* = (z_1^*, \ldots, z_N^*) \in \mathbb{R}^N$ be such that $\mathcal{T}_c(P,Q) = g_c(P,Q,\mathbf{z}^*)$. Denote as $l = \arg\inf_i z_i^*$ and $u = \arg\sup_i z_i^*$, which are different—otherwise the potentials are constant and we conclude that $K = 0$. Therefore

$$\begin{aligned}\mathcal{T}_c(P,Q) &\le \sum_{i=1}^{N} z_i^* p_i + \int \{c(\mathbf{y},\mathbf{x}_u) - z_u^*\} dQ(\mathbf{y}) \\ &\le (1-p_l) z_u^* + p_l z_l^* + \int \{c(\mathbf{y},\mathbf{x}_u) - z_u^*\} dQ(\mathbf{y}) \\ &\le -p_l z_u^* + p_l z_l^* + \int c(\mathbf{y},\mathbf{x}_u) dQ(\mathbf{y}) \\ &\le p_l (z_l^* - z_u^*) + \int c(\mathbf{y},\mathbf{x}_u) dQ(\mathbf{y}),\end{aligned}$$

which implies $p_l(z_u^* - z_l^*) \le \int c(\mathbf{y},\mathbf{x}_u) dQ(\mathbf{y}) - \mathcal{T}_c(P,Q)$ and

$$\sup_{i,j=1,\dots,N} |z_i^* - z_j^*| \le \frac{1}{\inf_i p_i} \left(\sup_{i=1,\dots,N} \int c(\mathbf{y},\mathbf{x}_i) dQ(\mathbf{y}) - \mathcal{T}_c(P,Q) \right),$$

since adding additive constant does not change $g_c(P,Q,\mathbf{z})$, then we conclude. □

Lemma 3..7. *Under the assumptions of Theorem 3.1.4, the class $\mathcal{F}_c^K$ is Q-Donsker.*

Proof of Lemma 3..7.

We use bracketing numbers, see Definition 2.1.6 in Vaart and Wellner (1996). Lemma 3.1.3 implies that

$$N_{[]}(2\epsilon, \mathcal{F}_c^K, ||\cdot||_{L^2(Q)}) \le N(\epsilon, \mathbb{B}_K(\mathbf{0}), |\cdot|).$$

Therefore, Lemma 4.14 in Massart and Picard (2007) implies that

$$\int_0^\infty \sqrt{\log\left(N_{[]}(\epsilon, \mathcal{F}_c^K, ||\cdot||_{L^2(Q)})\right)} d\epsilon < \infty. \tag{3.49}$$

The envelope function of the class $\mathcal{F}_c^K$ can be taken as the function F defined as $F(\mathbf{y}) = \sup_{i=1,\dots,m} c(\mathbf{x}_i,\mathbf{y}) + K$. Note that

$$\int F(\mathbf{y})^2 dQ(\mathbf{y}) \le 2K + 2 \int \sup_{i=1,\dots,m} c(\mathbf{x}_i,\mathbf{y})^2 dQ(\mathbf{y}) < \infty.$$

Using Theorem 3.7.38 in Giné and Nickl (2015) we obtain the desired result. □

Bibliography

Gabriel Acosta and Ricardo G. Durán. An optimal poincaré inequality in l1 for convex domains. *Proceedings of the American Mathematical Society*, 132(1):195–202, 2004. ISSN 00029939, 10886826. URL http://www.jstor.org/stable/1193994.

Mohit Bansil and Jun Kitagawa. Quantitative stability in the geometry of semi-discrete optimal transport. *International Mathematics Research Notices*, 2020.

J. F. Bonnans and A. Shapiro. Perturbation analysis of optimization problems. In *Springer Series in Operations Research*, 2000.

Luis A. Caffarelli. Boundary regularity of maps with convex potentials–ii. *Annals of Mathematics*, 144(3):453–496, 1996. ISSN 0003486X. URL http://www.jstor.org/stable/2118564.

Dario Cordero-Erausquin and Alessio Figalli. Regularity of monotone transport maps between unbounded domains. *Discrete and Continuous Dynamical Systems*, 39(0947–1078):7101–7112, 2019. ISSN 1078-0947.

Javier Cárcamo, Antonio Cuevas, and Luis-Alberto Rodríguez. Directional differentiability for supremum-type functionals: Statistical applications. *Bernoulli*, 26(3):2143 – 2175, 2020. doi: 10.3150/19-BEJ1188. URL https://doi.org/10.3150/19-BEJ1188.

Eustasio del Barrio, Alberto González-Sanz, and Marc Hallin. A note on the regularity of optimal-transport-based center-outward distribution and quantile functions. *Journal of Multivariate Analysis*, 180:104671, 2020. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2020.104671. URL https://www.sciencedirect.com/science/article/pii/S0047259X20302529.

Eustasio del Barrio, Alberto González-Sanz, and Jean-Michel Loubes. Central limit theorems for general transportation costs. *Pre-Print*, 2021.

Eustasio del Barrio, Alberto González-Sanz, and Jean-Michel Loubes. Central limit theorems for semidiscrete wasserstein distances. *Pre-Print*, 2022.

Zheng Fang and Andres Santos. Inference on Directionally Differentiable Functions. *The Review of Economic Studies*, 86(1):377–412, 09 2018. ISSN 0034-6527. doi: 10.1093/restud/rdy049. URL https://doi.org/10.1093/restud/rdy049.

Alfred Galichon. *Optimal Transport Methods in Economics*. Princeton University Press, 09 2016. ISBN 9780691172767. doi: 10.23943/princeton/9780691172767.001.0001. URL https://doi.org/10.23943/princeton/9780691172767.001.0001.

Wilfrid Gangbo and Robert J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996. doi: 10.1007/BF02392620. URL https://doi.org/10.1007/BF02392620.

Eva Giné and Richard Nickl. Mathematical foundations of infinite-dimensional statistical models. In *Cambridge Series in Statistical and Probabilistic Mathematics, [40]. Cambridge University Press, New York.*, 2015.

Marc Hallin, Eustasio del Barrio, Juan Cuesta-Albertos, and Carlos Matrán. Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *The Annals of Statistics*, 49(2):1139 – 1165, 2021. doi: 10.1214/20-AOS1996.

Jun Kitagawa, Quentin Mérigot, and Boris Thibert. Convergence of a newton algorithm for semi-discrete optimal transport. *J. Eur. Math. Soc.*, 21:2603–2651, 2019. doi: 10.4171/JEMS/889.

Nathan Linial, Yishay Mansour, and Ronald L. Rivest. Results on learnability and the vapnik-chervonenkis dimension. *Information and Computation*, 90(1):33–49, 1991. ISSN 0890-5401. doi: https://doi.org/10.1016/0890-5401(91)90058-A. URL https://www.sciencedirect.com/science/article/pii/089054019190058A.

Pascal Massart and Jean Picard. Concentration inequalities and model selection. In *École d'été de probabilités de Saint-Flour*, 2007.

Emanuel Milman. On the role of convexity in isoperimetry, spectral gap and concentration. *Inventiones mathematicae*, 177:1–43, 2007.

D. Pollard. Quantization and the method of k -means. *IEEE Trans. Inf. Theor.*, 28(2): 199–205, sep 2006. ISSN 0018-9448. doi: 10.1109/TIT.1982.1056481. URL https://doi.org/10.1109/TIT.1982.1056481.

David Pollard. A Central Limit Theorem for k-Means Clustering. *The Annals of Probability*, 10(4):919 – 926, 1982. doi: 10.1214/aop/1176993713. URL https://doi.org/10.1214/aop/1176993713.

M. Rathmair. On how poincaré inequalities imply weighted ones. *Monatsh Math*, 118: 753–763, 2019. URL https://doi.org/10.1007/s00605-019-01266-w.

R. Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series. Princeton University Press, Princeton, N. J., 1970.

R. Tyrrell Rockafellar and Roger J.-B. Wets. *Variational Analysis*. Springer Verlag, Heidelberg, Berlin, New York, 1998.

Werner Römisch. *Delta Method, Infinite Dimensional*. American Cancer Society, 2014. ISBN 9781118445112. doi: https://doi.org/10.1002/9781118445112.stat01949. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat01949.

Johan Segers. Graphical and uniform consistency of estimated optimal transport plans, 2022. URL https://arxiv.org/abs/2208.02508.

Max Sommerfeld and Axel Munk. Inference for empirical wasserstein distances on finite spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1): 219–238, 2018.

Aad W. Van Der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer, New York, NY, 1996.

Richard A Vitale. Lp metrics for compact, convex sets. *Journal of Approximation Theory*, 45(3):280–287, 1985. ISSN 0021-9045. doi: https://doi.org/10.1016/0021-9045(85)90051-6. URL https://www.sciencedirect.com/science/article/pii/0021904585900516.

CHAPTER 4

An improved central limit theorem and fast convergence rates for entropic transportation costs

The content of this chapter is fully available online in del Barrio et al. (2022).

Contents

We prove a central limit theorem for the entropic transportation cost between subgaussian probability measures, centered at the population cost. This is the first result which allows for asymptotically valid inference for entropic optimal transport between measures which are not necessarily discrete. In the compactly supported case, we complement these results with new, faster, convergence rates for the expected entropic transportation cost between empirical measures. Our proof is based on strengthening convergence results for dual solutions to the entropic optimal transport problem.

4.1 Introduction.

Optimal transport has emerged as a leading methodology in many areas of data science, machine learning, and statistics Rigollet and Weed (2018b); Hütter and Rigollet (2021); Hallin et al. (2021a); Ghosal and Sen (2019); Deb and Sen (2019); Hallin et al. (2021b); Móri and Székely (2020); Abadie and Imbens (2006); Todd (2010); Morgan and Harding (2006); Blanchet et al. (2019); Kuhn et al. (2019); Flamary et al. (2018); Gao and Kleywegt (2016); Gordaliza et al. (2019); Barrio et al. (2019); Black et al. (2020); Chiappa and Pacchiano (2021); Chiappa et al. (2020); Li et al. (2020); Courty et al. (2017); Redko et al. (2017); Grave et al. (2019); Alvarez-Melis and Jaakkola (2018); de Lara et al. (2021); Shi et al.

(2022, 2021b,a); Risser et al. (2021), with applications in fields ranging from high-energy physics Romão et al. (2020); Komiske et al. (2019) to computational biology Schiebinger et al. (2019); Yang et al. (2020). Central to its recent success in practice is the paradigm of entropic regularization, popularized by Cuturi (2013), which leads to a highly efficient parallelizable algorithm suitable for large-scale data analysis Peyré and Cuturi (2019). This regularization is defined by augmenting the standard optimal transportation problem by a penalization term based on relative entropy, defined between two probability measures α and β as $H(\alpha|\beta) = \int \log(\frac{d\alpha}{d\beta}(x))d\alpha(x)$ if α is absolutely continuous with respect to β, $\alpha \ll \beta$, and $+\infty$ otherwise. Given $P, Q \in \mathcal{P}(\mathbb{R}^d)$ and $\epsilon > 0$, the resulting problem reads

$$S_\epsilon(P,Q) = \min_{\pi \in \Pi(P,Q)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \tfrac{1}{2}\|x-y\|^2 d\pi(x,y) + \epsilon H(\pi | P \times Q), \tag{4.1}$$

where $\Pi(P,Q)$ denotes the set of couplings between P and Q.

Alongside its computational virtues, entropic regularization also brings substantial statistical benefits: entropically regularized transportation costs enjoy faster convergence rates than their unregularized counterparts, even in high dimensions, making them useful for estimation tasks Genevay et al. (2019); Mena and Niles-Weed (2019); Chizat et al. (2020); Pooladian and Niles-Weed (2021). Moreover, entropic regularization seems well suited to problems involving data corrupted with Gaussian noise Rigollet and Weed (2018a). Together, this body of results suggests the strengths of entropic optimal transport as an applied and theoretical statistical tool.

Obtaining limit theorems for the unregularized transportation costs is a longstanding question in probability theory and statistics. (See the recent work Hundrieser et al. (2022) for references and an account of the history of this problem.) Under relatively stringent assumptions on the measures, it is known that the empirical unregularized transport costs possesses asymptotically Gaussian fluctuations around its expectation del Barrio and Loubes (2019); del Barrio et al. (2021a); stronger results can be obtained when one or both of the measures is discrete del Barrio et al. (2021b); Sommerfeld and Munk (2018); Tameling et al. (2019), when the measures are smooth Manole et al. (2021), and in one dimension del Barrio et al. (1999, 2005).

The strict convexity and differentiability of the regularized optimal transportation problem makes it possible to prove significantly more general results. A central limit theorem for entropy regularized transportation costs, centered at the expectation of the empirical cost, was first obtained by Mena and Niles-Weed (2019) (see (4.2) below). Generalizations and extensions for discrete measures have been proved by Klatt et al. (2020); Bigot et al. (2019). A growing body of work investigates the properties of the entropy regularized optimal transport problem from the perspective of probability and analysis, including its asymptotic properties as $\epsilon \to 0$ Nutz and Wiesel (2021); Eckstein and Nutz (2021); Ghosal et al. (2021); Nutz and Wiesel (2022); Altschuler et al. (2022); Berman (2020); Chernozhukov et al. (2017), opening the door to further statistical applications of entropy regularised transport.

A crucial question in statistical applications of entropic optimal transport costs is the construction of asymptotic confidence intervals, to permit asymptotically valid inference.

The most general results known in this direction are due to Mena and Niles-Weed (2019), who showed that if P and Q are subgaussian probabilities on $\mathbb{R}^d$, then

$$\sqrt{n}(S_\epsilon(P_n, Q) - \mathbb{E}S_\epsilon(P_n, Q)) \xrightarrow{w} N(0, \mathrm{Var}_P(f_\epsilon^*)), \tag{4.2}$$

with $S_\epsilon(\cdot, \cdot)$ as in (4.1). (See Section 4.2 for further background and definitions.) A limitation of this result in practical inference problems is the centering at $\mathbb{E}S_\epsilon(P_n, Q)$ rather than at the population quantity $S_\epsilon(P, Q)$. This result parallels known results for the unregularized transport cost: del Barrio and Loubes (2019); del Barrio et al. (2021a) show that, under suitable technical conditions on P and Q, there exists $\sigma \geq 0$ such that

$$\sqrt{n}(W_p^p(P_n, Q) - \mathbb{E}W_p^p(P_n, Q)) \xrightarrow{w} N(0, \sigma^2)\,, \tag{4.3}$$

where W_p denotes the unregularized p-transportation cost, $p > 1$. See also González-Delgado et al. (2021) for its generalization to the flat torus. In this case, it is know that the centering at $\mathbb{E}W_p^p(P_n, Q)$ is unavoidable, and that it is not possible in general to replace $\mathbb{E}W_p^p(P_n, Q)$ by $W_p^p(P, Q)$, in view of the fact that known lower bounds on convergence rates of the Wasserstein distance imply that $\sqrt{n}(W_p^p(P_n, Q) - W_p^p(P, Q))$ is typically not stochastically bounded when $d > 2p$.

However, this limitation does *not* apply to the entropically regularized transport costs. Indeed, the results of Mena and Niles-Weed (2019) imply that

$$|S_\epsilon(P, Q) - \mathbb{E}S_\epsilon(P_n, Q)| \leq C_{P,Q} n^{-1/2} \tag{4.4}$$

for a positive constant $C_{P,Q}$ depending on the two measures. As a consequence, prior work does not rule out the possibility that $\sqrt{n}(S_\epsilon(P_n, Q) - S_\epsilon(P, Q))$ enjoys a central limit theorem, but nor does it provide a proof that such a theorem holds.

In this paper, we close this gap. We show a central limit theorem of the form

$$\sqrt{n}(S_\epsilon(P_n, Q) - S_\epsilon(P, Q)) \xrightarrow{w} N(0, \mathrm{Var}_P(f_\epsilon^*)), \tag{4.5}$$

valid for any subgaussian probabilities P and Q in any dimension. Prior to our work, such a bound was known only when P and Q were supported on a finite or countable set Klatt et al. (2020); Bigot et al. (2019). Also Harchaoui et al. (2020) provides central limit theorems for a different entropic regularization, where the solution is explicit. Our results represent a significant generalization of these results, and imply that, under sufficiently strong moment conditions, asymptitcally valid inference is always possible for the entropic transportation cost.

Our proof of (4.5) is based on an important strengthening of (4.4). Specifically, we show that, for subgaussian probability measures,

$$|\mathbb{E}S_\epsilon(P_n, Q) - S_\epsilon(P, Q)| = o(n^{-1/2}) \quad \text{as } n \to \infty. \tag{4.6}$$

Combining this result with (4.2) yields (4.5).

When P and Q are supported on a bounded set Ω, we are able to obtain substantially more precise results, which are of independent interest. Our techniques imply that for compactly supported P and Q,

$$|\mathbb{E}S_\epsilon(P_n, Q) - S_\epsilon(P, Q)| \leq C_{P,Q} n^{-1} .$$

(See Remark 4.3.5.) This result implies that the bias of $S_\epsilon(P_n, Q)$ decays at the fast n^{-1} rate, thereby recovering the rate typically obtained for *parametric* estimation problems. Our proof also yields new sample complexity results for the Sinkhorn divergence, defined as $D_\epsilon(P, Q) = S_\epsilon(P, Q) - \frac{1}{2}(S_\epsilon(P, P) + S_\epsilon(Q, Q))$. For probability measures on compact sets, convergence in Sinkhorn divergence is equivalent to weak convergence Feydy et al. (2019), implying that $D_\epsilon(P_n, P) \to 0$ a.s. In Theorem 4.5, we show the quantitative bound

$$\mathbb{E}D_\epsilon(P_n, P) \leq C_P n^{-1}, \tag{4.7}$$

valid for all compactly supported P. This convergence rate could have been anticipated from known distributional limits for Sinkhorn divergences between finitely supported measures Bigot et al. (2019); Klatt et al. (2020), but was unknown prior to our work.

In the bounded case, these results are all derived as corollaries of new convergence results for the *optimal dual potentials* in the entropic transport problem. In Theorem 4.4.5, we prove that, when P and Q are bounded, the entropic potentials converge fast in Hölder norm:

$$\mathbb{E}\|g_n - g^*\|^2_{\mathcal{C}^s(\Omega)},\ \mathbb{E}\|f_n - f^*\|^2_{\mathcal{C}^s(\Omega)} \leq C_{P,Q} n^{-1} , \tag{4.8}$$

where $s = [d/2] + 1$. We prove this result, as well as its two-sample analogue, in Section 4.4. To our knowledge, these bounds are new, even for finitely supported probability measures. When P and Q are not necessarily bounded but have subgaussian tails, we prove a non-quantitative analogue of (4.8), showing that f_n and g_n converge to f^* and g^* almost surely in a suitably strong topology. This result is a strengthening of a similar convergence result obtained by Mena and Niles-Weed (2019).

The remaining sections of this paper are organized as follows. Section 4.2 provides some background results on entropic transportation costs. The central limit theorem (4.5) and the faster rate (4.6) are given in Section 4.3. Section 4.4 contains the announced results about the convergence rates of the potentials. The bounds for Sinkhorn divergences are proved in Section 4.5. Finally we include a section with some numerical illustration of our limit theorems.

4.2 Preliminaries on entropic transportation costs

This selection collects several background results on the entropic transportation problem (4.1).

We say that a distribution ν is the pushforward by a map T of a distribution μ, if $\nu = \mu \circ T^{-1}$. A simple computation shows that if P^ε and Q^ε denote the pushforwards of P and Q under the map $x \mapsto \varepsilon^{-\frac{1}{2}} x$ then $S_\epsilon(P, Q) = \varepsilon S_1(P^\varepsilon, Q^\varepsilon)$. Hence, we focus on the case $\varepsilon = 1$ and write simply $S(P,Q)$ instead of $S_1(P,Q)$. The minimisation problem (4.1) admits a dual formulation. In fact, if $\pi \in \Pi(P,Q)$ and $r = \frac{d\pi}{d(P\times Q)}$, then, for any $f \in L_1(P)$, $g \in L_1(Q)$

$$\int [\tfrac{1}{2}\|x-y\|^2 + \log r(x,y)] r(x,y) dP(x) dQ(y) \geq \int f(x) dP(x) + \int g(y) dQ(y) \\ - \int e^{f(x)+g(y)-\frac{1}{2}\|x-y\|^2} dP(x) dQ(y) + 1,$$

with equality if and only if $r(x,y) = e^{f(x)+g(y)-\frac{1}{2}\|x-y\|^2}$ $P\times Q$-almost surely. (This follows from the elementary fact that $s\log s \geq s-1$, $s>0$, with equality if and only if $s=1$). This inequality implies the following version of weak duality:

$$S(P,Q) \geq \sup_{f\in L_1(P), g\in L_1(Q)} \Big\{ \int_{\mathbb{R}^d} f(x) dP(x) + \int_{\mathbb{R}^d} g(y) dQ(y) \\ - \int_{\mathbb{R}^d\times\mathbb{R}^d} e^{f(x)+g(y)-\frac{1}{2}\|x-y\|^2} dP(x) dQ(y) + 1 \Big\}.$$

It shows also that if $\frac{d\pi}{d(P\times Q)} = e^{f(x)+g(y)-\frac{1}{2}\|x-y\|^2}$ for some $f \in L_1(P_1)$ and $g \in L_1(Q)$, then π is a minimizer for the entropic transportation problem (indeed, by the strict convexity of H, it is the unique minimizer). The theory of entropic optimal transportation (see Csiszar (1975); Nutz (2021)) shows that the last inequality is, in fact, an equality, namely,

$$S(P,Q) = \sup_{f\in L_1(P), g\in L_1(Q)} \Big\{ \int_{\mathbb{R}^d} f(x) dP(x) + \int_{\mathbb{R}^d} g(y) dQ(y) \\ - \int_{\mathbb{R}^d\times\mathbb{R}^d} e^{f(x)+g(y)-\frac{1}{2}\|x-y\|^2} dP(x) dQ(y) + 1 \Big\}. \tag{4.9}$$

Maximizing pairs in (4.9) are called optimal potentials. These optimal potentials exist and satisfy some regularity conditions under integrability assumptions on P and Q.

Following the framework in Mena and Niles-Weed (2019), we say that a probability P is σ^2-subgaussian if $\mathbb{E}\big(e^{\frac{\|X\|^2}{2d\sigma^2}}\big) \leq 2$ when $X \sim P$. When P and Q are subgaussian there exist optimal potentials, denoted by f^*, g^*, satisfying the *optimality conditions*, i.e.

$$\int e^{f^*(x)+g^*(y)-\frac{1}{2}\|x-y\|^2} dQ(y) = 1, \quad \text{for all } x \in \mathbb{R}^d, \\ \int e^{f^*(x)+g^*(y)-\frac{1}{2}\|x-y\|^2} dP(x) = 1, \quad \text{for all } y \in \mathbb{R}^d, \tag{4.10}$$

see Proposition 6 in Mena and Niles-Weed (2019). Moreover, the pair (f^*, g^*) satisfying (4.10) is unique up to constant shifts, and is uniquely specified by adopting the normalization convention

$$\int f^*(x) dP(x) = \int g^*(y) dQ(y). \tag{4.11}$$

In what follows, we tacitly assume that (4.11) holds unless we explicitly specify an alternate convention.

The above considerations imply that the minimizer in the primal formulation is

$$d\pi^* = e^{f^*(x)+g^*(y)-\frac{1}{2}\|x-y\|^2} dQ(y)dP(x),$$

where f^* and g^* satisfy

$$\begin{aligned} f^*(x) &= -\log\left(\int e^{g^*(y)-\frac{1}{2}\|x-y\|^2} dQ(y)\right), \\ g^*(y) &= -\log\left(\int e^{f^*(y)-\frac{1}{2}\|x-y\|^2} dP(y)\right). \end{aligned} \tag{4.12}$$

Let $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$ be a multi-index. If $P, Q \in \mathcal{P}(\mathbb{R}^d)$ are σ^2-subgaussian then (see Proposition 1 in Mena and Niles-Weed (2019)), the optimal potential f^* specified above is such that

$$|D^\alpha(f^* - \tfrac{1}{2}\|\cdot\|^2)(x)| \leq C_{k,d} \begin{cases} 1+\sigma^4 & \text{if } k=0 \\ \sigma^k(\sigma+\sigma^2)^k & \text{otherwise,} \end{cases} \quad \text{if } \|x\| \leq \sqrt{d}\sigma, \tag{4.13}$$

$$|D^\alpha(f^* - \tfrac{1}{2}\|\cdot\|^2)(x)| \leq C_{k,d} \begin{cases} 1+(1+\sigma^2)\|x\|^2 & \text{if } k=0 \\ \sigma^k(\sqrt{\sigma\|x\|}+\sigma\|x\|)^k & \text{otherwise,} \end{cases} \text{if } \|x\| \geq \sqrt{d}\sigma, \tag{4.14}$$

and likewise for g^*, where in both cases $k := |\alpha|$, and the constant $C_{k,d}$ depends only on d and k.

Throughout the paper, we will assume that $P, Q \in \mathcal{P}(\mathbb{R}^d)$ are σ^2-subgaussian probabilities and $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ are independent samples of i.i.d r.v.'s with laws P and Q, respectively. We will denote by P_n and Q_m the associated empirical measures. We will require that the measures P_n and Q_n are also subgaussian, which is guaranteed by the following result, which summarizes Lemma 2 and Lemma 4 in Mena and Niles-Weed (2019).

Lemma 4.2.1. *Let $X_1, \ldots, X_n$ be i.i.d random variables with σ^2-subgaussian law $P \in \mathcal{P}(\mathbb{R}^d)$ and let P_n be the associated empirical measure. Then, there exists a random variable $\tilde{\sigma}$, such that*

(i) for every $n \in \mathbb{N}$, the probabilities P and P_n are uniformly $\tilde{\sigma^2}$-subgaussian almost surely,

(ii) for any $k \in \mathbb{N}$, we have $\mathbb{E}\left(\tilde{\sigma}^{2k}\right) \leq 2k^k\sigma^{2k}$.

4.3 An improved central limit theorem for subgaussian probability measures

This section shows that, for subgaussian probability measures, the expected empirical entropic transportation cost converges to its population counterpart with rate $o(n^{-1/2})$. This is an improvement over the bound (4.4) derived in Mena and Niles-Weed (2019) and has, as a main consequence, a CLT for the empirical entropic transportation cost with the natural centering constants (see Theorem 4.3.6 below), which, in turn, yields an asymptotically valid confidence interval for $S_\varepsilon(P,Q)$ regardless the dimension, d.

Let s be a nonnegative integer. To prove the main result in this section, we introduce the class $\mathcal{G}^s(C)$, consisting of all $f \in \mathcal{C}^s(\mathbb{R}^d)$ such that

$$\begin{aligned} &|f(x)| \leq C(1+\|x\|^3), \\ &|D^\alpha f(x)| \leq C(1+\|x\|^{s+1}), \ \ |\alpha| \leq s. \end{aligned} \tag{4.15}$$

Our next results gives an estimate of the complexity of this class, in terms of covering numbers with respect to the random metric $L_2(P_n)$. The proof can be easily adapted from the proof of Proposition 3 in Mena and Niles-Weed (2019). We omit further details.

Lemma 4.3.1. *Assume $\mathcal{G}^s(C)$ is as above. If $X_1,\dots,X_n$ are i.i.d random variables with σ^2-subgaussian law $P \in \mathcal{P}(\mathbb{R}^d)$, P_n is the associated empirical measure and $L = \frac{1}{n}\sum_{i=1}^n e^{-\|X_i\|^2/(4d\sigma^2)}$ then, for a constant $C_{s,d}$ depending only on s and d,*

$$\log \mathcal{N}(\epsilon, \mathcal{G}^s(C), L_2(P_n)) \leq C_{s,d} L^{\frac{d}{2s}} \epsilon^{\frac{-d}{s}} (1+\sigma^d)(1+\sigma^s)^{\frac{d}{s}}. \tag{4.16}$$

Finally, we introduce the space $\mathcal{G}^s = \bigcup_{C\geq 0} \mathcal{G}^s(C)$ endowed with the norm

$$\|f\|_s = \left\| \frac{f}{1+\|\cdot\|^3} \right\|_\infty + \sum_{i=1}^{s} \sum_{|k|=i} \left\| \frac{D^k f}{1+\|\cdot\|^{s+1}} \right\|_\infty$$

Let $(\mathcal{G}^s)'$ denote the dual space of $\mathcal{G}^s$, endowed with the dual norm

$$\|G\|_s' = \sup_{f\in\mathcal{G}^s,\ \|f\|_s\leq 1} |G(f)|.$$

With these ingredients we are ready to prove the main technical result of this section, from which we obtain the CLT for the entropic transportation cost with natural centering constants (Theorem 4.3.6 below).

Lemma 4.3.2. *If $P, Q \in \mathcal{P}(\mathbb{R}^d)$ be σ^2-subgaussian probabilities, then*

$$\sqrt{n}\, |\mathbb{E}S(P_n,Q) - S(P,Q)| \to 0. \tag{4.17}$$

Moreover, if $m = m(n)$ and $\lambda := \lim_{n\to\infty} \frac{n}{n+m} \in (0,1)$, then

$$\sqrt{\frac{nm}{n+m}}\, |\mathbb{E}S(P_n,Q_m) - S(P,Q)| \to 0. \tag{4.18}$$

Proof. Let $(f_n, g_n) \in L_1(P_n) \times L_1(Q)$ be the unique pair of optimal potentials satisfying (4.10) and (4.11) for P_n, Q. As noted above, by Proposition 6 in Mena and Niles-Weed (2019), this pair satisfies (4.13) and (4.14). We observe that, by optimality of the potentials,

$$S(P,Q) \geq \int_{\mathbb{R}^d} f_n(x)dP(x)+\int_{\mathbb{R}^d} g_n(y)dQ(y)-\int_{\mathbb{R}^d\times\mathbb{R}^d} e^{f_n(x)+g_n(y)-\frac{1}{2}\|x-y\|^2}dP(x)dQ(y)+1,$$

which yields

$$\begin{aligned}0 \leq \sqrt{n}\,(\mathbb{E}S(P_n,Q)-S(P,Q)) \leq & \mathbb{E}\int_{\mathbb{R}^d} f_n(x)\sqrt{n}(dP_n-dP)(x)\\ & -\mathbb{E}\int_{\mathbb{R}^d\times\mathbb{R}^d} e^{f_n(x)+g_n(y)-\frac{1}{2}\|x-y\|^2}\sqrt{n}(dP_n-dP)(x)dQ(y).\end{aligned}$$

Now the optimality condition

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} e^{f_n(x)+g_n(y)-\frac{1}{2}\|x-y\|^2}dQ(y)=1, \text{ for all } x\in\mathbb{R}^d$$

implies that

$$0 \leq \sqrt{n}\,(\mathbb{E}S_1(P_n,Q)-S_1(P,Q)) \leq \mathbb{E}\int_{\mathbb{R}^d} f_n(x)\sqrt{n}(dP_n-dP)(x).$$

Set $s=[d/2]+1$ and let $(f^*,g^*)\in L_1(P)\times L_1(Q)$ be the unique pair of optimal potentials satisfying (4.10) and (4.11) for P,Q. Since $\mathbb{E}\int_{\mathbb{R}^d} f^*(x)\sqrt{n}(dP_n-dP)(x)=0$,

$$0 \leq \sqrt{n}\,(\mathbb{E}S_1(P_n,Q)-S_1(P,Q)) \leq \mathbb{E}\int_{\mathbb{R}^d}\{f_n(x)-f^*(x)\}\sqrt{n}(dP_n-dP)(x)$$

We write now $\mathbb{G}_n$ for the empirical process indexed by $\mathcal{G}^s$, that is, $\mathbb{G}_n(f)=\sqrt{n}(P_n(f)-P(f))$, $f\in\mathcal{G}^s$, and note that

$$\begin{aligned}|\mathbb{G}_n(f)| &\leq \sqrt{n}(P_n+P)(|f|) \leq \left\|\frac{f}{1+\|\cdot\|^3}\right\|_\infty \sqrt{n}(P_n+P)(1+\|\cdot\|^3)\\ &\leq \sqrt{n}(P_n+P)(1+\|\cdot\|^3)\|f\|_s \leq \sqrt{n}(2+48(d\tilde{\sigma}^2)^3)\|f\|_s,\end{aligned}$$

where the last inequality comes from Lemma 1 in Mena and Niles-Weed (2019). Consequently, we deduce that $\mathbb{G}_n$ belongs to the dual space $(\mathcal{G}^s)'$, for all $n\in\mathbb{N}$, and we get the bound

$$\sqrt{n}\,(\mathbb{E}S_1(P_n,Q)-S_1(P,Q)) \leq \mathbb{E}\left\{\|\mathbb{G}_n\|_s'\|f^*-f_n\|_s\right\}.$$

Using Cauchy–Schwarz's inequality we see that

$$\sqrt{n}\,(\mathbb{E}S_1(P_n,Q)-S_1(P,Q)) \leq \sqrt{\mathbb{E}\|\mathbb{G}_n\|_s'^2\,\mathbb{E}\|f^*-f_n\|_s^2}. \tag{4.19}$$

Note that $\|\mathbb{G}_n\|'_s$ is the sup taken on the unit ball of $\mathcal{G}^s$, which is contained in $\mathcal{G}^s(1)$. We can conclude, by using (4.16) and Theorem 3.5.1 and Exercise 2.3.1 in Giné and Nickl (2015), that there exists a constant $C_{s,d} > 0$ such that

$$\mathbb{E}\|\mathbb{G}_n\|'^2_s \leq C_{s,d}\mathbb{E}\left(\int_0^{\max_{\|f\|_s\leq 1}\|f\|_{L_2(P_n)}} \sqrt{L^{\frac{d}{2s}}\epsilon^{\frac{-d}{s}}(1+\sigma^d)(1+\sigma^s)^{\frac{d}{s}}}d\epsilon\right)^2$$
$$\leq (1+\sigma^d)(1+\sigma^s)^{\frac{d}{s}}C_{s,d}\mathbb{E}\left(\int_0^{1+4\sqrt{3}d^{3/2}\tilde{\sigma}^3} L^{\frac{d}{4s}}\epsilon^{\frac{-d}{2s}}d\epsilon\right)^2$$
$$\leq C'_{s,d}(1+\sigma^{2d})\mathbb{E}L^{\frac{d}{2s}}(1+\tilde{\sigma}^3)^{\frac{2s-d}{s}},$$

where we have used first Lemma 1 in Mena and Niles-Weed (2019) to bound

$$\max_{\|f\|_s\leq 1}\|f\|_{L_2(P_n)} \leq 1 + \left(\int \|x\|^6 dP_n(x)\right)^{1/2} \leq 1 + 4\sqrt{3}d^{3/2}\tilde{\sigma}^3$$

and then the fact that $s = [d/2]+1$. Using the Cauchy-Schwarz inequality we see that

$$\mathbb{E}L^{\frac{d}{2s}}(1+\tilde{\sigma}^3)^{\frac{2s-d}{s}} \leq \sqrt{\mathbb{E}L^{\frac{d}{s}}\mathbb{E}(1+\tilde{\sigma}^3)^{\frac{2(2s-d)}{s}}}$$

where we can use the fact that $\mathbb{E}L < C$ for a positive constant C independent of n and Lemma 4.2.1 to conclude that $\limsup \mathbb{E}\|\mathbb{G}_n\|'^2_s < \infty$.

To deal with the second term in (4.19) we denote $\Delta_n = f^* - f_n$. We prove next that $\|\Delta_n\|_s \to 0$ almost surely, and then that it is dominated by a random variable with finite second moment. Together, these facts imply that $\mathbb{E}\|f^* - f_n\|_s^2 \to 0$ and conclude the proof. The first claim is given by the following result.

Lemma 4.3.3. *Let $P, Q \in \mathcal{P}(\mathbb{R}^d)$ be σ-subgaussian probabilities, and P_n, Q_n associated empirical measures. Then, the optimal transport potentials (f_n, g_n) for P_n, Q_n satisfy $\|f_n - f^*\|_s \to 0$ and $\|g_n - g^*\|_s \to 0$ almost surely.*

Proof. We prove the result for f_n, with the same conclusion following for g_n by symmetry. First, we use induction to prove convergence of the derivatives up to order s. We follow classical arguments in real analysis, see Rudin (1987):

(1) *For $J = 0$,* Proposition 4 in Mena and Niles-Weed (2019) shows that, almost surely, $\Delta_n := f^* - f_n \to 0$ uniformly in compact sets.

(2) *Assume that for every k with $|k| \leq J-1$, we have $D^k\Delta_n \to 0$, uniformly in compact sets.* Let $k = (k_1, \ldots, k_d)$ be such that $|k| = J$ and let $B_R \subset \mathbb{R}^d$ be the ball of radius R centered at 0. Using the fact that all the derivatives of $D^k\Delta_n$ are bounded and $D^k\Delta_n$ is itself pointwise bounded, see Proposition 1 and Lemma 2 in Mena and Niles-Weed (2019), we derive that the sequence $D^k\Delta_n$ is equicontinuous and bounded for all points. We can then apply the Arzelà-Ascoli theorem on B_R to deduce that, up to subsequences, $D^k\Delta_n \to \Delta^k$ uniformly on B_R. Suppose, without losing generality, that $k_1 \geq 1$, set $k' = (k_1 - 1, \dots, k_d)$ and note that

$$D^{k'}\Delta_n(x) = \int_0^{x_1} D^k\Delta_n(t, , x_2, \dots, x_d)dt + D^{k'}\Delta_n(0, x_2, \dots, x_d),$$

which implies that

$$\begin{aligned}&|D^{k'}\Delta_n(x) - \int_0^{x_1} \Delta^k(t, , x_2, \dots, x_d)dt|\\ &\leq \int_0^{x_1} |D^k\Delta_n(t, , x_2, \dots, x_d) - \Delta^k(t, , x_2, \dots, x_d)|dt + |D^{k'}\Delta_n(0, x_2, \dots, x_d)|.\end{aligned}$$

As a consequence,

$$\begin{aligned}&\sup_{x\in B_R} |D^{k'}\Delta_n(x) - \int_0^{x_1} \Delta^k(t, , x_2, \dots, x_d)dt|\\ &\leq R \sup_{x\in B_R} |D^k\Delta_n(x) - \Delta^k(x)| + |D^{k'}\Delta_n(0, x_2, \dots, x_d)| \to 0,\end{aligned}$$

where the limit follows from the induction hypothesis (recall that $\sup_{x\in B_R} |D^{k'}\Delta_n(x)| \to 0$). By uniqueness of the limit we conclude that $0 = \int_0^x \Delta^k dx_1$, which implies that $\Delta^k = 0$. By taking $R \to \infty$ we conclude that $D^k\Delta_n \to 0$ uniformly on the compact sets of $\mathbb{R}^d$.

To show convergence in the norm $\|\cdot\|_s$, it suffices to show that for any $\epsilon > 0$, there exists an n_0 such that $\|\Delta_n\|_s \leq \epsilon$ for all $n \geq n_0$. Recall that by Lemma 4.2.1 (i) and Proposition 1 in Mena and Niles-Weed (2019), there exists an almost surely finite random variable $\tilde{\sigma}$ and a constant $K_{s,d}$ such that for all $n \in \mathbb{N}$ and $x \in \mathbb{R}^d$,

$$\begin{aligned}\frac{|\Delta_n(x)|}{1 + \|x\|^3} &\leq K_{s,d}\frac{1 + \tilde{\sigma}^4}{1 + \|x\|}\\ \frac{|D^k\Delta_n(x)|}{1 + \|x\|^{s+1}} &\leq K_{s,d}\frac{1 + \tilde{\sigma}^{3s}}{1 + \|x\|} \qquad \forall |k| \leq s.\end{aligned} \tag{4.20}$$

We obtain that there exists a finite random variable $\tilde{K}$ such that

$$\frac{|\Delta_n(x)|}{1 + \|x\|^3} + \sum_{i=1}^s \sum_{|k|=i} \frac{|D^k\Delta_n(x)|}{1 + \|x\|^{s+1}} \leq \epsilon/2 \qquad \forall \|x\| > \tilde{K}\epsilon^{-1}, n \geq 0\,.$$

Since Δ_n and $D^k\Delta_n$ converge uniformly to zero on the compact set $\{x \in \mathbb{R}^d : \|x\| \leq c_{d,s}\tilde{K}\epsilon^{-1}\}$, there exists an n_0 for which

$$\frac{|\Delta_n(x)|}{1 + \|x\|^3} + \sum_{i=1}^s \sum_{|k|=i} \frac{|D^k\Delta_n(x)|}{1 + \|x\|^{s+1}} \leq \epsilon/2 \qquad \forall \|x\| \leq \tilde{K}\epsilon^{-1}, n \geq n_0\,.$$

Combining these claims, we obtain that $\|\Delta_n\|_s \leq \epsilon$ for all $n \geq n_0$, as desired. □

To complete the proof of Lemma 4.3.2 it only remains to prove that $\|f_n - f^*\|_s$ can be dominated by a random variable with finite second moment. We have from (4.20) above that $\|\Delta_n\|_s^2 \leq K'_{s,d}(1+\tilde{\sigma}^{3s})^2$ for some constant $K'_{s,d}$. It only remains to show that $\mathbb{E}\left(1+\tilde{\sigma}^{3s}\right)^2 < \infty$. But Lemma 4.2.1 implies that all moments of $\tilde{\sigma}$ are finite, which completes the proof.

To deal with the two-sample case, we split the difference as follows

$$\sqrt{\frac{nm}{n+m}}|\mathbb{E}S_1(P_n,Q_m) - S_1(P,Q)|$$
$$\leq \sqrt{\frac{nm}{n+m}}|\mathbb{E}S_1(P_n,Q_m) - S_1(P,Q_m)| + \sqrt{\frac{nm}{n+m}}|\mathbb{E}S_1(P,Q_m) - S_1(P,Q)|.$$

The second term tends to 0 by using (4.17). For the first one we denote $g_{n,m}$ a potential of $S_1(P_n,Q_m)$ and g_m a potential of $S_1(P,Q_m)$. Applying (4.19) we derive

$$\sqrt{m}|\left(\mathbb{E}S(P_n,Q_m) - S(P,Q_m)\right)| \leq \sqrt{\mathbb{E}\|F_m\|_s'^2\,\mathbb{E}\|g_{n,m} - g_m\|_s^2}$$
$$\leq 2\sqrt{\mathbb{E}\|F_m\|_s'^2\left(\mathbb{E}\|g_{n,m} - g^*\|_s^2 + \mathbb{E}\|g_m - g^*\|^2\right)}.$$

We conclude using Lemma 4.3.3 (which can be trivially adapted to this setup) and the subsequent argument. □

As a consequence of Lemma 4.3.2 by simply considering the change of variables $x \mapsto x\epsilon^{-\frac{1}{2}}$ (recall the comments at the beginning of this section) we obtain the generalization to any $\epsilon > 0$.

Corollary 4.3.4. *Let $P, Q \in \mathcal{P}(\mathbb{R}^d)$ be σ-subgaussian probabilities and P_n, Q_m associated empirical measures. Then*

$$\sqrt{n}\,|\mathbb{E}S_\epsilon(P_n,Q) - S_\epsilon(P,Q)| \to 0$$

and

$$\sqrt{\frac{nm}{n+m}}\,|\mathbb{E}S_\epsilon(P_n,Q_m) - S_\epsilon(P,Q)| \to 0.$$

As announced, Corollary 4.3.4 improves over Corollary 1 in Mena and Niles-Weed (2019), which implied $|\mathbb{E}S_\epsilon(P_n,Q) - S_\epsilon(P,Q)| = O(n^{-1/2})$ rather than $|\mathbb{E}S_\epsilon(P_n,Q) - S_\epsilon(P,Q)| = o(n^{-1/2})$.

Remark 4.3.5. *In some cases we can go much further in this direction. In fact, if P and Q are compactly supported then (see Theorem 4.4.5 below)*

$$\mathbb{E}\|f_n - f^*\|_{\mathcal{C}^s(\Omega)}^2 \leq \frac{c_\Omega}{n}$$

for some constant c_Ω*. Plugging this into* (4.19) *and using again the fact that* $\limsup_n \mathbb{E}\|\mathbb{G}_n\|_s'^2 < \infty$ *we conclude that*

$$\big|\mathbb{E}S_\epsilon(P_n, Q) - S_\epsilon(P, Q)\big| \leq \frac{C_\Omega}{n} \tag{4.21}$$

for some constant $C_\Omega > 0$*. A similar conclusion holds for the two-sample problem. Whether this improved rate remains valid for general subgaussian probabilities is an open question.*

The following central limit becomes a direct consequence of Theorem 3 in Mena and Niles-Weed (2019), which shows that the fluctuations around the mean are asymptotically Gaussian, i.e.

$$\sqrt{n}(S_\epsilon(P_n, Q) - \mathbb{E}S_\epsilon(P_n, Q)) \xrightarrow{w} N(0, \mathrm{Var}_P(f_\epsilon^*)). \tag{4.22}$$

Here $(f_\epsilon^*, g_\epsilon^*)$ are optimal potentials for $S_\epsilon(P, Q)$. We observe that, while the pair of optimal potentials is not uniquely defined, it follows from the uniqueness of the minimizer in (4.1) that if $(\tilde{f}_\epsilon, \tilde{g}_\epsilon)$ is another pair of optimal potentials then $\tilde{f}_\epsilon = f_\epsilon^* + K$ P-a.s. for some constant K. Hence, $\mathrm{Var}_P(f_\epsilon^*)$ is well defined in the sense that it does not depend on the choice of optimal potential.

Theorem 4.3.6. *Let* $P, Q \in \mathcal{P}(\mathbb{R}^d)$ *be* σ*-subgaussian probabilities, then*

$$\sqrt{n}(S_\epsilon(P_n, Q) - S_\epsilon(P, Q)) \xrightarrow{w} N(0, \mathrm{Var}_P(f_\epsilon^*)),$$

where $(f_\epsilon^*, g_\epsilon^*)$ *are optimal potentials for* $S_\epsilon(P, Q)$*. Moreover, if* $\lambda := \lim_{n,m\to\infty} \frac{n}{n+m} \in (0, 1)$*,*

$$\sqrt{\frac{nm}{n+m}}(S_\epsilon(P_n, Q_m) - S_\epsilon(P, Q)) \xrightarrow{w} N(0, (1-\lambda)\,\mathrm{Var}_P(f_\epsilon^*) + \lambda\,\mathrm{Var}_Q(g_\epsilon^*)).$$

One important advantage of Theorem 4.3.6 over (4.22) is that it can be exploited for inferential purposes. For instance, it enables to build confidence intervals for $S_\epsilon(P, Q)$ as follows.

Note that we can estimate the asymptotic variance in the one-sample CLT by

$$\hat{\sigma}_n^2 := \mathrm{Var}_{P_n}(f_{n,\epsilon}) = \frac{1}{n}\sum_{i=1}^n f_{n,\epsilon}^2(X_i) - \Big(\frac{1}{n}\sum_{i=1}^n f_{n,\epsilon}(X_i)\Big)^2, \tag{4.23}$$

where $(f_{n,\epsilon}, g_{n,\epsilon})$ is a pair of optimal potentials for $S_\epsilon(P_n, Q)$. It follows from the proof of Lemma 4.3.2 that $\mathbb{E}\|f_{n,\epsilon} - f_\epsilon^*\|_s^2 \to 0$. Hence, $\|f_{n,\epsilon} - f_\epsilon^*\|_s \to 0$ in probability. Using the elementary bound $|a^2 - b^2 \leq |a-b|^2 + 2|b||a-b|$ we see that that

$$\Big|\frac{1}{n}\sum_{i=1}^n f_{n,\epsilon}^2(X_i) - \frac{1}{n}\sum_{i=1}^n f_\epsilon^{*2}(X_i)\Big| \leq (\|f_{n,\epsilon} - f_\epsilon^*\|_s^2 + 2\|f_\epsilon^*\|_s\|f_{n,\epsilon} - f_\epsilon^*\|_s)\frac{1}{n}\sum_{i=1}^n (1 + \|X_i\|^3).$$

Since $\frac{1}{n}\sum_{i=1}^n f_\epsilon^{*2}(X_i) \to \mathbb{E}_P(f_\epsilon^{*2})$ a.s. we conclude that $\frac{1}{n}\sum_{i=1}^n f_{n,\epsilon}^2(X_i) \to \mathbb{E}_P(f_\epsilon^{*2})$ in probability and, arguing similarly for $\frac{1}{n}\sum_{i=1}^n f_{n,\epsilon}(X_i)$, that

$$\hat{\sigma}_n^2 \to \mathrm{Var}_P(f_\epsilon^*) \quad \text{in probability.}$$

We conclude that

$$\frac{\sqrt{n}}{\hat{\sigma}_n}(S_\epsilon(P_n, Q) - S_\epsilon(P, Q)) \xrightarrow{w} N(0,1)$$

and, as a consequence, that, writing z_β for the β quantile of the standard normal distribution,

$$\left[S_\epsilon(P_n, Q) \pm \frac{\hat{\sigma}_n}{\sqrt{n}} z_{1-\alpha/2}\right] \tag{4.24}$$

is a confidence interval for $S_\epsilon(P, Q)$ of asymptotic level $1-\alpha$. A similar confidence interval can be constructed from the two-sample statistic. Such results will be illustrated in the Simulations Section.

4.4 Convergence rates for optimal potentials

The goal of this section is to prove a bound on the difference between empirical potentials and their population counterparts. In this section we assume that both measures, P, Q, are supported in a compact set $\Omega \subset \mathbb{R}^d$. By translation invariance of the optimal transport problem, we may assume without loss of generality that $0 \in \Omega$. We write D_Ω for the diameter of Ω and let (f^*, g^*) be a pair of optimal potentials (maximizers of (4.9) for P and Q) and (f_n, g_n) their empirical counterpart (maximizers of (4.9) for P_n and Q). As noted above, these optimal potentials are unique up to an additive constant. In this section, we adopt the following normalization convention:

$$\int g^*(y)dQ(y) = \int g_n(y)dQ(y) = 0, \tag{4.25}$$

We show below that derivatives af the optimal potentials are uniformly bounded (see Lemma 4.4.1). Additionally, the choice of optimal potentials in (4.25) allows to control uniformly the optimal potentials, as we show in Lemma 4.4.4. These are key ingredients for the aim of the section, namely, showing that the convergence rate of f_n (resp. g_n) towards f^* (resp. g^*), is $O_p\left(\frac{1}{\sqrt{n}}\right)$.

The optimal potentials belong to the space $\mathcal{C}^s(\Omega)$, for $s = \lceil \frac{d}{2} \rceil + 1$, in which we consider the norm

$$\|f\|_{\mathcal{C}^s(\Omega)} = \sum_{i=0}^{s} \sum_{|\alpha|=i} \|D^\alpha f\|_\infty .$$

In this section, we use the notation $c_{d,s}, c'_{d,s}, \dots$ to indicate unspecified positive constants depending on d and s whose value may change from line to line. The optimality conditions (4.12) imply the following bounds (see Proposition 1 in Genevay et al. (2019)).

Lemma 4.4.1. *Let $\Omega \subset \mathbb{R}^d$ be a compact set and $P, Q \in \mathcal{P}(\Omega)$. Then the optimal potentials (f^*, g^*) satisfy:*

(i) $\min_{y\in\Omega}\{\frac{1}{2}\|x-y\|^2 - g^*(y)\} \le f^*(x) \le \max_{y\in\Omega}\{\frac{1}{2}\|x-y\|^2 - g^*(y)\}$,

(ii) $f^*(x)$ *is* D_Ω*-Lipschitz,*

(iii) $f^* \in \mathcal{C}^\infty(\Omega)$ *and* $\|D^\alpha f^*\|_\infty \leq C_{d,\alpha}(1 + D_\Omega^{|\alpha|})$ *for all multi-indices* α *with* $|\alpha| \geq 1$, *for some constant* $C_{d,\alpha}$ *depending only on* d *and* α.

Proof. The first two claims are proven in Proposition 1 of Genevay et al. (2019), so it suffices to consider the last claim. We prove it for f^*, the case of g^* being similar. Define $\bar{f}^*(x) = f^*(x) - \frac{1}{2}\|x\|^2$. As in the proof of Proposition 1 in Mena and Niles-Weed (2019), the Faà di Bruno formula yields

$$-D^\alpha \bar{f}^*(x) = \sum_{\beta_1+\cdots+\beta_k=\alpha} \lambda_{\alpha,\beta_1,\ldots,\beta_k} \prod_{j=1}^k \mu_{\beta_j,g}, \quad \text{for all } x \in \Omega,$$

where $\lambda_{\alpha,\beta_1,\ldots,\beta_k}$ are combinatorial quantities and for a multi-index β we define

$$\begin{aligned}\mu_{\beta,g} &= \frac{\int y^\beta e^{g^*(y)-\frac{1}{2}\|y\|^2+\langle x,y\rangle} dQ(y)}{\int e^{g^*(y)-\frac{1}{2}\|y\|^2+\langle x,y\rangle} dQ(y)} \\ &= \int y^\beta e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2} dQ(y) \\ &= \int \prod_{i=1}^d y_i^{\beta_i} e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2} dQ(y).\end{aligned}$$

By the optimality condition (4.12), $\int e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2} dQ(y) = 1$. As a consequence, there exists $C'_{d,\alpha}$ such that $\|D^\alpha \bar{f}^*\|_\infty \leq C'_{d,\alpha} D_\Omega^{|\alpha|}$. Since $\|D^\alpha \frac{1}{2}\|x\|^2\|_\infty \leq 1 + D_\Omega$ for $|\alpha| \geq 1$, we obtain $\|D^a f^*\| \leq C'_{d,\alpha} D_\Omega^\alpha + 1 + D_\Omega \leq C_{d,\alpha}(1 + D_\Omega^{|\alpha|})$. □

Remark 4.4.2. *Since the probabilities* P_n *and* Q_n *are also supported on the same compact set* Ω, *Lemma 4.4.1 holds also for* f_n *and* g_n.

We also obtain bounds on the derivatives of $e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}$.

Lemma 4.4.3. *For any multi-index* β, *the function* $x \mapsto x^\beta e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}$ *has* $\mathcal{C}^s(\Omega)$ *norm at most* $c_{d,s} e^{D_\Omega^2}(1 + D_\Omega^{s+|\beta|})$.

Proof. By Lemma 4.4.1, $\|x^\beta e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}\|_\infty \leq D_\Omega^{|\beta|} e^{D_\Omega^2}$. For any $1 \leq |\alpha| \leq s$, the Faà di Bruno formula implies

$$D^\alpha e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2} = e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2} \sum_{\lambda_1+\cdots+\lambda_s=\alpha} \gamma_{\alpha,\lambda_1,\ldots,\lambda_s} \prod_{j=1}^s D^{\lambda_j}(h^*(x,y) - \frac{1}{2}\|x-y\|^2)$$

for some combinatorial coefficients $\gamma_{\alpha,\lambda_1,\ldots,\lambda_s}$, where the derivative operators are taken with respect to the x variable. By Lemma 4.4.1, this quantity is bounded in magnitude by $c_{d,s} e^{D_\Omega^2}(1 + D_\Omega^{|\alpha|})$ for some constant $c'_{d,s}$. This implies

$$|D^\alpha x^\beta e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}| \leq c''_{d,s} e^{D_\Omega^2}(1 + D_\Omega^{s+|\beta|}) \qquad \text{for all } 1 \leq |\alpha| \leq s.$$

Therefore, choosing $c_{d,s}$ to be a sufficiently large constant depending on d and s yields the claim. $\square$

For our particular choice of optimal potentials we can also control the uniform norm, as follows.

Lemma 4.4.4. *Under (4.25), we have*

$$\|f^*\|_\infty, \|f_n\|_\infty, \|g^*\|_\infty, \|g_n\|_\infty \leq \tfrac{1}{2}D_\Omega^2.$$

Proof. Since $S_\epsilon(P,Q) = \int_{\mathbb{R}^d} f^*(x)dP(x) + \int_{\mathbb{R}^d} g^*(y)dQ(y) \geq 0$, (4.25) implies $\int_{\mathbb{R}^d} f^*(x)dP(x) \geq 0$. Therefore, using first the optimality conditions, then Jensen's inequality and finally (4.25), we obtain

$$g^*(y) = -\log\left(\int e^{f^*(y)-\frac{1}{2}\|x-y\|^2} dP(y)\right) \leq \int \{\tfrac{1}{2}\|x-y\|^2 - f^*(y)\} dP(y) \leq \tfrac{1}{2}D_\Omega^2,$$

for all $y \in \Omega$. By the same argument $f^*(x) \leq \frac{1}{2}D_\Omega^2$, for all $x \in \Omega$. Set $x \in \Omega$, by Lemma 4.4.1,

$$f^*(x) \geq \min_{y\in\Omega}\{\tfrac{1}{2}\|x-y\|^2 - g^*(y)\} \geq -\max_{y\in\Omega} g^*(y) \geq -\tfrac{1}{2}D_\Omega^2,$$

and the same for g^*. $\square$

For any $a, b \in \mathcal{C}^s(\Omega)$ denote

$$L(a,b) = \int_{\mathbb{R}^d} a(x)dP(x) + \int_{\mathbb{R}^d} b(y)dQ(y) - \int_{\mathbb{R}^d\times\mathbb{R}^d} e^{a(x)+b(y)-\frac{1}{2}\|x-y\|^2} dP(x)dQ(y) + 1$$

and its semi-empirical counterpart

$$L_n(a,b) = \int_{\mathbb{R}^d} a(x)dP_n(x) + \int_{\mathbb{R}^d} b(y)dQ(y) - \int_{\mathbb{R}^d\times\mathbb{R}^d} e^{a(x)+b(y)-\frac{1}{2}\|x-y\|^2} dP_n(x)dQ(y) + 1.$$

Let us denote by h^* and h_n the functions, belonging to $\mathcal{C}^s(\Omega\times\Omega)$, defined by

$$h^*(x,y) = f^*(x) + g^*(y),\ h_n(x,y) = f_n(x) + g_n(y), \tag{4.26}$$

and $\pi^* \in \mathcal{P}(\Omega\times\Omega)$ the optimal coupling defined by $d\pi^* = e^{h^*(x,y)-\|x-y\|^2}dP(x)dQ(y)$. Abusing notation, we write $L(f_n, g_n) = L(h_n)$ and $L(f^*, g^*) = L(h^*)$. As a consequence of Lemma 4.4.4 we obtain the following useful bound,

$$\|h^*\|_\infty,\ \|h_n\|_\infty \leq D_\Omega^2, \text{ for all } n \in \mathbb{N}. \tag{4.27}$$

At this point, we can state the main theorem of this section.

Theorem 4.4.5. *Let $\Omega \subset \mathbb{R}^d$ be a compact set and $P, Q \in \mathcal{P}(\Omega)$. Assume (f^*, g^*) are optimal potentials for P, Q and (f_n, g_n) for P_n, Q satisfying (4.25). Then there exists a constant C_d, depending only on d, such that*

$$\mathbb{E}\|g_n - g^*\|^2_{\mathcal{C}^s(\Omega)},\ \mathbb{E}\|f_n - f^*\|^2_{\mathcal{C}^s(\Omega)} \leq \frac{C_d}{n} D_\Omega^{5(d+1)} e^{15D_\Omega^2}$$

Moreover, if $(f_{n,m}, g_{n,m})$ are optimal potentials for P_n, Q_m satisfying (4.25), then

$$\mathbb{E}\|g_{n,m} - g^*\|^2_{\mathcal{C}^s(\Omega)},\ \mathbb{E}\|f_{n,m} - f^*\|^2_{\mathcal{C}^s(\Omega)} \leq \frac{C_d}{\min\{n,m\}} D_\Omega^{5(d+1)} e^{15D_\Omega^2}.$$

The proof is divided in a sequence of technical lemmas, of some independent interest. We show first (Lemma 4.4.6) that the functional L is well-behaved in the sense of being strongly concave near its maximum. Then (in Lemma 4.4.7) we show that the functional $L_n - L$ is Lipsichtz. Typically, these two results are enough to prove convergence at the fast n^{-1} rate (see, e.g., Theorem 3.2.5 of Vaart and Wellner (1996)). Unfortunately, in our case, the norms appearing in Lemmas 4.4.6 and 4.4.7 are different. This technical issue can be handled thanks to Lemmas 4.4.8 and 4.4.9.

Lemma 4.4.6. *Let $\Omega \subset \mathbb{R}^d$ be a compact set and $P, Q \in \mathcal{P}(\Omega)$, then*

$$L(h_n) - L(h^*) \leq -\tfrac{1}{2}\|h_n - h^*\|^2_{L^2(d\pi^*)} e^{-\|h_n - h^*\|_\infty}, \tag{4.28}$$

where h^*, h_n and π^* are defined in (4.26).

Proof. The inequality $e^x \geq 1 + x + \frac{e^{-|x|}}{2}x^2$, which can be checked by elementary means, implies that

$$\begin{aligned}
&\int e^{h_n(x,y) - \frac{1}{2}||x-y||^2} dP(x)dQ(y) \\
&\quad = \int e^{h_n(x,y) - h^*(x,y)} e^{h^*(x,y) - \frac{1}{2}||x-y||^2} dP(x)dQ(y) \\
&\quad \geq \int \left\{1 + (h_n(x,y) - h^*(x,y)) + \tfrac{1}{2}(h_n(x,y) - h^*(x,y))^2 e^{-|h_n(x,y) - h^*(x,y)|}\right\} d\pi^*(x,y) \\
&\quad \geq \int \left\{1 + (h_n(x,y) - h^*(x,y)) + \tfrac{1}{2}(h_n(x,y) - h^*(x,y))^2 e^{-||h_n - h^*||_\infty}\right\} d\pi^*(x,y).
\end{aligned}$$

The optimality conditions yield $L(h^*) = \int h^*(x,y)dP(x)dQ(y)$. Hence,

$$\begin{aligned}
&\int e^{h_n(x,y) - \frac{1}{2}||x-y||^2} dP(x)dQ(y) \\
&\quad \geq -L(h^*) + \int \left\{1 + h_n(x,y) + \tfrac{1}{2}(h_n(x,y) - h^*(x,y))^2 e^{-||h_n - h^*||_\infty}\right\} d\pi^*(x,y).
\end{aligned}$$

We conclude by using the relation $\int h_n(x,y)d\pi^* = \int h_n(x,y)dP(x)dQ(y)$, which follows from the optimality conditions. □

Lemma 4.4.7. *Under the assumptions of Lemma 4.4.6, we have*

$$L_n(h_n) - L(h_n) - L_n(h^*) + L(h^*) \leq \|P - P_n\|_{\mathcal{C}_1^s(\Omega)} \|h_n - h^*\|_{\mathcal{C}^s(\Omega^2)}, \quad a.s. \tag{4.29}$$

where

$$\|P - P_n\|_{\mathcal{C}_1^s(\Omega)} := \sup_{\|f\|_{\mathcal{C}^s(\Omega)} \leq 1} \int f(x)(dP_n(x) - dP(x)). \tag{4.30}$$

Proof. As noted above, the optimality conditions imply that

$$L_n(h_n) = \int h_n(x,y) dP_n(x) dQ(y), \quad L(h_n) = \int h_n(x,y) dP(x) dQ(y),$$
$$L_n(h^*) = \int h^*(x,y) dP_n(x) dQ(y), \quad L(h^*) = \int h^*(x,y) dP(x) dQ(y).$$

Therefore we have

$$\begin{aligned}
&L_n(h_n) - L(h_n) - L_n(h^*) + L(h^*) \\
&= \int h_n(x,y) dQ(y)(dP_n(x) - dP(x)) - \int h^*(x,y) dQ(y)(dP_n(x) - dP(x)) \\
&= \int (h_n(x,y) - h^*(x,y)) dQ(y)(dP_n(x) - dP(x)) \\
&\leq \|h_n - h^*\|_{\mathcal{C}^s(\Omega)} \sup_{\substack{\|h\|_{\mathcal{C}^s(\Omega^2)} \leq 1 \\ h(x,y) = f(x) + g(y)}} \int h(x,y) dQ(y)(dP_n(x) - dP(x)).
\end{aligned}$$

Note that

$$\begin{aligned}
&\sup_{\substack{\|h\|_{\mathcal{C}^s(\Omega^2)} \leq 1 \\ h(x,y) = f(x) + g(y)}} \int h(x,y) dQ(y)(dP_n(x) - dP(x)) \\
&\quad = \sup_{\substack{\|h\|_{\mathcal{C}^s(\Omega^2)} \leq 1 \\ h(x,y) = f(x) + g(y)}} \int g(y) dQ(y)(dP_n(x) - dP(x)) + \int f(x) dQ(y)(dP_n(x) - dP(x)).
\end{aligned}$$

Since the first term is 0 and the second is not affected by adding constant to f, we see that it equals

$$\sup_{\|f\|_{\mathcal{C}^s(\Omega)} \leq 1} \int f(x)(dP_n(x) - dP(x)).$$

□

As anticipated, Lemma 4.4.7 works with the norm $\|\cdot\|^2_{\mathcal{C}^s(\Omega^2)}$ and Lemma 4.4.6 with $\|\cdot\|_{L^2(d\pi^*)}$. Both norms are different, but the next technical results show how these norms are related in the present setup.

Lemma 4.4.8. *Under the assumptions of Lemma 4.4.6,*

$$\|D^\alpha f^* - D^\alpha f_n\|_\infty^2 \leq c_{d,s} D_\Omega^{2|\alpha|} \|h_n - h^*\|_\infty^2,$$
$$\|D^\alpha g^* - D^\alpha g_n\|_\infty^2 \leq c_{d,s} D_\Omega^{2|\alpha|} e^{2D_\Omega^2} \left(\|h_n - h^*\|_\infty^2 + D_\Omega^{2s} \|P - P_n\|_{\mathcal{C}_1^s(\Omega)}^2 \right),$$

for every multi-index α, with $1 \leq |\alpha| \leq s$.

Proof. We let $c_{d,s}$ denote a positive constant depending on d and s whose value may change from line to line. We note first that $f^*(x) - f_n(x) = \bar{f}^*(x) - \bar{f}_n(x)$, where

$$\bar{f}^*(x) = f^*(x) - \tfrac{1}{2}\|x\|^2 \text{ and } \bar{f}_n(x) = f_n(x) - \tfrac{1}{2}\|x\|^2. \tag{4.31}$$

As in the proof of Lemma 4.4.1, the Faá di Bruno formula implies

$$D^\alpha f_n(x) - D^\alpha f^*(x) = D^\alpha \overline{f}_n(x) - D^\alpha \overline{f}^*(x)$$
$$= \sum_{\beta_1 + \cdots + \beta_s = \alpha} \lambda_{\alpha,\beta_1,\ldots,\beta_s} \left(\prod_{j=1}^s \int y^{\beta_j} e^{h_n(x,y) - \frac{1}{2}\|x-y\|^2} dQ(y) - \prod_{j=1}^s \int y^{\beta_j} e^{h^*(x,y) - \frac{1}{2}\|x-y\|^2} dQ(y) \right).$$

Splitting the product, this last term equals

$$\sum_{\beta_1 + \cdots + \beta_s = \alpha} \lambda_{\alpha,\beta_1,\ldots,\beta_s} \sum_{i=1}^s \prod_{j<i} \int y^{\beta_j} e^{h_n(x,y) - \frac{1}{2}\|x-y\|^2} dQ(y)$$
$$\prod_{j>i} \int y^{\beta_j} e^{h^*(x,y) - \frac{1}{2}\|x-y\|^2} dQ(y) \int y^{\beta_i} e^{-\frac{1}{2}\|x-y\|^2} \{e^{h_n(x,y)} - e^{h^*(x,y)}\} dQ(y).$$

Since $0 \in \Omega$, it follows that $|y^{\beta_j}| \leq D_\Omega^{|\beta_j|}$. Using that $|e^x - e^y| \leq (e^y + e^x)|x - y|$ we upper bound $|D^\alpha f_n(x) - D^\alpha f^*(x)|$ by

$$D_\Omega^{|\alpha|} \sum_{\beta_1 + \cdots + \beta_s = \alpha} |\lambda_{\alpha,\beta_1,\ldots,\beta_s}| \sum_{i=1}^s \int (e^{h^*(x,y) - \frac{1}{2}\|x-y\|^2} + e^{h_n(x,y) - \frac{1}{2}\|x-y\|^2}) |h_n(x,y) - h^*(x,y)| dQ(y)$$
$$\leq 2s D_\Omega^{|\alpha|} \sum_{\beta_1 + \cdots + \beta_s = \alpha} |\lambda_{\alpha,\beta_1,\ldots,\beta_s}| \|h_n - h^*\|_\infty,$$

where we have used (4.12) to bound the integral. We conclude that $\|D^a f_n(x) - D^a f^*(x)\|_\infty^2 \leq c_{d,s} D_\Omega^{2|\alpha|} \|h_n - h^*\|_\infty^2$.

Turning to g_n and g_n^*, we can argue similarly to obtain

$$
\begin{aligned}
&|D^\alpha g_n(y) - D^\alpha g^*(y)| = |D^\alpha \overline{g}_n(y) - D^\alpha \overline{g}^*(y)| \\
\leq & \sum_{\beta_1+\cdots+\beta_s=\alpha} |\lambda_{\alpha,\beta_1,\ldots,\beta_s}| \left(\prod_{j=1}^s \int x^{\beta_j} e^{h_n(x,y)-\frac{1}{2}\|x-y\|^2} dP_n(x) - \prod_{j=1}^s \int x^{\beta_j} e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2} dP(x) \right) \\
= & \sum_{\beta_1+\cdots+\beta_s=\alpha} |\lambda_{\alpha,\beta_1,\ldots,\beta_s}| \sum_{i=1}^s \prod_{j<i} \int x^{\beta_j} e^{h_n(x,y)-\frac{1}{2}\|x-y\|^2} dP_n(x) \prod_{j>i} \int x^{\beta_j} e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2} dP(x) \cdot \\
& \left(\int x^{\beta_j} e^{h_n(x,y)-\frac{1}{2}\|x-y\|^2} dP_n(x) - \int x^{\beta_j} e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2} dP(x) \right).
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\left| \int x^{\beta_j} e^{h_n(x,y)-\frac{1}{2}\|x-y\|^2} dP_n(x) - \int x^{\beta_j} e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2} dP(x) \right| \\
\leq & \left| \int x^{\beta_j} (e^{h_n(x,y)-\frac{1}{2}\|x-y\|^2} - e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}) dP_n(x) \right| + \left| \int x^{\beta_j} e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2} (dP(x) - P_n(x)) \right|
\end{aligned}
$$

Since $\|h_n\|_\infty, \|h^*\|_\infty \leq D_\Omega^2$ by (4.27), the first term can be bounded by $2D_\Omega^{|\beta_j|} e^{D_\Omega^2} \|h_n - h^*\|_\infty$. For the other term observe that by Lemma 4.4.3, the function $x \mapsto x^{\beta_j} e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2}$ belongs to $\mathcal{C}^s(\Omega)$, with norm at most $c_{d,s} e^{D_\Omega^2}(1 + D_\Omega^{s+|\beta_j|})$. We conclude that there exists some constant $c_{d,s}$ such that

$$
\left| \int x^{\beta_j} e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2} (dP_n(x) - dP(x)) \right| \leq c_{d,s} D_\Omega^{|\beta_j|+s} e^{D_\Omega^2} \|P - P_n\|_{\mathcal{C}_1^s(\Omega)}. \tag{4.32}
$$

Combining the last two estimates we finally have

$$
\|D^a g^* - D^a g_n\|_\infty^2 \leq c_{d,s} D_\Omega^{2|\alpha|} e^{2D_\Omega^2} \left(\|h_n - h^*\|_\infty^2 + D_\Omega^{2s} \|P - P_n\|_{\mathcal{C}_1^s(\Omega)}^2 \right),
$$

which allows us to conclude. □

Now we need to compare the norms $\|\cdot\|_\infty$ and $\|\cdot\|_{L^2(d\pi^*)}$. We set $C = e^{-\frac{3}{2}D_\Omega^2}$, and note that (4.25) implies

$$
\begin{aligned}
\int (h_n(x,y) - & h^*(x,y))^2 e^{h^*(x,y)-\frac{1}{2}\|x-y\|^2} dP(x) dQ(y) \\
\geq \; & C \int (h_n(x,y) - h^*(x,y))^2 dP(x) dQ(y) \\
= \; & C\{ \int (f_n(x) - f^*(x))^2 dP(x) + \int (g_n(y) - g^*(y))^2 dQ(y) \\
& +2 \int (f_n(x) - f^*(x))(g_n(y) - g^*(y)) dP(x) dQ(y) \}.
\end{aligned}
$$

Since the last term equals 0, we obtain the bound

$$\|h_n - h^*\|^2_{L^2(d\pi^*)} \geq e^{-\frac{3}{2}D_\Omega^2}\left(\|f_n - f^*\|^2_{L^2(dP)} + \|g_n - g^*\|^2_{L^2(dQ)}\right). \tag{4.33}$$

Finally, we prove the last technical result, which relates the L^2 and L^∞ norms for the difference of the potentials.

Lemma 4.4.9. *Under the assumptions of Lemma 4.4.6, we have*

$$\begin{aligned}\|f_n - f^*\|^2_{L^2(dP)} &+ \|g_n - g^*\|^2_{L^2(dQ)} \\ &\geq \tfrac{1}{2}e^{-2D_\Omega^2}\left(\|f_n - f^*\|^2_\infty + \|g_n - g^*\|^2_\infty\right) - c_{d,s}D_\Omega^{2s}\|P - P_n\|^2_{\mathcal{C}_1^s(\Omega)}.\end{aligned}$$

Proof. We will work separately with f^* and g^*. Fixing $x \in \Omega$, Jensen's inequality yields

$$|e^{-f^*(x)} - e^{-f_n(x)}|^2 \leq \int \left(|e^{g^*(y)-\frac{1}{2}\|x-y\|^2} - e^{g_n(y)-\frac{1}{2}\|x-y\|^2}|\right)^2 dQ(y).$$

Now, the mean value theorem implies

$$|x-y|e^{\min\{x,y\}} \leq |e^x - e^y| \leq e^{\max\{x,y\}}|x-y|, \quad x, y \in \mathbb{R},$$

yielding

$$\begin{aligned}e^{-2\max\{\|f^*\|_\infty,\|f_n\|_\infty\}}|f_n(x) - f^*(x)|^2 &\leq |e^{-f^*(x)} - e^{-f_n(x)}|^2 \\ &\leq e^{2\max\{\|g^*\|_\infty,\|g_n\|_\infty\}}\|g_n - g^*\|^2_{L^2(dQ)}.\end{aligned}$$

Consequently, using Lemma 4.4.4, we have proved that

$$\|g_n - g^*\|^2_{L^2(dQ)} \geq e^{-2D_\Omega^2}\|f_n - f^*\|^2_\infty,$$

Now we deal with $\|g_n - g^*\|^2_\infty$. We fix $y \in \Omega$. By the triangle inequality we have

$$\begin{aligned}&|e^{-g^*(y)} - e^{-g_n(y)}| \\ &\leq \int \left|e^{f^*(x)+\frac{1}{2}\|x-y\|^2} - e^{f_n(y)+\frac{1}{2}\|x-y\|^2}\right| dP(y) + \left|\int e^{f_n(x)+\frac{1}{2}\|x-y\|^2}(dP(x) - dP_n)\right|.\end{aligned}$$

Squaring both sides we see that

$$\begin{aligned}&|e^{-g^*(y)} - e^{-g_n(y)}|^2 \\ &\leq 2\int \left|e^{f^*(x)+\frac{1}{2}\|x-y\|^2} - e^{f_n(y)+\frac{1}{2}\|x-y\|^2}\right|^2 dP(y) + 2\left|\int e^{f_n(x)+\frac{1}{2}\|x-y\|^2}(dP(x) - dP_n)\right|^2.\end{aligned}$$

The first term is bounded by $2e^{D_\Omega^2}\|f_n - f^*\|^2_{L^2(dP)}$ as in the previous case. Repeating the arguments which led to the bound (4.32), the second term is at most $c_{d,s}e^{D_\Omega^2}(1 + D_\Omega^{2s})\|P - P_n\|^2_{\mathcal{C}_1^s(\Omega)}$. Together, these estimates yield

$$e^{-D_\Omega^2}\|g_n - g^*\|^2_\infty \leq 2e^{D_\Omega^2}\|f_n - f^*\|^2_{L^2(dP)} + c_{d,s}D_\Omega^{2s}e^{D_\Omega^2}\|P - P_n\|^2_{\mathcal{C}_1^s(\Omega)}.$$

We conclude by rearranging this inequality and combining it with the bound on $\|g_n - g^*\|^2_{L^2(dQ)}$ derived above. □

We are ready now for the proof of the main result in this section.

Proof of Theorem 4.4.5. Combining Lemma 4.4.6, (4.33) and Lemma 4.4.9 we see that

$$\begin{aligned} L(h^*) - L(h_n) &\geq \tfrac{1}{2} e^{-\|h_n - h^*\|_\infty} e^{-\frac{3}{2}D_\Omega^2} \left(\|f_n - f^*\|^2_{L^2(dP)} + \|g_n - g^*\|^2_{L^2(dQ)} \right) \\ &\geq \tfrac{1}{2} e^{-\frac{7}{2}D_\Omega^2} \left(\|f_n - f^*\|^2_{L^2(dP)} + \|g_n - g^*\|^2_{L^2(dQ)} \right) \\ &\geq \tfrac{1}{2} e^{-\frac{7}{2}D_\Omega^2} \left(\frac{1}{2} e^{-2D_\Omega^2} \left(\|f_n - f^*\|^2_\infty + \|g_n - g^*\|^2_\infty \right) - c_{d,s} D_\Omega^{2s} \|P - P_n\|^2_{\mathcal{C}^s_1(\Omega)} \right). \end{aligned}$$

Moreover, since $\|f_n - f^*\|^2_\infty + \|g_n - g^*\|^2_\infty \geq \frac{1}{2}\|h_n - h^*\|^2_\infty$, we obtain

$$L(h^*) - L(h_n) \geq \tfrac{1}{2} e^{-\frac{7}{2}D_\Omega^2} \left(\frac{1}{4} e^{-2D_\Omega^2} \|h_n - h^*\|^2_\infty - D_\Omega^{2s} \|P - P_n\|^2_{\mathcal{C}^s_1(\Omega)} \right).$$

Lemma 4.4.8 implies the existence of some constant $c_{d,s}$ such that

$$\|h_n - h^*\|^2_\infty \geq \frac{1}{c_{d,s} D_\Omega^{2s} e^{2D_\Omega^2}} \left(\|f_n - f^*\|^2_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|^2_{\mathcal{C}^s(\Omega)} \right) - D_\Omega^{2s} \|P - P_n\|^2_{\mathcal{C}^s_1(\Omega)},$$

which yields

$$\begin{aligned} L(h^*) - L(h_n) \geq c_{d,s} e^{-\frac{15}{2}D_\Omega^2} D_\Omega^{-2s} &\left(\|f_n - f^*\|^2_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|^2_{\mathcal{C}^s(\Omega)} \right) \\ &- c'_{d,s} e^{-\frac{7}{2}D_\Omega^2} D_\Omega^{2s} \|P - P_n\|^2_{\mathcal{C}^s_1(\Omega)}. \end{aligned} \tag{4.34}$$

On the other hand, Lemma 4.4.7 yields

$$\begin{aligned} \|P - P_n\|_{\mathcal{C}^s_1(\Omega)} \left(\|f_n - f^*\|_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|_{\mathcal{C}^s(\Omega)} \right) &\geq L_n(h_n) - L(h_n) - L_n(h^*) + L(h^*) \\ &\geq L_n(h^*) - L(h_n) - L_n(h^*) + L(h^*) \\ &= L(h^*) - L(h_n) \end{aligned}$$

The previous bound and (4.34) yield

$$\begin{aligned} &\|P - P_n\|_{\mathcal{C}^s_1(\Omega)} \left(\|f_n - f^*\|_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|_{\mathcal{C}^s(\Omega)} \right) \geq \\ &c_{d,s} e^{-\frac{15}{2}D_\Omega^2} D_\Omega^{-2s} \left(\|f_n - f^*\|^2_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|^2_{\mathcal{C}^s(\Omega)} \right) - c'_{d,s} e^{-\frac{7}{2}D_\Omega^2} D_\Omega^{2s} \|P - P_n\|^2_{\mathcal{C}^s_1(\Omega)}. \end{aligned}$$

which, by using the inequality $(a+b)^2 \leq 2(a^2 + b^2)$, implies

$$\begin{aligned} &\sqrt{2}\|P - P_n\|_{\mathcal{C}^s_1(\Omega)} \left(\|f_n - f^*\|^2_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|^2_{\mathcal{C}^s(\Omega)} \right)^{1/2} \geq \\ &c_{d,s} e^{-\frac{15}{2}D_\Omega^2} D_\Omega^{-2s} \left(\|f_n - f^*\|^2_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|^2_{\mathcal{C}^s(\Omega)} \right) - c'_{d,s} e^{-\frac{7}{2}D_\Omega^2} D_\Omega^{2s} \|P - P_n\|^2_{\mathcal{C}^s_1(\Omega)}. \end{aligned}$$

Denoting $\Delta_n = \left(\|f_n - f^*\|^2_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|^2_{\mathcal{C}^s(\Omega)}\right)^{\frac{1}{2}}$, we get

$$\|P - P_n\|_{\mathcal{C}^s_1(\Omega)}\Delta_n \geq c_{d,s}e^{-\frac{15}{2}D^2_\Omega}D_\Omega^{-2s}\Delta_n^2 - c'_{d,s}e^{-\frac{7}{2}D^2_\Omega}D^{2s}_\Omega\|P - P_n\|^2_{\mathcal{C}^s_1(\Omega)}. \tag{4.35}$$

From this we obtain

$$\begin{aligned}\Delta_n &\leq c_{d,s}D^{2s}_\Omega e^{\frac{15}{2}D^2_\Omega}\left(\|P - P_n\|_{\mathcal{C}^s_1(\Omega)} + \sqrt{\|P - P_n\|^2_{\mathcal{C}^s_1(\Omega)} + e^{-11D^2_\Omega}\|P - P_n\|^2_{\mathcal{C}^s_1(\Omega)}}\right)\\ &\leq c_{d,s}D^{2s}_\Omega e^{\frac{15}{2}D^2_\Omega}\|P - P_n\|_{\mathcal{C}^s_1(\Omega)}.\end{aligned} \tag{4.36}$$

Next, we analyze $\|P - P_n\|_{\mathcal{C}^s_1(\Omega)}$. Theorem 3.5.1 and Exercise 2.3.1 in Giné and Nickl (2015) imply that there exists a numerical constant C such that

$$n\mathbb{E}\|P - P_n\|^2_{\mathcal{C}^s_1(\Omega)} \leq C\mathbb{E}\left(\int_0^{\max_{\|f\|_{\mathcal{C}^s(\Omega)}\leq 1}\|f\|_{L_2(P_n)}}\sqrt{\log\left(2N(\epsilon, \mathcal{C}^s_1(\Omega), \|\cdot\|_\infty)\right)}d\epsilon\right)^2.$$

By Proposition 1.1. in van der Vaart (1994),

$$\log\left(2N(\epsilon, \mathcal{C}^s_1(\Omega), \|\cdot\|_\infty)\right) \leq c_{s,d}D^d_\Omega\left(\frac{1}{\epsilon}\right)^{\frac{d}{s}}.$$

Since $\|f\|_{L_2(P_n)} \leq \|f\|_{\mathcal{C}^s(\Omega)}$ almost surely, the choice $s = [d/2] + 1$ yields the bound

$$n\mathbb{E}\|P - P_n\|^2_{\mathcal{C}^s_1(\Omega)} \leq c_{s,d}D^d_\Omega\left(\int_0^1\left(\frac{1}{\epsilon}\right)^{\frac{d}{d+1}}d\epsilon\right)^2 = c_{s,d}D^d_\Omega, \tag{4.37}$$

which completes the proof for the one-sample case.

The two-sample case can be handled with the same argument plus some minor modifications, as follows. Let $f_{n,m}$ be the optimal potential for P_n and Q_m. Then,

$$\|f_{n,m} - f^*\|^2_{\mathcal{C}^s(\Omega)} \leq 2\|f_n - f^*\|^2_{\mathcal{C}^s(\Omega)} + 2\|f_{n,m} - f_n\|^2_{\mathcal{C}^s(\Omega)}.$$

The first term can be controlled by (4.36). Moreover, observe that the derivation of (4.36) did not use any facts about the measure Q apart from the fact that it is supporte don Ω. Since P_n is also supporte din Ω, this implies that $\|f_{n,m} - f_n\|_{\mathcal{C}^s(\Omega)}$ can also be controlled by (4.36), so that the bound

$$\|f_{n,m} - f^*\|^2_{\mathcal{C}^s(\Omega)} \leq c_{d,s}D^{2s}_\Omega e^{\frac{15}{2}D^2_\Omega}\left(\|P - P_n\|^2_{\mathcal{C}^s_1(\Omega)} + \|Q - Q_m\|^2_{\mathcal{C}^s_1(\Omega)}\right) \tag{4.38}$$

holds. This and (4.37) complete the proof. □

4.5 Convergence rates for Sinkhorn divergences

In this section, we develop faster convergence rates for the *Sinkhorn divergence*. The entropic transportation cost, $S_\epsilon(P,Q)$ is not symmetric in P,Q and does not satisfy $S_\epsilon(P,P)=0$. These observations motivated the introduction of Sinkhorn divergences Genevay et al. (2018): For probabilities $P,Q\in\mathcal{P}(\mathbb{R}^d)$ the quadratic Sinkhorn divergence is defined as

$$D_\epsilon(P,Q)=S_\epsilon(P,Q)-\tfrac{1}{2}\left(S_\epsilon(P,P)+S_\epsilon(Q,Q)\right).$$

Clearly, $D_\epsilon(P,Q)$ is symmetric in P,Q and $D_\epsilon(P,P)=0$. In fact (see Theorem 1 in Feydy et al. (2019)), $D_\epsilon(P,Q)\geq 0$, with $D_\epsilon(P,Q)=0$ if and only if $P=Q$, and for measures supported on a compact set, convergence in Sinkhorn distance is equivalent to weak convergence. This makes the Sinkhorn divergence a suitable measure of dissimilarity in applications.

In this section we obtain rates of convergence for empirical Sinkhorn divergences. More precisely, we consider independent samples $X_1,\ldots,X_n$, $Y_1,\ldots,Y_m$ of i.i.d. r.v.'s with law $P\in\mathcal{P}(\Omega)$ and associated empirical measures P_n and P'_m, respectively. Since P_n and P'_m converge weakly to P, the Sinkhorn divergence satisifes $D_\epsilon(P_n,P'_m)\to 0$ a.s. The main result of this section gives a rate for this convergence.

Theorem 4.5.1. *Assume $\Omega\subset\mathbb{R}^d$ is compact, $P\in\mathcal{P}(\Omega)$ and P_n and P'_m are empirical measures as above. Then there exist constants c_d and c'_d, depending only on d, such that*

(i) (one-sample case)

$$\mathbb{E}D_1(P_n,P)\leq\frac{c_d}{n}D_\Omega^{\frac{3d}{2}+1}\frac{32}{(d+1)^2}e^{\frac{19}{2}D_\Omega^2}.$$

(ii) (two-sample case)

$$\mathbb{E}D_1(P_n,P'_m)\leq\frac{c'_d}{\min\{n,m\}}D_\Omega^{\frac{3d}{2}+1}\frac{32}{(d+1)^2}e^{\frac{19}{2}D_\Omega^2}.$$

Proof. We deal first with the one-sample case. We denote by $(f_{n,n},g_{n,n})$ the optimal potentials for $S_1(P_n,P_n)$, set $h_{n,n}(x,y)=f_{n,n}(x)+g_{n,n}(y)$ and write $d\pi_{n,n}(x,y)=e^{h_{n,n}(x,y)-\frac{1}{2}\|x-y\|^2}$ for the optimal measure and, as in (4.26), we write h^*, π^* for the corresponding objects in the case of $S_1(P,P)$ and h_n, π_n in the case of $S_1(P_n,P)$. Then we can write

$$D_1(P_n,P)=\int h_n(x,y)d\pi_n(x,y)-\tfrac{1}{2}\left(\int h_{n,n}(x,y)d\pi_{n,n}(x,y)+\int h^*(x,y)d\pi^*(x,y)\right).\tag{4.39}$$

Moreover, using the optimality conditions, we have

$$\int\left(h_n(x,y)-h^*(x,y)\right)d\pi^*(x,y)=L(h_n)-L(h^*)\leq 0,$$

and

$$\int (h_n(x,y) - h_{n,n}(x,y))\, d\pi_{n,n}(x,y) = L_{n,n}(h_n) - L_{n,n}(h_{n,n}) \leq 0,$$

where L is defined as in the previous section and

$$L_{n,n}(h) = \int \left\{ h(x,y) - e^{h(x,y) - \frac{1}{2}||x-y||^2} + 1 \right\} dP_n(x) dP_n(x).$$

Therefore, from (4.39) we obtain

$$D_1(P_n, P) \leq \int h_n(x,y) d\pi_n(x,y) - \tfrac{1}{2} \left(\int h_n(x,y) d\pi_{n,n}(x,y) + \int h_n(x,y) d\pi^*(x,y) \right). \tag{4.40}$$

Note, moreover, that the upper bound in (4.40) can be rewritten as

$$\begin{aligned}
&\int f_n(x) dP_n(x) + \int g_n(y) dP(y) \\
&\qquad - \tfrac{1}{2} \left(\int f_n(x) dP_n(x) + \int g_n(y) dP_n(y) + \int f_n(x) dP(x) + \int g_n(y) dP(y) \right) \\
&\qquad = \tfrac{1}{2} \int (f_n(x) - g_n(x))\, dP_n(x) + \tfrac{1}{2} \int (g_n(x) - f_n(x))\, dP(x) \\
&\qquad = \tfrac{1}{2} \int (f_n(x) - g_n(x))\, (dP_n(x) - dP(x)) \\
&\qquad = \tfrac{1}{2} \int (f_n(x) - g^*(x))\, (dP_n(x) - dP(x)) \\
&\qquad + \tfrac{1}{2} \int (g^*(x) - g_n(x))\, (dP_n(x) - dP(x)),
\end{aligned} \tag{4.41}$$

where (f_n, g_n) are optimal entropic potentials for P_n, P and (f^*, g^*) are optimal transport potentials for (P, P), where adopt the normalization convention $\int g^*(y) dP(y) = \int g_n(y) dP(y) = 0$. The symmetry of $S_1(P, P)$ and the uniqueness of the entropic potentials up to additive constants implies that $f^* = g^* + a$ for some constant $a \in \mathbb{R}$. Plugging this into (4.41) we obtain from (4.40) that

$$D_1(P_n, P) \leq \tfrac{1}{2} \|P - P_n\|_{\mathcal{C}_1^s(\Omega)} \left(\|f_n - f^*\|_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|_{\mathcal{C}^s(\Omega)} \right). \tag{4.42}$$

From (4.36), we obtain, for some constant $c_{d,s}$, the bound

$$\begin{aligned}
\left(\|f_n - f^*\|_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|_{\mathcal{C}^s(\Omega)} \right) &\leq 2 \left(\|f_n - f^*\|^2_{\mathcal{C}^s(\Omega)} + \|g_n - g^*\|^2_{\mathcal{C}^s(\Omega)} \right)^{\frac{1}{2}} \\
&\leq c_{d,s} D_\Omega^{2s} e^{\frac{15}{2} D_\Omega^2} \|P - P_n\|_{\mathcal{C}_1^s(\Omega)}.
\end{aligned}$$

We conclude as in the proof of Theorem 4.4.5.

For the two sample case we can adapt the argument above without much effort. Indeed, observe that we can write

$$D_1(P_n, P'_m) = \int h_{n,m}(x,y)d\pi_{n,m}(x,y) - \tfrac{1}{2}\Big(\int h_{n,n}(x,y)d\pi_{n,n}(x,y) + \int h_{m,m}(x,y)d\pi_{m,m}(x,y)\Big) \tag{4.43}$$

and argue as in (4.40) to get

$$D_1(P_n, P'_m) \leq \int h_{n,m}(x,y)d\pi_{n,m}(x,y) - \tfrac{1}{2}\left(\int h_{n,m}(x,y)d\pi_{n,n}(x,y) + \int h_{n,m}(x,y)d\pi_{m,m}(x,y)\right). \tag{4.44}$$

Now we can reuse the same arguments leading to (4.40)—just replacing P by P'_m—to upper bound $D_1(P_n, P'_m)$ by

$$\tfrac{1}{2}\int (f_{n,m}(x) - g^*(x))\left(dP_n(x) - dP'_m(x)\right) + \tfrac{1}{2}\int (g^*(x) - g_{n,m}(x))\left(dP_n(x) - dP'_m(x)\right). \tag{4.45}$$

Once again, since (f^*, g^*) agree up to an additive constant, (4.45) is equivalent to

$$\tfrac{1}{2}\int (f_{n,m}(x) - f^*(x))\left(dP_n(x) - dP'_m(x)\right) + \tfrac{1}{2}\int (g^*(x) - g_{n,m}(x))\left(dP_n(x) - dP'_m(x)\right).$$

Finally, the two sample case can be deduced directly from the following inequality and

$$\begin{aligned} D_1(P_n, P) &\leq \frac{1}{2}\|P_n - P'_m\|_{\mathcal{C}_1^s(\Omega)}\left(\|f_{n,m} - f^*\|_{\mathcal{C}^s(\Omega)} + \|g_{n,m} - g^*\|_{\mathcal{C}^s(\Omega)}\right) \\ &\leq \frac{1}{2}\left(\|P'_m - P\|_{\mathcal{C}_1^s(\Omega)} + \|P_n - P\|_{\mathcal{C}_1^s(\Omega)}\right)\left(\|f_{n,m} - f^*\|_{\mathcal{C}^s(\Omega)} + \|g_{n,m} - g^*\|_{\mathcal{C}^s(\Omega)}\right) \end{aligned} \tag{4.46}$$

and (4.38), which yields

$$\|f_{n,m} - f^*\|_{\mathcal{C}^s(\Omega)} + \|g_{n,m} - g^*\|_{\mathcal{C}^s(\Omega)} \leq c_{d,s} D_\Omega^{2s} e^{\frac{15}{2}D_\Omega^2}\left(\|P - P_n\|_{\mathcal{C}_1^s(\Omega)} + \|P - P'_m\|_{\mathcal{C}_1^s(\Omega)}\right)$$

for a constant $c_{d,s}$ depending on d and s. We conclude as above. □

4.6 Implementation issues and empirical results

In this section we provide details about the practical implementation and statistical performance of the two-sample analog of the confidence interval (4.24).

Recall from Theorem 4.3.6 that

$$\sqrt{\frac{nm}{n+m}}\frac{1}{\sigma_{\epsilon,\lambda}(P,Q)}(S_\epsilon(P_n,Q_m)-S_\epsilon(P,Q)) \xrightarrow{w} N(0,1), \tag{4.47}$$

where $\sigma^2_{\epsilon,\lambda}(P,Q)$ is the asymptotic variance of the two-sample case. This variance can be consistently estimated by

$$\hat{\sigma}^2_{n,m} := \frac{m}{n+m}\left(\frac{1}{n}\sum_{i=1}^n f^2_{n,\epsilon}(X_i) - \left(\frac{1}{n}\sum_{i=1}^n f_{n,\epsilon}(X_i)\right)^2\right) + \frac{n}{n+m}\left(\frac{1}{m}\sum_{i=1}^m g^2_{n,\epsilon}(Y_i) - \left(\frac{1}{m}\sum_{i=1}^m g_{n,\epsilon}(Y_i)\right)^2\right), \tag{4.48}$$

where $(f_{n,\epsilon}, g_{n,\epsilon})$ is a pair of empirical potentials. Hence, writing z_β for the β quantile of the standard normal distribution and arguing as in Section 4.3, we can conclude that the interval

$$CI^{n,m}_\alpha = \left[S_\epsilon(P_n,Q_m) \pm \hat{\sigma}_{n,m}\sqrt{\frac{n+m}{nm}}z_{1-\alpha/2}\right]$$

is an asymptotic confidence interval of level $1-\alpha$.

We investigate here the finite sample performance of this confidence interval. We consider the scenario where $P \sim N(0, I_d/2)$ and $Q \sim N((1,\dots,1)^t, I_d/2)$. The population entropy regularized cost has a closed form for Gaussian measures (see del Barrio and Loubes (2020), Janati et al. (2020) or Mallasto et al. (2021)), which, for our case, is

$$S_\epsilon(P,Q) = 2\,d - \frac{\epsilon}{2}\left(d\sqrt{1+\frac{4}{\epsilon^2}} - d\log\left(1+\sqrt{1+\frac{4}{\epsilon^2}}\right) + d\,\log(2) - d\right).$$

We focus on the case $n=m$, for several choices of $n = 50, 100, 250, 500, 1000, 5000$, and study the influence of the parameters d and ϵ on the rate of convergence of the true confidence level to the nominal level $1-\alpha$, for $\alpha = 0.05$. To approximate this true confidence level we use Monte Carlo simulation, with 1000 replicates of the interval. The results are reported on Table 4.6.1. In particular, we compute $CI^{n,n}_{0.05}$ for different values of $\epsilon \in [0.5, 2, 5, 10]$ and $d \in [2, 10, 15]$. To calculate $S_\epsilon(P_n, Q_n)$ and the empirical potentials—which allows us to compute $CI^{n,n}_{0.05}$—we use the python library POT, see Flamary et al. (2021). We observe that, effectively, both d and ϵ affect the estimation of the asymptotic confidence interval $CI^{n,n}_{0.05}$. Actually a large sample size is required to achieve the nominal confidence interval for small values of ϵ and large dimension. This is more or less expected, in view of Remark 4.3.5, the value $n\left|\mathbb{E}S_\epsilon(P_n,Q) - S_\epsilon(P,Q)\right|$ can be upper bounded by a constant C_Ω, which depends exponentially on the support's diameter— extrapolating this argument to the case where the probabilities are not supported in a compact set — and it provides a possible explanation of the inaccuracy produced by the choice of small values ϵ or large d. Note that this exponential dependency on the diameter is translated directly to an exponential dependence on $\frac{1}{\epsilon}$ by the change of variables $x \mapsto \varepsilon^{-\frac{1}{2}}x$. Moreover, the convergence, when $\epsilon \to 0$, of the entropic regularised potentials towards the optimal transport ones (see Nutz and Wiesel (2022)), which are cursed by the dimension (see Weed and Bach (2019)), explains also the results of Table 4.6.1.

		$\mathbb{P}\left(S_\epsilon(P,Q) \in CI^{n,n}_{0.05}\right)$			
	n	$\epsilon = 0.5$	$\epsilon = 2$	$\epsilon = 5$	$\epsilon = 10$
$d = 2$	50	0.935	0.936	0.932	0.941
	100	0.937	0.952	0.929	0.941
	250	0.95	0.94	0.935	0.949
	500	0.954	0.947	0.95	0.958
	1000	0.944	0.954	0.947	0.96
	5000	0.939	0.957	0.947	0.955
$d = 10$	50	0.781	0.945	0.958	0.932
	100	0.787	0.937	0.951	0.945
	250	0.775	0.941	0.948	0.943
	500	0.785	0.955	0.953	0.947
	1000	0.803	0.94	0.945	0.954
	5000	0.862	0.944	0.946	0.951
$d = 15$	50	0.487	0.944	0.933	0.944
	100	0.396	0.944	0.957	0.944
	250	0.271	0.938	0.943	0.953
	500	0.194	0.94	0.941	0.947
	1000	0.173	0.938	0.945	0.955
	5000	0.134	0.942	0.943	0.943

Table 4.6.1: Evolution of the Monte Carlo estimation (number of iterations equals 1000) of $\mathbb{P}\left(S_\epsilon(P,Q) \in CI^{n,n}_{0.05}\right)$ for different values of the dimension d and regularization factor ϵ.

Acknowledgements

The research of Eustasio del Barrio is partially supported by FEDER, Spanish Ministerio de Economía y Competitividad, grant MTM2017-86061-C2-1-P and Junta de Castilla y León, grants VA005P17 and VA002G18. The research of Alberto González-Sanz and Jean-Michel Loubes is partially supported by the AI Interdisciplinary Institute ANITI, which is funded by the French "Investing for the Future – PIA3" program under the Grant agreement ANR-19-PI3A-0004. The research of Jonathan Niles-Weed is partially funded by National Science Foundation, Grant DMS-2015291.

Bibliography

Alberto Abadie and Guido W Imbens. Large sample properties of matching estimators for average treatment effects. *econometrica*, 74(1):235–267, 2006.

Jason M. Altschuler, Jonathan Niles-Weed, and Austin J. Stromme. Asymptotics for semidiscrete entropic optimal transport. *SIAM J. Math. Anal.*, 54(2):1718–1741, 2022. ISSN 0036-1410. doi: 10.1137/21M1440165. URL https://doi.org/10.1137/21M1440165.

David Alvarez-Melis and Tommi S. Jaakkola. Gromov-Wasserstein alignment of word embedding spaces. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 1881–1890, 2018. URL https://aclanthology.info/papers/D18-1214/d18-1214.

E. Barrio, Paula Gordaliza, and Jean-Michel Loubes. A central limit theorem for lp transportation cost on the real line with application to fairness assessment in machine learning. *Information and Inference: A Journal of the IMA*, 2019.

Robert J. Berman. The Sinkhorn algorithm, parabolic optimal transport and geometric Monge-Ampère equations. *Numer. Math.*, 145(4):771–836, 2020. ISSN 0029-599X. doi: 10.1007/s00211-020-01127-x. URL https://doi.org/10.1007/s00211-020-01127-x.

Jérémie Bigot, Elsa Cazelles, and Nicolas Papadakis. Central limit theorems for entropy-regularized optimal transport on finite spaces and statistical applications. *Electronic Journal of Statistics*, 13(2):5120 – 5150, 2019. doi: 10.1214/19-EJS1637. URL https://doi.org/10.1214/19-EJS1637.

Emily Black, Samuel Yeom, and Matt Fredrikson. Fliptest: fairness testing via optimal transport. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 111–121, 2020.

Jose Blanchet, Yang Kang, and Karthyek Murthy. Robust Wasserstein profile inference and applications to machine learning. *J. Appl. Probab.*, 56(3):830–857, 2019. ISSN 0021-9002. doi: 10.1017/jpr.2019.49. URL https://doi.org/10.1017/jpr.2019.49.

Victor Chernozhukov, Alfred Galichon, Marc Hallin, and Marc Henry. Monge-Kantorovich depth, quantiles, ranks and signs. *Ann. Statist.*, 45(1):223–256, 2017. ISSN 0090-5364. doi: 10.1214/16-AOS1450. URL https://doi.org/10.1214/16-AOS1450.

Silvia Chiappa and Aldo Pacchiano. Fairness with continuous optimal transport. *arXiv preprint arXiv:2101.02084*, 2021.

Silvia Chiappa, Ray Jiang, Tom Stepleton, Aldo Pacchiano, Heinrich Jiang, and John Aslanides. A general approach to fairness with optimal transport. In *AAAI*, pages 3633–3640, 2020.

Lenaic Chizat, Pierre Roussillon, Flavien Léger, François-Xavier Vialard, and Gabriel Peyré. Faster wasserstein distance estimation with the sinkhorn divergence. *Advances in Neural Information Processing Systems*, 33, 2020.

Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(9):1853–1865, 2017.

I. Csiszar. I-Divergence Geometry of Probability Distributions and Minimization Problems. *The Annals of Probability*, 3(1):146 – 158, 1975. doi: 10.1214/aop/1176996454. URL https://doi.org/10.1214/aop/1176996454.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pages 2292–2300, 2013. URL http://papers.nips.cc/paper/4927-sinkhorn-distances-lightspeed-computation-of-optimal-transport.

Lucas de Lara, Alberto González-Sanz, Nicholas Asher, and Jean-Michel Loubes. Transport-based counterfactual models. *Pre-Print*, 2021.

Nabarun Deb and Bodhisattva Sen. Multivariate rank-based distribution-free nonparametric testing using measure transportation. *J. Amer. Statist. Assoc.*, 2019. To appear.

Eustasio del Barrio and Jean-Michel Loubes. Central limit theorems for empirical transportation cost in general dimension. *The Annals of Probability*, 47(2):926 – 951, 2019. doi: 10.1214/18-AOP1275. URL https://doi.org/10.1214/18-AOP1275.

Eustasio del Barrio and Jean-Michel Loubes. The statistical effect of entropic regularization in optimal transportation. *Pre-Print*, abs/2006.05199, 2020.

Eustasio del Barrio, Evarist Giné, and Carlos Matrán. Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *Ann. Probab.*, 27 (2):1009–1071, 1999. ISSN 0091-1798. doi: 10.1214/aop/1022677394. URL https://doi.org/10.1214/aop/1022677394.

Eustasio del Barrio, Evarist Giné, and Frederic Utzet. Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli*, 11(1):131–189, 2005. ISSN 1350-7265. doi: 10.3150/bj/1110228245. URL https://doi.org/10.3150/bj/1110228245.

Eustasio del Barrio, Alberto González-Sanz, and Jean-Michel Loubes. Central limit theorems for general transportation costs. *Pre-Print*, 2021a.

Eustasio del Barrio, Alberto González-Sanz, and Jean-Michel Loubes. Central limit theorems for semi-discrete wasserstein distances. *Pre-Print*, 2021b.

Eustasio del Barrio, Alberto Gonzalez-Sanz, Jean-Michel Loubes, and Jonathan Niles-Weed. An improved central limit theorem and fast convergence rates for entropic transportation costs. *arXiv preprint arXiv:2204.09105*, 2022.

Stephan Eckstein and Marcel Nutz. Quantitative stability of regularized optimal transport and convergence of sinkhorn's algorithm. *Pre-Print*, 10 2021. URL https://arxiv.org/pdf/2110.06798.pdf.

Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trouvé, and Gabriel Peyré. Interpolating between optimal transport and mmd using sinkhorn divergences. In *AISTATS*, 2019.

Rémi Flamary, Marco Cuturi, Nicolas Courty, and Alain Rakotomamonjy. Wasserstein discriminant analysis. *Mach. Learn.*, 107(12):1923–1945, 2018. ISSN 0885-6125. doi: 10.1007/s10994-018-5717-1. URL https://doi.org/10.1007/s10994-018-5717-1.

Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T.H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021. URL http://jmlr.org/papers/v22/20-451.html.

Rui Gao and Anton J Kleywegt. Distributionally robust stochastic optimization with wasserstein distance. *arXiv preprint arXiv:1604.02199*, 2016.

Aude Genevay, Gabriel Peyré, and Marco Cuturi. Learning generative models with sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain*, pages 1608–1617, 2018. URL http://proceedings.mlr.press/v84/genevay18a.html.

Aude Genevay, Lénaïc Chizat, Francis Bach, Marco Cuturi, and Gabriel Peyré. Sample complexity of sinkhorn divergences. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pages 1574–1583. PMLR, 16–18 Apr 2019. URL http://proceedings.mlr.press/v89/genevay19a.html.

Promit Ghosal and Bodhisattva Sen. Multivariate ranks and quantiles using optimal transport: Consistency, rates, and nonparametric testing. *arXiv preprint arXiv:1905.05340*, 2019.

Promit Ghosal, Marcel Nutz, and Espen Bernton. Stability of entropic optimal transport and schrödinger bridges. *Pre-Print*, 06 2021. URL https://arxiv.org/pdf/2106.03670.pdf.

Eva Giné and Richard Nickl. Mathematical foundations of infinite-dimensional statistical models. In *Cambridge Series in Statistical and Probabilistic Mathematics, [40]. Cambridge University Press, New York.*, 2015.

Javier González-Delgado, Alberto González-Sanz, Juan Cortés, and Pierre Neuvial. Two-sample goodness-of-fit tests on the flat torus based on wasserstein distance and their relevance to structural biology. *arXiv: Probability*, 2021.

Paula Gordaliza, Eustasio Del Barrio, Gamboa Fabrice, and Jean-Michel Loubes. Obtaining fairness using optimal transport theory. In *International Conference on Machine Learning*, pages 2357–2365. PMLR, 2019.

Edouard Grave, Armand Joulin, and Quentin Berthet. Unsupervised alignment of embeddings with wasserstein procrustes. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1880–1890. PMLR, 2019.

Marc Hallin, Eustasio del Barrio, Juan Cuesta-Albertos, and Carlos Matrán. Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *The Annals of Statistics*, 49(2):1139 – 1165, 2021a. doi: 10.1214/20-AOS1996.

Marc Hallin, Gilles Mordant, and Johan Segers. Multivariate goodness-of-fit tests based on Wasserstein distance. *Electron. J. Stat.*, 15(1):1328–1371, 2021b. doi: 10.1214/21-ejs1816. URL https://doi.org/10.1214/21-ejs1816.

Zaïd Harchaoui, Lang Liu, and Soumik Pal. Asymptotics of entropy-regularized optimal transport via chaos decomposition. *arXiv: Probability*, 2020.

Shayan Hundrieser, Marcel Klatt, Thomas Staudt, and Axel Munk. A unifying approach to distributional limits for empirical optimal transport. *Pre-Print*, 02 2022. URL https://arxiv.org/pdf/2202.12790.pdf.

Jan-Christian Hütter and Philippe Rigollet. Minimax estimation of smooth optimal transport maps. *Ann. Statist.*, 49(2):1166–1194, 2021. ISSN 0090-5364. doi: 10.1214/20-aos1997. URL https://doi.org/10.1214/20-aos1997.

Hicham Janati, Boris Muzellec, Gabriel Peyré, and Marco Cuturi. Entropic optimal transport between unbalanced gaussian measures has a closed form. *Advances in neural information processing systems*, 33:10468–10479, 2020.

Marcel Klatt, Carla Tameling, and Axel Munk. Empirical regularized optimal transport: statistical theory and applications. *SIAM J. Math. Data Sci.*, 2(2):419–443, 2020. doi: 10.1137/19M1278788. URL https://doi.org/10.1137/19M1278788.

Patrick T Komiske, Eric M Metodiev, and Jesse Thaler. Metric space of collider events. *Physical review letters*, 123(4):041801, 2019.

Daniel Kuhn, Peyman Mohajerin Esfahani, Viet Anh Nguyen, and Soroosh Shafieezadeh-Abadeh. Wasserstein distributionally robust optimization: Theory and applications in machine learning. In *Operations Research & Management Science in the Age of Analytics*, pages 130–166. INFORMS, 2019.

Mengxue Li, Yi-Ming Zhai, You-Wei Luo, Peng-Fei Ge, and Chuan-Xian Ren. Enhanced transport distance for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13936–13944, 2020.

Anton Mallasto, Augusto Gerolin, and Hà Quang Minh. Entropy-regularized 2-wasserstein distance between gaussian measures. *Information Geometry*, pages 1–35, 2021.

Tudor Manole, Sivaraman Balakrishnan, Jonathan Niles-Weed, and Larry Wasserman. Plugin estimation of smooth optimal transport maps. *Pre-Print*, 07 2021. URL https://arxiv.org/pdf/2107.12364.pdf.

Gonzalo Mena and Jonathan Niles-Weed. Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/5acdc9ca5d99ae66afdfe1eea0e3b26b-Paper.pdf.

Stephen L Morgan and David J Harding. Matching estimators of causal effects: Prospects and pitfalls in theory and practice. *Sociological methods & research*, 35(1):3–60, 2006.

Tamás F. Móri and Gábor J. Székely. The earth mover's correlation. *Ann. Univ. Sci. Budapest. Sect. Comput.*, 50:249–268, 2020. ISSN 0138-9491. doi: 10.1007/s10701-018-0208-6. URL https://doi.org/10.1007/s10701-018-0208-6.

Marcel Nutz. Introduction to entropic optimal transport, 2021.

Marcel Nutz and Johannes Wiesel. Entropic optimal transport: Convergence of potentials. *Probability Theory and Related Fields*, pages 1–24, 2021.

Marcel Nutz and Johannes Wiesel. Stability of schrödinger potentials and convergence of sinkhorn's algorithm. *Pre-Print*, 01 2022. URL https://arxiv.org/pdf/2201.10059.pdf.

Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019. ISSN 1935-8237. doi: 10.1561/2200000073. URL http://dx.doi.org/10.1561/2200000073.

Aram-Alexandre Pooladian and Jonathan Niles-Weed. Entropic estimation of optimal transport maps. *Pre-Print*, 09 2021. URL https://arxiv.org/pdf/2109.12004.pdf.

Ievgen Redko, Amaury Habrard, and Marc Sebban. Theoretical analysis of domain adaptation with optimal transport. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 737–753. Springer, 2017.

Philippe Rigollet and Jonathan Weed. Entropic optimal transport is maximum-likelihood deconvolution. *C. R. Math. Acad. Sci. Paris*, 356(11-12):1228–1235, 2018a. ISSN 1631-073X. doi: 10.1016/j.crma.2018.10.010. URL https://doi.org/10.1016/j.crma.2018.10.010.

Philippe Rigollet and Jonathan Weed. Uncoupled isotonic regression via minimum Wasserstein deconvolution. *Information and Inference: A Journal of the IMA*, 2018b.

Laurent Risser, Alberto González Sanz, Quentin Vincenot, and Jean-Michel Loubes. Tackling algorithmic bias in neural-network classifiers using wasserstein-2 regularization. *J. Math Imaging Vis*, 2021.

M Crispim Romão, NF Castro, JG Milhano, R Pedro, and T Vale. Use of a generalized energy mover's distance in the search for rare phenomena at colliders. *arXiv preprint arXiv:2004.09360*, 2020.

Walter Rudin. Real and complex analysis 3rd ed., 1987.

Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943, 2019.

Hongjian Shi, Mathias Drton, Marc Hallin, and Fang Han. Center-outward sign-and rank-based quadrant, Spearman, and Kendall tests for multivariate independence. Available at arXiv:2111.15567v1, 2021a.

Hongjian Shi, Marc Hallin, Mathias Drton, and Fang Han. On universally consistent and fully distribution-free rank tests of vector independence. *Ann. Statist.*, 2021b. ISSN 0090-5364. URL https://arxiv.org/abs/2007.02186. (in press).

Hongjian Shi, Mathias Drton, and Fang Han. Distribution-free consistent independence tests via center-outward ranks and signs. *J. Amer. Statist. Assoc.*, 117(537):395–410, 2022. ISSN 0162-1459. doi: 10.1080/01621459.2020.1782223. URL https://doi.org/10.1080/01621459.2020.1782223.

Max Sommerfeld and Axel Munk. Inference for empirical wasserstein distances on finite spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1): 219–238, 2018.

Carla Tameling, Max Sommerfeld, and Axel Munk. Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications. *Ann. Appl. Probab.*, 29 (5):2744–2781, 10 2019. doi: 10.1214/19-AAP1463. URL https://doi.org/10.1214/19-AAP1463.

Petra E Todd. Matching estimators. In *Microeconometrics*, pages 108–121. Springer, 2010.

Aad W. Van Der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer, New York, NY, 1996.

Aad van der Vaart. Bracketing smooth functions. *Stochastic Processes and their Applications*, 52(1):93–105, 1994. ISSN 0304-4149. doi: https://doi.org/10.1016/0304-4149(94)90102-3. URL https://www.sciencedirect.com/science/article/pii/0304414994901023.

Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A):2620 – 2648, 2019. doi: 10.3150/18-BEJ1065. URL https://doi.org/10.3150/18-BEJ1065.

Karren Dai Yang, Karthik Damodaran, Saradha Venkatachalapathy, Ali C Soylemezoglu, GV Shivashankar, and Caroline Uhler. Predicting cell lineages using autoencoders and optimal transport. *PLoS computational biology*, 16(4):e1007828, 2020.

CHAPTER 5

Weak limits of entropy regularized Optimal Transport; potentials, plans and divergences

The content of this chapter is fully available online in González-Sanz et al. (2022).

Contents

This work deals with the asymptotic distribution of both potentials and couplings of entropic regularized optimal transport for compactly supported probabilities in $\mathbb{R}^d$. We first provide the central limit theorem of the Sinkhorn potentials—the solutions of the dual problem—as a Gaussian process in $\mathcal{C}^s(\Omega)$. Then we obtain the weak limits of the couplings—the solutions of the primal problem—evaluated on integrable functions, proving a conjecture of Harchaoui et al. (2020). In both cases, their limit is a real Gaussian random variable. Finally we consider the weak limit of the entropic Sinkhorn divergence under both assumptions $H_0 : \mathrm{P} = \mathrm{Q}$ or $H_1 : \mathrm{P} \neq \mathrm{Q}$. Under H_0 the limit is a quadratic for applied to a Gaussian process in a Sobolev space, while under H_1, the limit is Gaussian. Such results enable statistical inference based on entropic regularized optimal transport.

5.1 Introduction

Optimal transport has proven its effectiveness as a powerful tool in statistical data analysis. Formulated as the minimization problem, it reads

$$\mathcal{T}_2(\mathrm{P}, \mathrm{Q}) = \min_{\pi \in \Pi(\mathrm{P},\mathrm{Q})} \int_{\mathbb{R}^d \times \mathbb{R}^d} \tfrac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2 d\pi(\mathbf{x}, \mathbf{y}),$$

where $\Pi(\mathrm{P}, \mathrm{Q})$ denotes the set of couplings between the probabilities P and Q, optimal transport provides a notion of discrepancy between distributions (eg. Chapter 7 in Villani (2003)) useful for testing similarity between probabilities (eg. del Barrio et al. (1999, 2005); González-Delgado et al. (2021); Del Barrio et al. (2019) among others) and making inference. For this, it is necessary to know the weak limit of $\mathcal{T}_2(\mathrm{P}_n, \mathrm{Q}_m)$ when the empirical measures P_n and Q_m are used in the place of the population ones, P and Q. Unfortunately, in general dimension—with the exception of perhaps a few simplified cases (eg. del Barrio et al. (2021b); Klatt et al. (2020))—the limit is unknown. Moreover, the rate is slower than the usual parametric rate $\sqrt{n}$ (eg. Weed and Bach (2019); Fournier and Guillin (2013)). This motivates the use of regularization methods for optimal transport since they are not affected by the curse of dimension, such as entropic regularization (eg. Cuturi (2013)) or sliced version (eg. Rabin et al. (2011)). Regularized optimal transport is nowadays used for many practical applications such as domain adaptation (eg. Courty et al. (2017)), counterfactual explanations (eg. de Lara et al. (2021)), music transcription (eg. Flamary et al. (2016)), diffeomorphic registration (eg. Feydy et al. (2017), De Lara et al. (2022)) or measure colocalization in super-resolution images (eg. Klatt et al. (2020)).

Thought weak limits are known for the optimal value in the regularized optimal transport problem, less is known about the distributional limits of the optimizers themselves. In this paper we provide the limits of the empirical solutions of the primal problem (plans/couplings), the dual problem (potentials) and the celebrated Sinkhorn divergence (eg. Genevay et al. (2018)).

Let $\Omega \subset \mathbb{R}^d$ be a compact set. The entropic regularized optimal transport cost between two probabilities $\mathrm{P}, \mathrm{Q} \in \mathcal{P}(\Omega)$ is defined as the solution to the optimization problem

$$S_\epsilon(\mathrm{P}, \mathrm{Q}) = \min_{\pi \in \Pi(\mathrm{P},\mathrm{Q})} \int_{\mathbb{R}^d \times \mathbb{R}^d} \tfrac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 d\pi(\mathbf{x},\mathbf{y}) + \epsilon H(\pi|\mathrm{P}\times\mathrm{Q}), \tag{5.1}$$

where the relative entropy between two probability measures α and β is written as $H(\alpha|\beta) = \int \log(\frac{d\alpha}{d\beta}(x))d\alpha(x)$ if α is absolutely continuous with respect to β, $\alpha \ll \beta$, and $+\infty$ otherwise. We denote the solution of (5.1) by $\pi_{\mathrm{P},\mathrm{Q}}$. This problem can also be written in its dual formulation

$$S_\epsilon(\mathrm{P}, \mathrm{Q}) = \sup_{\substack{f \in L_1(\mathrm{P}) \\ g \in L_1(\mathrm{Q})}} \mathbb{E}\left(f(\mathbf{X}) + g(\mathbf{Y}) - \epsilon\, e^{\frac{f(\mathbf{X})+g(\mathbf{Y})-\frac{1}{2}||\mathbf{X}-\mathbf{Y}||^2}{\epsilon}} \right) + \epsilon, \tag{5.2}$$

with $\mathbf{X} \sim \mathrm{P},\ \mathbf{Y} \sim \mathrm{Q}$. Without any loss of generality we can assume that

$$\mathbb{E}g(\mathbf{Y}) = 0. \tag{5.3}$$

The optimal solutions of this problem, defined as $(f_{\mathrm{P},\mathrm{Q}}, g_{\mathrm{P},\mathrm{Q}})$, are uniquely $\mathrm{P}\times\mathrm{Q}$-a.s. determined by the relation

$$\mathbb{E}\left((h_1(\mathbf{X}) + h_2(\mathbf{Y}))\, e^{\frac{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X})+g_{\mathrm{P},\mathrm{Q}}(\mathbf{Y})-\frac{1}{2}||\mathbf{X}-\mathbf{Y}||^2}{\epsilon}} \right) = \mathbb{E}\left(h_1(\mathbf{X}) + h_2(\mathbf{Y})\right), \tag{5.4}$$

for all $h_1, h_2 \in \mathcal{C}(\Omega)$, which is usually called *optimality condition*. There exists a unique extension of the solutions of (5.4) to the space $\mathcal{C}^\alpha(\Omega) \times \mathcal{C}^\alpha(\Omega)$, given by the relations

$$\begin{aligned} f_{\mathrm{P,Q}} &= -\epsilon \log \left(\int e^{\frac{g_{\mathrm{P,Q}}(\mathbf{y}) - \frac{1}{2}\|\cdot - \mathbf{y}\|^2}{\epsilon}} d\mathrm{Q}(\mathbf{y}) \right) \\ g_{\mathrm{P,Q}} &= -\epsilon \log \left(\int e^{\frac{f_{\mathrm{P,Q}}(\mathbf{x}) - \frac{1}{2}\|\mathbf{x} - \cdot\|^2}{\epsilon}} d\mathrm{P}(\mathbf{x}) \right) \end{aligned} \tag{5.5}$$

This relation has an important consequence—the optimization class can be reduced from $L_1(\mathrm{Q})$ to $\mathcal{C}^s(\Omega)$, with, moreover, uniformly (for all $\mathrm{P}, \mathrm{Q} \in \mathcal{P}(\Omega)$) bounded derivatives, see Genevay et al. (2019); Mena and Niles-Weed (2019). Since the class $\mathcal{C}^s(\Omega)$, with an appropriate choice of s, is uniformly Donsker (eg. Section 2.7.1. in Vaart and Wellner (1996)), one can obtain the bound $\sqrt{n}\,\mathbb{E}|S_\epsilon(\mathrm{P}_n, \mathrm{Q}) - S_\epsilon(\mathrm{P}, \mathrm{Q})| \leq C_\Omega$, where the constant C_Ω depends polynomially on $\mathrm{diam}(\Omega)$, see Mena and Niles-Weed (2019). Moreover, since the convergence of $\mathbb{E}S_1(\mathrm{P}_n, \mathrm{Q})$ towards its population counterpart is faster (eg. del Barrio et al. (2022)) and the fluctuations are asymptotically Gaussian (del Barrio and Loubes (2019); Mena and Niles-Weed (2019); del Barrio et al. (2021a)), the weak limit

$$\sqrt{n}(S_\epsilon(\mathrm{P}_n, \mathrm{Q}) - S_1(\mathrm{P}, \mathrm{Q})) \xrightarrow{w} N(0, \mathrm{Var}_{\mathbf{X}\sim\mathrm{P}}(f^\epsilon_{\mathrm{P,Q}}(\mathbf{X}))) \tag{5.6}$$

holds. (5.6) gives a weak limit for $S_\epsilon(\mathrm{P}_n, \mathrm{Q})$, but does not give information about the coupling $\pi_{\mathrm{P,Q}}$ or the optimal dual variables $(f_{\mathrm{P,Q}}, g_{\mathrm{P,Q}})$.

This provides asymptotic confidence intervals for the population Sinkhorn cost. However, $S_1(\mathrm{P}, \mathrm{Q})$ approaches 0 does not imply that both probabilities are close, hence the asymptotic behaviour of the Sinkhorn cost given in (5.6) does not provide a useful hypothesis contrast for statistical inference.

For statistical inference, obtaining the asymptotic behaviour of optimal regularized couplings, i.e. the limits of

$$\sqrt{\frac{n\,m}{n+m}} \int \eta\,(d\pi^\epsilon_{\mathrm{P}_n,\mathrm{Q}_m} - d\pi^\epsilon_{\mathrm{P,Q}}), \quad \text{where } \eta \in L^2(\mathrm{P}\times\mathrm{Q}), \tag{5.7}$$

is very helpful. The knowledge of the asymptotic behaviour of (5.7) provides consistent confidence intervals for the inference on $\pi^\epsilon_{\mathrm{P}_n,\mathrm{Q}_m}$, which gives theoretical support to the previously cited applications. With regards to the limit of (5.7), the recent work Harchaoui et al. (2020) proved, for modified regularization procedure inspired by Schrodinger's lazy gas experiment, the convergence towards

$$\sqrt{\frac{n\,m}{n+m}} \left(\int \eta d\pi^\epsilon_{\mathrm{P}_n,\mathrm{Q}_m} - \int \eta d\pi^\epsilon_{\mathrm{P,Q}} \right) \xrightarrow{w} N(0, \sigma^2_{\lambda,\epsilon}(\eta)), \quad \eta \in L^2(\mathrm{P}\times\mathrm{Q}), \tag{5.8}$$

where the variance $\sigma^2_\lambda(\eta)$ is

$$\begin{aligned} \lambda\,\mathrm{Var}_{\mathbf{X}\sim\mathrm{P}} &\left((1 - \mathcal{A}^\epsilon_{\mathrm{Q}}\mathcal{A}^\epsilon_{\mathrm{P}})^{-1}(\eta^\epsilon_{\mathbf{x}} - \mathcal{A}^\epsilon_{\mathrm{Q}}\eta^\epsilon_{\mathbf{y}})(\mathbf{X})\right) \\ &+ (1-\lambda)\,\mathrm{Var}_{\mathbf{Y}\sim\mathrm{Q}}\left((1 - \mathcal{A}^\epsilon_{\mathrm{P}}\mathcal{A}^\epsilon_{\mathrm{Q}})^{-1}(\eta^\epsilon_{\mathbf{y}} - \mathcal{A}^\epsilon_{\mathrm{P}}\eta^\epsilon_{\mathbf{x}})(\mathbf{Y})\right), \end{aligned}$$

see section 5.3 for the precise definitions of the operators $\mathcal{A}_{\mathrm{P}}^{\epsilon}, \mathcal{A}_{\mathrm{Q}}^{\epsilon}$ and section 5.3 for the ones of $\eta_{\mathbf{x}}^{\epsilon}, \eta_{\mathbf{y}}^{\epsilon}$. Moreover, the authors of Harchaoui et al. (2020) conjectured that the distributional limit (5.8) holds also for the classic Sinkhorn regularization. Gunsilius and Xu (2021) proved that (5.7) is tight and the limit is centered, however the conjecture remained opened. In Theorem 5.3.1 we prove that (5.8) holds for compactly supported measures, and therefore the conjecture of Harchaoui et al. (2020) is true.

Theorem 5.3.1 is derived as a consequence of the first order linearization of the potentials, described in Theorem 5.2.1, whose proof is based on a reformulation of the optimality conditions (5.5) as a Z-estimation problem (see for instance Vaart and Wellner (1996)). Differentiating in the Fréchet sense the objective function and using the uniform bounds provided by del Barrio et al. (2022), the problem is reduced to the continuity and existence of the following self-operator in $\mathcal{C}_0^s(\Omega)$

$$\begin{pmatrix} (1-\mathcal{A}_{\mathrm{Q}}^{\epsilon}\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon})^{-1} & -(1-\mathcal{A}_{\mathrm{Q}}^{\epsilon}\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon})^{-1}\mathcal{A}_{\mathrm{Q}}^{\epsilon} \\ -\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon}(1-\mathcal{A}_{\mathrm{Q}}^{\epsilon}\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon})^{-1} & (1-\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon}\mathcal{A}_{\mathrm{Q}}^{\epsilon})^{-1} \end{pmatrix}, \tag{5.9}$$

which follows from Fredholm alternative (cf. Theorem 6.6. in Brezis (2010) eg.). As a consequence, Theorem 5.2.1 yields the limits, if $m=m(n)\to\infty$ and $\frac{m}{n+m}\to\lambda\in(0,1)$,

$$\sqrt{\frac{nm}{n+m}}\begin{pmatrix} f_{\mathrm{P}_n,\mathrm{Q}_m}-f_{\mathrm{P},\mathrm{Q}} \\ g_{\mathrm{P}_n,\mathrm{Q}_m}-g_{\mathrm{P},\mathrm{Q}} \end{pmatrix} \to \begin{pmatrix} \sqrt{1-\lambda}(1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathbb{G}_{\mathrm{Q},s}-\sqrt{\lambda}(1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathcal{A}_{\mathrm{Q}}\mathbb{G}_{\mathrm{P},s} \\ \sqrt{\lambda}(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\mathbb{G}_{\mathrm{P},s}-\sqrt{1-\lambda}\mathcal{A}_{\mathrm{P}}(1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathbb{G}_{\mathrm{Q},s} \end{pmatrix},$$

weakly in $\mathcal{C}^s(\Omega)\times\mathcal{C}_0^s(\Omega)$, where $\mathbb{G}_{\mathrm{P},s}$ and $\mathbb{G}_{\mathrm{Q},s}$ are the centered processes with covariance functions (5.12). Moreover, in the one-sample case;

$$\sqrt{n}\begin{pmatrix} f_{\mathrm{P}_n,\mathrm{Q}}-f_{\mathrm{P},\mathrm{Q}} \\ g_{\mathrm{P}_n,\mathrm{Q}}-g_{\mathrm{P},\mathrm{Q}} \end{pmatrix} \to -\begin{pmatrix} (1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathcal{A}_{\mathrm{Q}}\mathbb{G}_{\mathrm{P},s} \\ -(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\mathbb{G}_{\mathrm{P},s} \end{pmatrix},$$

weakly in $\mathcal{C}^s(\Omega)\times\mathcal{C}_0^s(\Omega)$. Theorem 5.2.1, apart from being interesting in itself, has many applications since, among other things, its derivative approximates the transport map (eg. Pooladian and Niles-Weed (2021)), which is also a useful tool for inference.

The regularized transport cost is easier to compute than the usual optimal transport cost but is unsuitable for two-sample testing, since $S_\epsilon(\mathrm{P},\mathrm{P})\neq 0$. In Genevay et al. (2018), the authors provides to remedy this deficiency by defining the quadratic Sinkhorn's divergence:

$$D_\epsilon(\mathrm{P},\mathrm{Q}) = S_\epsilon(\mathrm{P},\mathrm{Q}) - \tfrac{1}{2}\left(S_\epsilon(\mathrm{P},\mathrm{P}) + S_\epsilon(\mathrm{Q},\mathrm{Q})\right).$$

This definition satisfies several atractive properties: it is symmetric in P,Q and $D_\epsilon(\mathrm{P},\mathrm{Q})\geq 0$, with $D_\epsilon(\mathrm{P},\mathrm{Q})=0$ if and only if $\mathrm{P}=\mathrm{Q}$ (see Theorem 1 in Feydy et al. (2019)). Now, apparently this is an effective way to measure discrepancies between distributions.

The weak limit of the empirical Sinkhorn's divergence, described in Theorem 5.4.3, has different rates depending on the hypotheses $H_0:\ \mathrm{P}=\mathrm{Q}$ or $H_1:\ \mathrm{P}\neq\mathrm{Q}$. Under H_1 the limit can be derived by means of Efron-Stein linearization (del Barrio and Loubes (2019); Mena and Niles-Weed (2019); del Barrio et al. (2021a); González-Delgado et al. (2021)), giving

$$\sqrt{\frac{n\,m}{n+m}}(D_\epsilon(\mathrm{P}_n,\mathrm{Q}_m)-D_\epsilon(\mathrm{P},\mathrm{Q}))\xrightarrow{w} N(0,\lambda\,\mathrm{Var}_{\mathrm{P}}(\psi^\epsilon_{\mathrm{P},\mathrm{Q}})+(1-\lambda)\,\mathrm{Var}_{\mathrm{Q}}(\psi^\epsilon_{\mathrm{Q},\mathrm{P}})),$$

and

$$\sqrt{n}(D_\epsilon(\mathrm{P}_n,\mathrm{Q})-D_\epsilon(\mathrm{P},\mathrm{Q}))\xrightarrow{w} N(0,\mathrm{Var}_{\mathrm{P}}(\psi^\epsilon_{\mathrm{P},\mathrm{Q}})),$$

where $\psi^\epsilon_{\mathrm{P},\mathrm{Q}}=f^\epsilon_{\mathrm{P},\mathrm{Q}}-\frac{1}{2}(f^\epsilon_{\mathrm{P},\mathrm{P}}+g^\epsilon_{\mathrm{P},\mathrm{P}})$ and $\psi^\epsilon_{\mathrm{Q},\mathrm{P}}=f^\epsilon_{\mathrm{Q},\mathrm{P}}-\frac{1}{2}(f^\epsilon_{\mathrm{Q},\mathrm{Q}}+g^\epsilon_{\mathrm{Q},\mathrm{Q}})$. Under H_0, however, $\psi^\epsilon_{\mathrm{P},\mathrm{Q}}=\psi^\epsilon_{\mathrm{Q},\mathrm{P}}=0$ so the limit is degenerate, and in fact $D_\epsilon(\mathrm{P}_n,\mathrm{Q})=O_P(\frac{1}{n})$. To obtain a non-trivial limit, we therefore conduct a second order analysis. Our proof relies on Theorem 5.2.1 and use of the embedding theorem of the Hilbert space $W^{2s,2}(\Omega)$ into $\mathcal{C}^s(\Omega)$ (cf. Theorem 6.3 in Adams and Fournier (2003) eg.). The limit, in this case, behaves as

$$n\,D_1(\mathrm{P}_n,\mathrm{P})\xrightarrow{w}\frac{1}{4}\sum_{j=1}^\infty\lambda_j\left(\sum_{i=1}^\infty x_{i,j}N_i\right)\left(\sum_{i=1}^\infty x_{i,j}N_i\right)$$

and

$$\frac{n\,m}{n+m}(D_\epsilon(\mathrm{P}_n,\mathrm{P}'_m)-D_\epsilon(\mathrm{P},\mathrm{Q}))\xrightarrow{w}$$
$$\frac{1}{4}\sum_{j=1}^\infty\lambda_j\left(\sum_{i=1}^\infty x_{i,j}(\sqrt{\lambda}N_i+\sqrt{1-\lambda}N'_i)\right)\left(\sum_{i=1}^\infty x_{i,j}(\sqrt{\lambda}N_i+\sqrt{1-\lambda}N'_i)\right),$$

where $\{N_i\}_{i\in\mathbb{N}}$ and $\{N'_i\}_{i\in\mathbb{N}}$ are mutually i.i.d. with $N_i,N'_i\sim N(0,1)$ and the sequences $\{x_{i,j}\}_{i,j\in\mathbb{N}}\subset\mathbb{R}$ and $\{\lambda_j\}_{j\in\mathbb{N}}\subset[0,\infty)$ are positive sequences depending on P and ϵ.

In recent years there has been a substantial body of work studying the weak limits of the optimal transport problem. Since, for obvious reasons we cannot cite every one of them, we refer to Hundrieser et al. (2022) for a comprehensive survey. Focusing on the entropic regularized optimal transport, similar results for finitely supported measures are obtained by Klatt et al. (2020) and Bigot et al. (2019). The limits of the regularized cost (5.6) has been proven first in del Barrio et al. (2022) and Mena and Niles-Weed (2019)—using the Efron-Stein linearization–and then in Goldfeld et al. (2022)—using Hadamard linearization. The conjecture of Harchaoui et al. (2020) has been also investigated in Gunsilius and Xu (2021), under slightly different assumptions. That work shows the existence of a weak limit for (5.8), which they conjecture is Gaussian. We prove their conjecture. They also derive a similar result to Theorem 5.2.1, but in a weaker norm. We prove convergence in the space $\mathcal{C}^s(\Omega)\times\mathcal{C}^s_0(\Omega)$, which allows us to derive the limit of the Sinkhorns divergence.

5.1.1 Outline of the paper

The paper is organized as follows. The notation is given in Section (5.1.2). The central limit for the regularized optimal transport potentials and its proof can be found in Section 5.2, however auxiliary Lemmas are proven in Section 5.5. In Section 5.3 we enunciate and prove the central limit theorem for the couplings, Theorem 5.3.1, and its immediate consequence, Corollary 5.3.2. Section 5.4 deals with the weak limits of the Sinkhorn divergence, formally stated in Theorems 5.4.1 and 5.4.3 . Also in this section the reader can find a discussion about the simplification of the limit by embedding it into a Hilbert space and the proof of the main result. Auxiliary results and their proofs are postponed to Section 5.6.

5.1.2 Notations

For the reader's convenience, this section sets the notation used throughout this work. Unless otherwise specified, the probabilities P and Q are supported in the compact set $\Omega \subset \mathbb{R}^d$, meaning that $\mathrm{P}, \mathrm{Q} \in \mathcal{P}(\Omega)$.

For two probabilities $\mu, \nu \in \mathcal{P}(\Omega)$ the pair $(f^{\epsilon}_{\mu,\nu}, g^{\epsilon}_{\mu,\nu})$ is one solving (5.2) for $S_1(\mu, \nu)$, its direct sum $((\mathbf{x}, \mathbf{y}) \mapsto f^{\epsilon}_{\mu,\nu}(\mathbf{x}) + g^{\epsilon}_{\mu,\nu}(\mathbf{y}))$ is denoted by $h^{\epsilon}_{\mu,\nu}$. The solution of (5.1) is denoted by $\pi^{\epsilon}_{\mu,\nu}$ and its density w.r.t. $d\mu d\nu$ by $\xi^{\epsilon}_{\mu,\nu}$. For the particular case of $\epsilon = 1$, the super-index is avoided.

We set $s = \lceil \frac{d}{2} \rceil + 1$, and denote, for any $\alpha \in \mathbb{N}$, the space of all functions on Ω that possess uniformly bounded partial derivatives up to order α as $\mathcal{C}^{\alpha}(\Omega)$, in which we consider the norm

$$\|f\|_{\mathcal{C}^{\alpha}(\Omega)} = \|f\|_{\alpha} = \sum_{i=0}^{\alpha} \sum_{|\beta|=i} \|D^{\beta} f\|_{\infty}.$$

In the product space $\mathcal{C}^{\alpha}(\Omega) \times \mathcal{C}^{\alpha}(\Omega)$ we consider the norm

$$\|(f,g)\|_{\mathcal{C}^{\alpha}(\Omega)\times\mathcal{C}^{\alpha}(\Omega)} = \|(f,g)\|_{\alpha\times\alpha} = \|f\|_{\alpha} + \|g\|_{\alpha}.$$

The Sobolev space $W^{\alpha,2}(\Omega)$ consists of (equivalence classes of) functions $L^2(\Omega)$–the Hilbert space of square Lebesgue (ℓ_d) integrable functions–whose distributional derivatives up to order $|a| \leq \alpha$ also belong $L^2(\Omega)$. $W^{\alpha,2}(\Omega)$ is a separable Hilbert space with the norm

$$\|f\|_{W^{\alpha,2}(\Omega)} = \left(\sum_{i=0}^{\alpha} \sum_{|\beta|=i} \|D^{\beta} f\|^2_{L^2(\Omega)} \right)^{\frac{1}{2}},$$

and, when it does not cause confusion, we use $\|f\|_{W^{\alpha,2}(\Omega)} = \|f\|_W$ and $\langle \cdot, \cdot \rangle_W$ for its inner product.

Moreover, for a general Banach space $\mathcal{H}$, the norm is denoted as $\|\cdot\|_{\mathcal{H}}$. Unless otherwise stated, the random vectors $\mathbf{X}$ and $\mathbf{Y}$ are independent and follow respectively the distributions

P and Q. For a measurable function $f : \Omega \to \mathbb{R}$, the following expressions are equivalent:

$$\mathbb{E}(f(\mathbf{X})) = \mathbb{E}_{\mathbf{X}\sim \mathrm{P}}(f(\mathbf{X})) = \int f(\mathbf{x})d\mathrm{P}(\mathbf{x}) = \int f d\mathrm{P} = \mathrm{P}(f).$$

Sometimes, when the function f depends also in some others random variables $f = f(\mathbf{X}, \mathbf{Z}_1, \dots, \mathbf{Z}_m)$, we use

$$\mathbb{E}_{\mathbf{X}\sim \mathrm{P}}(f(\mathbf{X}, \mathbf{Z}_1, \dots, \mathbf{Z}_m)) = \int f(\mathbf{x}, \mathbf{Z}_1, \dots, \mathbf{Z}_m)d\mathrm{P}(\mathbf{x}) = \mathrm{P}^{\mathbf{X}}(f(\mathbf{X}, \mathbf{Z}_1, \dots, \mathbf{Z}_m)).$$

The function $\mathbf{y} \mapsto \int f(\mathbf{x}, \mathbf{y})d\mathrm{P}(\mathbf{x})$ is denoted by

$$\mathbb{E}_{\mathbf{X}\sim \mathrm{P}}(f(\mathbf{X}, \cdot) = \int f(\mathbf{x}, \cdot)d\mathrm{P}(\mathbf{x}) = \mathrm{P}^{\mathbf{X}}(f(\mathbf{X}, \cdot)).$$

Let $\{w_n\}_{n\in} \subset \mathcal{H}$ be a random sequence and $\{a_n\}_{n\in\mathbb{N}}$ be a real random sequence, the notation $w_n = o_P(a_n)$ means that the sequence w_n/a_n tends to 0 in probability, and $w_n = O_P(a_n)$ that w_n/a_n is stochastically bounded, for further details see Section 2.2 in Vaart (1998). Finally, for a Borel measure μ in $\mathbb{R}^d$, $L^2(\mu)$ denotes the space of square integrable functions. The spacial case of the Lebesgue measure ℓ_d in Ω is denoted by $L^2(\Omega)$

5.2 Central Limit Theorem of Sinkhorn potentials

The main result of this section is Theorem 5.2.1, which gives the first-order linearization of the difference between the empirical and population Sinkhorn potentials. As a consequence, Corollary 5.2.2 gives the central limit theorem of that difference with rate $\sqrt{\frac{nm}{m+n}}$. This section contains a summary of the proof of Theorem 5.2.1, highlighting the steps that are useful for further sections, as well as a full proof following the statement of Corollary 5.2.2. The proofs of auxiliary results are postponed, except that of Lemma 5.2.5, which is included in this section due to the interest of the technique, which is repeated several times in the rest of the work. Set $\mathrm{P}, \mathrm{Q} \in \mathcal{P}(\Omega)$, the shape of the first order development depends on the operators

$$\mathcal{A}^{\epsilon}_{\mathrm{P}} : L^2(\mathrm{P}) \ni f \mapsto \int \xi^{\epsilon}_{\mathrm{P},\mathrm{Q}}(\mathbf{x}, \cdot)f(\mathbf{x})d\mathrm{P}(\mathbf{x}) \in \mathcal{C}^{\alpha}(\Omega),$$
$$\mathcal{A}^{\epsilon}_{\mathrm{Q}} : L^2_0(\mathrm{Q}) \ni g \mapsto \int \xi^{\epsilon}_{\mathrm{P},\mathrm{Q}}(\cdot, \mathbf{y})g(\mathbf{y})d\mathrm{Q}(\mathbf{y}) \in \mathcal{C}^{\alpha}(\Omega).$$

Theorem 5.2.1. *Let $\Omega \subset \mathbb{R}^d$ be a compact set, $\mathrm{P}, \mathrm{Q} \in \mathcal{P}(\Omega)$ and P_n (resp. Q_m) be the empirical measure of the i.i.d. sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ (resp. $\mathbf{Y}_1, \dots, \mathbf{Y}_m$) distributed as P (resp. Q). If $m = m(n) \to \infty$ and $\frac{m}{n+m} \to \lambda \in (0, 1)$,*

$$\begin{pmatrix} f_{\mathrm{P}_n,\mathrm{Q}_m} - f_{\mathrm{P},\mathrm{Q}} \\ g_{\mathrm{P}_n,\mathrm{Q}_m} - g_{\mathrm{P},\mathrm{Q}} \end{pmatrix} = \begin{pmatrix} (1 - \mathcal{A}_{\mathrm{Q}}\mathcal{A}^{\epsilon}_{\mathrm{P}})^{-1}\mathcal{A}_{\mathrm{Q}}\mathbb{G}^{n}_{\mathrm{P},s} - (1 - \mathcal{A}_{\mathrm{Q}}\mathcal{A}^{\epsilon}_{\mathrm{P}})^{-1}\mathbb{G}^{m}_{\mathrm{Q},s} \\ \mathcal{A}^{\epsilon}_{\mathrm{P}}(1 - \mathcal{A}_{\mathrm{Q}}\mathcal{A}^{\epsilon}_{\mathrm{P}})^{-1}\mathbb{G}^{m}_{\mathrm{Q},s} - (1 - \mathcal{A}^{\epsilon}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\mathbb{G}^{n}_{\mathrm{P},s} \end{pmatrix} + o_P\left(\sqrt{\frac{n+m}{nm}}\right)$$

in $\mathcal{C}^\alpha(\Omega)\times\mathcal{C}^\alpha(\Omega)$*, for any* $\alpha\in\mathbb{N}$*, and*

$$\left(\begin{array}{c}\mathbb{G}^n_{\mathrm{P},s}\\ \mathbb{G}^m_{\mathrm{Q},s}\end{array}\right)=\left(\begin{array}{c}\frac{1}{n}\sum_{k=1}^n \xi_{\mathrm{P,Q}}(\mathbf{X}_k,\cdot)-\mathbb{E}\left(\xi_{\mathrm{P,Q}}(\mathbf{X},\cdot)\right)\\ \frac{1}{m}\sum_{k=1}^m \xi_{\mathrm{P,Q}}(\cdot,\mathbf{Y}_k)-\mathbb{E}\left(\xi_{\mathrm{P,Q}}(\cdot,\mathbf{Y})\right)\end{array}\right).$$

Both $\sqrt{n}\mathbb{G}^n_{\mathrm{P},s}$ and $\sqrt{m}\mathbb{G}^m_{\mathrm{Q},s}$ converge to Gaussian random variables—the proof is a consequence of the embedding theorem of the Hilbert space $W^{2s,2}(\Omega)$ into $\mathcal{C}^s(\Omega)$ (cf. Theorem 6.3. in Adams and Fournier (2003)) and the central limit theorem in Hilbert spaces (cf. Theorem 10.5. in Ledoux and Talagrand (1991)). The dependency relationship between these limit processes is inherited from that of the samples; for instance, if they are independent the limit of

$$\sqrt{\frac{n\,m}{n+m}}\left(\begin{array}{c}\mathbb{G}^n_{\mathrm{P},s}\\ \mathbb{G}^m_{\mathrm{Q},s}\end{array}\right)\tag{5.10}$$

is the pair of centered independent Gaussian processes

$$\left(\begin{array}{c}\sqrt{\lambda}\,\mathbb{G}_{\mathrm{Q},s}\\ \sqrt{1-\lambda}\,\mathbb{G}_{\mathrm{P},s}\end{array}\right)\in\mathcal{C}^s(\Omega)\times\mathcal{C}^s_0(\Omega),\tag{5.11}$$

uniquely characterized by the covariance functions

$$\begin{aligned}\mathbb{E}\left(\mathbb{G}_{\mathrm{P},s}(\mathbf{y})\mathbb{G}_{\mathrm{P},s}(\mathbf{y}')\right)&=\frac{\mathbb{E}\left(e^{f_{\mathrm{P,Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\mathbf{y}\|^2}e^{f_{\mathrm{P,Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\mathbf{y}'\|^2}\right)}{\mathbb{E}\left(e^{f_{\mathrm{P,Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\mathbf{y}\|^2}\right)\mathbb{E}\left(e^{f_{\mathrm{P,Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\mathbf{y}'\|^2}\right)}-1\quad\text{and}\\ \mathbb{E}\left(\mathbb{G}_{\mathrm{Q},s}(\mathbf{x})\mathbb{G}_{\mathrm{Q},s}(\mathbf{x}')\right)&=\frac{\mathbb{E}\left(e^{g_{\mathrm{P,Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{x}-\mathbf{Y}\|^2}e^{g_{\mathrm{P,Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{x}'-\mathbf{Y}\|^2}\right)}{\mathbb{E}\left(e^{g_{\mathrm{P,Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{x}-\mathbf{Y}\|^2}\right)\mathbb{E}\left(e^{g_{\mathrm{P,Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{x}'-\mathbf{Y}\|^2}\right)}-1.\end{aligned}\tag{5.12}$$

However, if $\mathrm{Q}_m=\mathrm{P}_n$, the limit of (5.10) is $\frac{1}{\sqrt{2}}\left(\begin{array}{c}\mathbb{G}_{\mathrm{P},s}\\ \mathbb{G}_{\mathrm{P},s}\end{array}\right)\in\mathcal{C}^s(\Omega)\times\mathcal{C}^s_0(\Omega)$. As a consequence we obtain the central limit theorem for the potentials.

Corollary 5.2.2. *Let* $\Omega\subset\mathbb{R}^d$ *be a compact set,* $\mathrm{P},\mathrm{Q}\in\mathcal{P}(\Omega)$ *and* P_n *(resp.* Q_m*) be the empirical measure of the i.i.d. sample* $\mathbf{X}_1,\dots,\mathbf{X}_n$ *(resp.* $\mathbf{Y}_1,\dots,\mathbf{Y}_m$*) distributed as* P *(resp.* Q*). Set* $\alpha\in\mathbb{N}$ *and suppose that both samples are mutually independent. Then, if* $m=m(n)\to\infty$ *and* $\frac{m}{n+m}\to\lambda\in(0,1)$*,*

$$\sqrt{\frac{n\,m}{n+m}}\left(\begin{array}{c}f_{\mathrm{P}_n,\mathrm{Q}_m}-f_{\mathrm{P,Q}}\\ g_{\mathrm{P}_n,\mathrm{Q}_m}-g_{\mathrm{P,Q}}\end{array}\right)\to\left(\begin{array}{c}\sqrt{1-\lambda}(1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathbb{G}_{\mathrm{Q},s}-\sqrt{\lambda}(1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathcal{A}_{\mathrm{Q}}\mathbb{G}_{\mathrm{P},s}\\ \sqrt{\lambda}(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\mathbb{G}_{\mathrm{P},s}-\sqrt{1-\lambda}\mathcal{A}_{\mathrm{P}}(1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathbb{G}_{\mathrm{Q},s}\end{array}\right),$$

weakly in $\mathcal{C}^\alpha(\Omega)\times\mathcal{C}^\alpha(\Omega)$*, where* $\mathbb{G}_{\mathrm{P},s}$ *and* $\mathbb{G}_{\mathrm{Q},s}$ *are the centered Gaussian processes with covariance functions (5.12). Moreover, in the one-sample case;*

$$\sqrt{n}\left(\begin{array}{c}f_{\mathrm{P}_n,\mathrm{Q}}-f_{\mathrm{P,Q}}\\ g_{\mathrm{P}_n,\mathrm{Q}}-g_{\mathrm{P,Q}}\end{array}\right)\to-\left(\begin{array}{c}(1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathcal{A}_{\mathrm{Q}}\mathbb{G}_{\mathrm{P},s}\\ -(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\mathbb{G}_{\mathrm{P},s}\end{array}\right),$$

weakly in $\mathcal{C}^\alpha(\Omega)\times\mathcal{C}^\alpha(\Omega)$*.*

We will give a brief outline of the proof of the Theorem 5.2.1. Due to (5.3), the potentials are uniquely determined in the Banach space $\mathcal{C}^\alpha(\Omega) \times \mathcal{C}_0^\alpha(\Omega)$, where, for $\alpha > 0$, $\mathcal{C}_0^\alpha(\Omega)$ denotes the subspace of $\mathcal{C}^\alpha(\Omega)$ with null expectation w.r.t. Q, i.e.

$$\{f \in \mathcal{C}^\alpha(\Omega) : \quad \int f(\mathbf{y}) d\mathrm{Q}(\mathbf{y}) = 0\}.$$

The relation (5.5) gives a precise characterization of the Sinkhorn potentials. This means that $(f_{\mathrm{P},\mathrm{Q}}, g_{\mathrm{P},\mathrm{Q}})$, $(f^\epsilon_{\mathrm{P}_n,\mathrm{Q}_m}, g^\epsilon_{\mathrm{P}_n,\mathrm{Q}_m}) \in \mathcal{C}^\alpha(\Omega) \times \mathcal{C}_0^\alpha(\Omega)$ are just the solutions of $\Psi(f^\epsilon_{\mathrm{P},\mathrm{Q}}, g^\epsilon_{\mathrm{P},\mathrm{Q}}) = \Psi_{n,m}(f^\epsilon_{\mathrm{P}_n,\mathrm{Q}_m}, g^\epsilon_{\mathrm{P}_n,\mathrm{Q}_m}) = 0$, where

$$\Psi, \Psi_{n,m} : \mathcal{C}^\alpha(\Omega) \times \mathcal{C}_0^\alpha(\Omega) \to \mathcal{C}^\alpha(\Omega) \times \mathcal{C}_0^\alpha(\Omega))$$

are respectively defined as

$$\Psi\begin{pmatrix} f \\ g \end{pmatrix} = \begin{pmatrix} f + \epsilon \log\left(\int e^{\frac{g(\mathbf{y}) - \frac{1}{2}\|\cdot - \mathbf{y}\|^2}{\epsilon}} d\mathrm{Q}(\mathbf{y})\right) \\ g + \epsilon \log\left(\int e^{\frac{f(\mathbf{x}) - \frac{1}{2}\|\mathbf{x} - \cdot\|^2}{\epsilon}} d\mathrm{P}(\mathbf{x})\right) - \int \epsilon \log\left(\int e^{\frac{f(\mathbf{x}) - \frac{1}{2}\|\mathbf{x} - \cdot\|^2}{\epsilon}} d\mathrm{P}(\mathbf{x})\right) d\mathrm{Q}(\mathbf{y}) \end{pmatrix}$$

and

$$\Psi_{n,m}\begin{pmatrix} f \\ g \end{pmatrix} =$$
$$\begin{pmatrix} f + \epsilon \log\left(\frac{1}{m}\sum_{i=1}^m \left(e^{g(\mathbf{Y}_i) - \frac{1}{2}\|\cdot - \mathbf{Y}_i\|^2}\right)\right) \\ g + \epsilon \log\left(\frac{1}{n}\sum_{i=1}^n \left(e^{\frac{f(\mathbf{X}_i) - \frac{1}{2}\|\mathbf{X}_i - \cdot\|^2}{\epsilon}}\right)\right) - \epsilon \int \log\left(\frac{1}{n}\sum_{i=1}^n \left(e^{\frac{f(\mathbf{X}_i) - \frac{1}{2}\|\mathbf{X}_i - \mathbf{y}\|^2}{\epsilon}}\right)\right) d\mathrm{Q}(\mathbf{y}) \end{pmatrix}.$$

Note that we are subtracting the expectation w.r.t. Q in the second component so that the image of Ψ and $\Psi_{n,m}$ lies in $\mathcal{C}^\alpha(\Omega) \times \mathcal{C}_0^\alpha(\Omega)$. The first step consists in differentiating Ψ in $\mathcal{C}^\alpha(\Omega) \times \mathcal{C}_0^\alpha(\Omega)$. The Fréchet derivative in the pair $(f_{\mathrm{P},\mathrm{Q}}, g_{\mathrm{P},\mathrm{Q}})$ gives

$$D_{(f_{\mathrm{P},\mathrm{Q}}, g_{\mathrm{P},\mathrm{Q}})}\Psi : \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \mapsto \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + \begin{pmatrix} \mathcal{A}^\epsilon_{\mathrm{Q}} h_2 \\ \bar{\mathcal{A}}^\epsilon_{\mathrm{P}} h_1 \end{pmatrix} \tag{5.13}$$

The existence of the Fréchet derivative (cf. Shapiro (1990) eg.) implies

$$\frac{\|\Psi((g_{\mathrm{P},\mathrm{Q}}, f_{\mathrm{P},\mathrm{Q}}) + \delta) - \Psi((g_{\mathrm{P},\mathrm{Q}}, f_{\mathrm{P},\mathrm{Q}})) - D_{(f_{\mathrm{P},\mathrm{Q}}, g_{\mathrm{P},\mathrm{Q}})}\Psi_{\mathrm{P},\mathrm{Q}}(\delta)\|_s}{\|\delta\|_{\alpha\times\alpha}} \xrightarrow[\|\delta\|_{\alpha\times\alpha}\to 0]{} 0, \tag{5.14}$$

and, since $\delta_{n,m} = (g_{\mathrm{P}_n,\mathrm{Q}_m}, f_{\mathrm{P}_n,\mathrm{Q}_m}) - (g_{\mathrm{P},\mathrm{Q}}, f_{\mathrm{P},\mathrm{Q}})$ satisfies $\|\delta_{n,m}\|_{\mathcal{C}^\alpha(\Omega)\times\mathcal{C}_0^\alpha(\Omega)} = O_P(\sqrt{\frac{n+m}{nm}})$ (cf. Theorem 4.5 in del Barrio et al. (2022) eg.), also

$$\frac{\|\Psi((g_{\mathrm{P}_n,\mathrm{Q}_m}, f_{\mathrm{P}_n,\mathrm{Q}_m})) - \Psi((g_{\mathrm{P},\mathrm{Q}}, f_{\mathrm{P},\mathrm{Q}})) - D_{(f_{\mathrm{P},\mathrm{Q}}, g_{\mathrm{P},\mathrm{Q}})}\Psi_{\mathrm{P},\mathrm{Q}}(\delta_{n,m})\|_{\alpha\times\alpha}}{\|\delta_{n,m}\|_{\alpha\times\alpha}} \xrightarrow{\mathbb{P}} 0.$$

Therefore, the weak limit of $\sqrt{\frac{nm}{n+m}} D_{(f_{\mathrm{P,Q}}, g_{\mathrm{P,Q}})} \Psi_{\mathrm{P,Q}}(\delta_{n,m})$ is that of

$$\sqrt{\frac{n\,m}{n+m}}(\Psi((g_{\mathrm{P}_n,\mathrm{Q}_m}, f_{\mathrm{P}_n,\mathrm{Q}_m})) - \Psi((g_{\mathrm{P,Q}}, f_{\mathrm{P,Q}}))),$$

which we show satisfies

$$\sqrt{\frac{n\,m}{n+m}}(\Psi\left(\begin{array}{c} g_{\mathrm{P}_n,\mathrm{Q}_m} \\ f_{\mathrm{P}_n,\mathrm{Q}_m} \end{array}\right) - \Psi(\left(\begin{array}{c} g_{\mathrm{P}_n,\mathrm{Q}_m} \\ f_{\mathrm{P}_n,\mathrm{Q}_m} \end{array}\right)) = \sqrt{\frac{n\,m}{n+m}}\left(\begin{array}{c} \mathbb{G}^n_{\mathrm{P},s} \\ \mathbb{G}^m_{\mathrm{Q},s} \end{array}\right) + o_P(1), \quad (5.15)$$

in $\mathcal{C}^\alpha(\Omega) \times \mathcal{C}^\alpha_0(\Omega)$, where $\mathbb{G}^n_{\mathrm{P},s}$ and $\mathbb{G}^m_{\mathrm{Q},s}$ are defined in Theorem 5.2.1.

The final step is the invertibility of the operator $D_{(f_{\mathrm{P,Q}}, g_{\mathrm{P,Q}})}\Psi$ in $\mathcal{C}^\alpha(\Omega) \times \mathcal{C}^\alpha_0(\Omega)$. The proof falls into the following steps. First, we prove that the operator $\mathcal{B} = \left(\begin{array}{c} \mathcal{A}^\epsilon_{\mathrm{Q}} \\ \bar{\mathcal{A}}^\epsilon_{\mathrm{P}} \end{array}\right)$ is compact (i.e. it maps bounded sequences into relatively compact ones); then the Fredholm alternative (cf. Theorem 6.6. in Brezis (2010)) implies then either, for every $h \in \mathcal{C}^\alpha(\Omega) \times \mathcal{C}^\alpha_0(\Omega)$, the equation $D_{(f_{\mathrm{P,Q}}, g_{\mathrm{P,Q}})}\Psi(u) = u + \mathcal{B}u = h$ has a unique solution, or $D_{(f_{\mathrm{P,Q}}, g_{\mathrm{P,Q}})}\Psi(u) = 0$ admits a finite number of non-zero solutions. Therefore, injectivity, which follows from Jensen's inequality, implies bijectivity and also the existence of a continuous inverse (cf. Corollary 2.7. in Brezis (2010)).

Proof of Theorem 5.2.1 and Corollary 5.2.2.

For the ease of notation we restrict the proof to $\epsilon = 1$, and thus the index related to ϵ is omitted. Also we assume that $\alpha \geq s$. The following result proves the Fréchet diferentiability of Ψ, which enables to write (5.14).

Lemma 5.2.3. *The functional Ψ is Fréchet differentiable in $(f_{\mathrm{P,Q}}, g_{\mathrm{P,Q}})$ with derivative* (5.13).

To derive (5.15) we realize that the optimality condition $\Psi_{n,m}\left(\begin{array}{c} f_{\mathrm{P}_n,\mathrm{Q}_m} \\ g_{\mathrm{P}_n,\mathrm{Q}_m} \end{array}\right) = \Psi\left(\begin{array}{c} f_{\mathrm{P,Q}} \\ g_{\mathrm{P,Q}} \end{array}\right) =$ 0 implies

$$(\Psi\left(\begin{array}{c} f_{\mathrm{P}_n,\mathrm{Q}_m} \\ g_{\mathrm{P}_n,\mathrm{Q}_m} \end{array}\right) - \Psi\left(\begin{array}{c} f_{\mathrm{P,Q}} \\ g_{\mathrm{P,Q}} \end{array}\right)) = (\Psi\left(\begin{array}{c} f_{\mathrm{P}_n,\mathrm{Q}_m} \\ g_{\mathrm{P}_n,\mathrm{Q}_m} \end{array}\right) - \Psi_{n,m}\left(\begin{array}{c} f_{\mathrm{P}_n,\mathrm{Q}_m} \\ g_{\mathrm{P}_n,\mathrm{Q}_m} \end{array}\right)) \quad (5.16)$$

and the proof of the following result uses the uniform control

$$\|f_{\mathrm{P}_n,\mathrm{Q}_m} - f_{\mathrm{P,Q}}\|_\alpha,\ \|g_{\mathrm{P}_n,\mathrm{Q}_m} - g_{\mathrm{P,Q}}\|_\alpha = O_P\left(\sqrt{\frac{n+m}{n\,m}}\right)$$

given by Theorem 4.5 in del Barrio et al. (2022) to exchange, up to additive $o_P\left(\sqrt{\frac{n+m}{nm}}\right)$ terms, $(f_{\mathrm{P}_n,\mathrm{Q}_m}, g_{\mathrm{P}_n,\mathrm{Q}_m})$ by $(f_{\mathrm{P,Q}}, g_{\mathrm{P,Q}})$ in (5.16).

Lemma 5.2.4. *The asymptotic equality*

$$(\Psi\begin{pmatrix} f_{\mathrm{P}_n,\mathrm{Q}_m} \\ g_{\mathrm{P}_n,\mathrm{Q}_m}\end{pmatrix} - \Psi\begin{pmatrix} f_{\mathrm{P},\mathrm{Q}} \\ g_{\mathrm{P},\mathrm{Q}}\end{pmatrix}) = (\Psi\begin{pmatrix} f_{\mathrm{P},\mathrm{Q}} \\ g_{\mathrm{P},\mathrm{Q}}\end{pmatrix} - \Psi_{n,m}\begin{pmatrix} f_{\mathrm{P},\mathrm{Q}} \\ g_{\mathrm{P},\mathrm{Q}}\end{pmatrix}) + o_P\left(\sqrt{\frac{n+m}{nm}}\right)$$

holds in $\mathcal{C}^{\alpha}(\Omega)\times\mathcal{C}_0^{\alpha}(\Omega)$.

To conclude we need only to show the invertibility in $\mathcal{C}^{\alpha}(\Omega)\times\mathcal{C}_0^{\alpha}(\Omega)$ of the operator $D_{(f_{\mathrm{P},\mathrm{Q}},g_{\mathrm{P},\mathrm{Q}})}\Psi$. In the following result we adopt matrix notation, i.e.

$$\begin{pmatrix} (1-\mathcal{A}_{\mathrm{Q}}^{\epsilon}\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon})^{-1} & -(1-\mathcal{A}_{\mathrm{Q}}^{\epsilon}\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon})^{-1}\mathcal{A}_{\mathrm{Q}}^{\epsilon} \\ -\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon}(1-\mathcal{A}_{\mathrm{Q}}^{\epsilon}\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon})^{-1} & (1-\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon}\mathcal{A}_{\mathrm{Q}}^{\epsilon})^{-1}\end{pmatrix}\begin{pmatrix} f \\ g\end{pmatrix}$$
$$= \begin{pmatrix} (1-\mathcal{A}_{\mathrm{P}}^{\epsilon}\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon})^{-1}f - (1-\mathcal{A}_{\mathrm{P}}^{\epsilon}\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon})^{-1}\mathcal{A}_{\mathrm{P}}^{\epsilon}g \\ -\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon}(1-\mathcal{A}_{\mathrm{P}}^{\epsilon}\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon})^{-1}f + (1-\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon}\mathcal{A}_{\mathrm{P}}^{\epsilon})^{-1}g\end{pmatrix}.$$

Lemma 5.2.5. *Let* $\Omega\subset\mathbb{R}^d$ *be a compact set,* $\mathrm{P},\mathrm{Q}\in\mathcal{P}(\Omega)$*, then*

(i) $(1-\mathcal{A}_{\mathrm{Q}}^{\epsilon}\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon})$ *and* $(1-\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon}\mathcal{A}_{\mathrm{Q}}^{\epsilon})$ *are continuously invertible operators in* $\mathcal{C}^{\alpha}(\Omega)$ *and* $\mathcal{C}_0^{\alpha}(\Omega)$ *respectively,*

(ii) $\mathcal{A}_{\mathrm{P}}^{\epsilon}=\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon}$ *in the space* $\{f\in\mathcal{C}^{\alpha}(\Omega):\quad \int f(\mathbf{x})d\mathrm{P}(\mathbf{x})=0\}$.

(iii) the relation

$$\begin{pmatrix} (1-\mathcal{A}_{\mathrm{Q}}^{\epsilon}\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon})^{-1} & -(1-\mathcal{A}_{\mathrm{Q}}^{\epsilon}\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon})^{-1}\mathcal{A}_{\mathrm{Q}}^{\epsilon} \\ -\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon}(1-\mathcal{A}_{\mathrm{Q}}^{\epsilon}\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon})^{-1} & (1-\bar{\mathcal{A}}_{\mathrm{P}}^{\epsilon}\mathcal{A}_{\mathrm{Q}}^{\epsilon})^{-1}\end{pmatrix} = (D_{(f_{\mathrm{P},\mathrm{Q}},g_{\mathrm{P},\mathrm{Q}})}\Psi)^{-1}$$

holds in $\mathcal{C}^{\alpha}(\Omega)\times\mathcal{C}_0^{\alpha}(\Omega)$.

Proof. If we show that the operators $\mathcal{A}_{\mathrm{Q}}$ $\bar{\mathcal{A}}_{\mathrm{P}}$ are compact and that $(1-\bar{\mathcal{A}}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})$ and $(1-\mathcal{A}_{\mathrm{Q}}\bar{\mathcal{A}}_{\mathrm{P}})$ are injective, then Fredholm alternative (Theorem 6.6. in Brezis (2010)) and continuous inverse theorem (Corollary 2.7. in Brezis (2010)) would prove (i). (ii) follows by (5.17) below and the last claim follows by basic algebra.
To prove the compactness of $\mathcal{A}_{\mathrm{P}}$, let $\{f_k\}_{k\in\mathbb{N}}$ be a bounded sequence in $\mathcal{C}^{\alpha}(\Omega)$. Since the sequence

$$\mathcal{A}_{\mathrm{P}}f_k = \int \xi_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\cdot)f_k(\mathbf{x})d\mathrm{P}(\mathbf{x})\in\mathcal{C}^{\alpha+1}(\Omega),\quad k\in N$$

has its derivatives up to order $\alpha+1$ uniformly bounded by some $C(\Omega,s,d)$ (cf. Lemma 4.3 in del Barrio et al. (2022)), the Ascoli-Arzelà theorem yields the relatively compactness of $\{\mathcal{A}_{\mathrm{P}}f_k\}_{k\in\mathbb{N}}$ in $\mathcal{C}^{\alpha}(\Omega)$. The same argument applies to $\mathcal{A}_{\mathrm{Q}}$.

The injectivity holds by *reductio ad absurdum*. Suppose that $\mathcal{A}_{\mathrm{Q}}\bar{\mathcal{A}}_{\mathrm{P}}f=\mathcal{A}_{\mathrm{Q}}(\mathcal{A}_{\mathrm{P}}-\mathrm{Q}\mathcal{A}_{\mathrm{P}})f=-f$, for some $f\in\mathcal{C}^{s}(\Omega)$. In this case, since the optimality contidion implies

$$\mathrm{Q}\mathcal{A}_{\mathrm{P}}f=\int\int\xi_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y})f(\mathbf{x})d\mathrm{P}(\mathbf{x})d\mathrm{Q}(\mathbf{y})=\int f d\mathrm{P},\tag{5.17}$$

we have

$$\mathrm{P}\mathcal{A}_{\mathrm{Q}}\bar{\mathcal{A}}_{\mathrm{P}}f=\int\int\xi_{\mathrm{P,Q}}(\mathbf{x},\mathbf{y})((\mathcal{A}_{\mathrm{P}}-\mathrm{Q}\mathcal{A}_{\mathrm{P}})f)(\mathbf{y})d\mathrm{Q}(\mathbf{y})d\mathrm{P}(\mathbf{x})=\mathrm{Q}(\mathcal{A}_{\mathrm{P}}-\mathrm{Q}\mathcal{A}_{\mathrm{P}})f=0,$$

and by assumption $0=-\int f d\mathrm{P}$. Therefore we have

$$(\mathrm{Q}\mathcal{A}_{\mathrm{P}}f)^2=\left(\int\xi_{\mathrm{P,Q}}(\mathbf{x}',\mathbf{y})\int\xi_{\mathrm{P,Q}}(\mathbf{x},\mathbf{y})f(\mathbf{x})d\mathrm{P}(\mathbf{x})d\mathrm{Q}(\mathbf{y})\right)^2=f(\mathbf{x}')^2,\quad\text{for all }\mathbf{x}'\in\Omega$$

for all $\mathbf{x}'\in\Omega$, where Jensen's inequality gives

$$f(\mathbf{x}')^2\leq\int\xi_{\mathrm{P,Q}}(\mathbf{x}',\mathbf{y})\left(\int\xi_{\mathrm{P,Q}}(\mathbf{x},\mathbf{y})f(\mathbf{x})d\mathrm{P}(\mathbf{x})\right)^2d\mathrm{Q}(\mathbf{y})\leq\|f\|_\infty^2,\quad\text{for all }\mathbf{x}'\in\Omega,$$

and the first inequality is strict unless

$$\int\xi_{\mathrm{P,Q}}(\mathbf{x},\mathbf{y})(f(\mathbf{x}))\mathrm{P}(\mathbf{x})=c\in\mathbb{R}\text{ for Q-a.e. }\mathbf{y}.\tag{5.18}$$

Since there exists $\mathbf{x}'\in\Omega$ such that $|f(\mathbf{x}')|=\|f\|_\infty$, (5.18) holds. This would imply that

$$\mathbf{x}\mapsto\log\left(\frac{e^{f_{\mathrm{P,Q}}(\mathbf{x})}(f(\mathbf{x})+2\|f\|_\infty)}{c+2\|f\|_\infty}\right)$$

is a solution of the dual problem (5.2). Since it is unique, we obtain $\left(\frac{e^{f_{\mathrm{P,Q}}(\mathbf{x})}(f(\mathbf{x})+\|f\|_\infty)}{c+\|f\|_\infty}\right)=e^{f_{\mathrm{P,Q}}(\mathbf{x})}$ and $f=c$ in Ω. Since f is centered, we conclude $f=0$. $\square$

Now we prove Corollary 5.2.2. Since separable Hilbert spaces are of type 2 (cf. p.215 in Ledoux and Talagrand (1991)), and $\|e^{f_{\mathrm{P,Q}}+g_{\mathrm{P,Q}}-\frac{1}{2}\|\cdot-\cdot\|^2}\|^2_{W^{2\,\alpha,2}(\Omega)}\leq C$ (cf. Lemmas 4.1. and 4.5 in del Barrio et al. (2022)), Theorem 10.5. in Ledoux and Talagrand (1991) gives the limit

$$\sqrt{\frac{n\,m}{n+m}}\begin{pmatrix}\int e^{f_{\mathrm{P,Q}}+g_{\mathrm{P,Q}}(\mathbf{y})-\frac{1}{2}\|\cdot-\mathbf{y}\|^2}(d\mathrm{Q}(\mathbf{x})-d\mathrm{Q}_m(\mathbf{y}))\\ \int e^{f_{\mathrm{P,Q}}(\mathbf{x})+g_{\mathrm{P,Q}}-\frac{1}{2}\|\mathbf{x}-\cdot\|^2}(d\mathrm{P}(\mathbf{x})-d\mathrm{P}_n(\mathbf{x}))\end{pmatrix}\xrightarrow{w}-\begin{pmatrix}\sqrt{\lambda}\,\mathbb{G}_{\mathrm{Q},W}\\ \sqrt{1-\lambda}\,\mathbb{G}_{\mathrm{P},W}\end{pmatrix}\tag{5.19}$$

in $W^{2\,\alpha,2}(\Omega)\times W^{2\,\alpha,2}(\Omega)$, where, for $\delta_1,\delta_2\in W^{2\,\alpha,2}(\Omega)$:

$$\begin{aligned}\langle\mathbb{G}_{\mathrm{Q},W},\,\delta_1\rangle_{W^{2\,\alpha,2}(\Omega)}&\sim N\left(0,\mathrm{Var}_{\mathbf{Y}}\left(\langle\delta_1,\,e^{\frac{f_{\mathrm{P,Q}}+g_{\mathrm{P,Q}}(\mathbf{Y})-\frac{1}{2}\|\cdot-\mathbf{Y}\|^2}{\epsilon}}\rangle_{W^{2\,\alpha,2}(\Omega)}\right)\right),\\ \langle\mathbb{G}_{\mathrm{P},W},\,\delta_2\rangle_{W^{2\,\alpha,2}(\Omega)})&\sim N\left(0,\mathrm{Var}_{\mathbf{X}}\left(\langle\delta_2,\,e^{\frac{f_{\mathrm{P,Q}}(\mathbf{X})+g_{\mathrm{P,Q}}-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}{\epsilon}}\rangle_{W^{2\,\alpha,2}(\Omega)}\right)\right).\end{aligned}\tag{5.20}$$

We prove now that this limit holds also in $\mathcal{C}^{\alpha}(\Omega)\times\mathcal{C}_0^{\alpha}(\Omega)$. Since $2\alpha\geq d+1$, the Sobolev embedding theorem (Theorem 6.3. in Adams and Fournier (2003)) ensures the existence of a continuous linear mapping $\tau: W^{2\alpha,2}(\Omega)\to\mathcal{C}^{\alpha}(\Omega)$, such that $\tau(f)=f\text{-}\ell_d$ a.s. The continuous linear mapping theorem and the fact that the potentials belong to $\mathcal{C}^{\alpha}(\Omega)\times\mathcal{C}_0^{\alpha}(\Omega)$, we obtain

$$\sqrt{\frac{n\,m}{n+m}}\begin{pmatrix}\tau\int e^{f_{\mathrm{P,Q}}+g_{\mathrm{P,Q}}(\mathbf{y})-\frac{1}{2}\|\cdot-\mathbf{y}\|^2}(d\mathrm{Q}(\mathbf{x})-d\mathrm{Q}_m(\mathbf{y}))\\ \tau\int e^{f_{\mathrm{P,Q}}(\mathbf{x})+g_{\mathrm{P,Q}}-\frac{1}{2}\|\mathbf{x}-\cdot\|^2}(d\mathrm{P}(\mathbf{x})-d\mathrm{P}_n(\mathbf{x}))\end{pmatrix}$$
$$=\sqrt{\frac{n\,m}{n+m}}\begin{pmatrix}\int e^{f_{\mathrm{P,Q}}+g_{\mathrm{P,Q}}(\mathbf{y})-\frac{1}{2}\|\cdot-\mathbf{y}\|^2}(d\mathrm{Q}(\mathbf{x})-d\mathrm{Q}_m(\mathbf{y}))\\ \int e^{f_{\mathrm{P,Q}}(\mathbf{x})+g_{\mathrm{P,Q}}-\frac{1}{2}\|\mathbf{x}-\cdot\|^2}(d\mathrm{P}(\mathbf{x})-d\mathrm{P}_n(\mathbf{x}))\end{pmatrix}\xrightarrow{w}-\begin{pmatrix}\sqrt{\lambda}\,\tau\mathbb{G}_{\mathrm{Q},W}\\ \sqrt{1-\lambda}\,\tau\mathbb{G}_{\mathrm{P},W}\end{pmatrix}$$
$$=-\begin{pmatrix}\sqrt{\lambda}\,\mathbb{G}_{\mathrm{Q},s}\\ \sqrt{1-\lambda}\,\tau\mathbb{G}_{\mathrm{P},s}\end{pmatrix}$$

in $\mathcal{C}^{\alpha}(\Omega)\times\mathcal{C}^{\alpha}(\Omega)$. The dependence relation between $\mathbb{G}_{\mathrm{Q},s}$ and $\mathbb{G}_{\mathrm{P},s}$ is directly inherited from the samples. In this case they are independent. Finally, Lemma 5.6.5 applied to

$$(\Psi\begin{pmatrix}f_{\mathrm{P,Q}}\\ g_{\mathrm{P,Q}}\end{pmatrix}-\Psi_{n,m}\begin{pmatrix}f_{\mathrm{P,Q}}\\ g_{\mathrm{P,Q}}\end{pmatrix})$$
$$=\begin{pmatrix}\log\left(\int e^{g_{\mathrm{P,Q}}(\mathbf{y})-\frac{1}{2}\|\cdot-\mathbf{y}\|^2}d\mathrm{Q}(\mathbf{x})\right)-\log\left(\int e^{g_{\mathrm{P,Q}}(\mathbf{y})-\frac{1}{2}\|\cdot-\mathbf{y}\|^2}d\mathrm{Q}_m(\mathbf{y})\right)\\ \log\left(\int e^{f_{\mathrm{P,Q}}(\mathbf{x})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}d\mathrm{P}(\mathbf{x})\right)-\log\left(\int e^{f_{\mathrm{P,Q}}(\mathbf{x})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}d\mathrm{P}_n(\mathbf{x})\right)\end{pmatrix}$$

yields

$$(D_{f_{\mathrm{P,Q}},g_{\mathrm{P,Q}}}\Psi)\sqrt{\frac{n\,m}{n+m}}\begin{pmatrix}f_{\mathrm{P}_n,\mathrm{Q}_m}-f_{\mathrm{P,Q}}\\ g_{\mathrm{P}_n,\mathrm{Q}_m}-g_{\mathrm{P,Q}}\end{pmatrix}\longrightarrow-\begin{pmatrix}\sqrt{\lambda}\,\mathbb{G}_{\mathrm{Q},s}\\ \sqrt{1-\lambda}\,\mathbb{G}_{\mathrm{P},s}\end{pmatrix}.$$

Using the continuous mapping theorem, we obtain the relation

$$\begin{pmatrix}f_{\mathrm{P}_n,\mathrm{Q}_m}-f_{\mathrm{P,Q}}\\ g_{\mathrm{P}_n,\mathrm{Q}_m}-g_{\mathrm{P,Q}}\end{pmatrix}=\begin{pmatrix}(1-\mathcal{A}_{\mathrm{Q}}\bar{\mathcal{A}}_{\mathrm{P}})^{-1}\mathcal{A}_{\mathrm{Q}}\mathbb{G}^n_{\mathrm{P},s}-(1-\mathcal{A}_{\mathrm{Q}}\bar{\mathcal{A}}_{\mathrm{P}})^{-1}\mathbb{G}^m_{\mathrm{Q},s}\\ \eta\bar{\mathcal{A}}_{\mathrm{P}}(1-\mathcal{A}_{\mathrm{Q}}\bar{\mathcal{A}}_{\mathrm{P}})^{-1}\mathbb{G}^m_{\mathrm{Q},s}-(1-\bar{\mathcal{A}}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\mathbb{G}^n_{\mathrm{P},s}\end{pmatrix}+o_P\left(\sqrt{\frac{n+m}{n\,m}}\right)\tag{5.21}$$

Since

$$\int\mathbb{G}^n_{\mathrm{P},s}d\mathrm{Q}=\frac{1}{n}\sum_{k=1}^{n}\int\xi_{\mathrm{P,Q}}(\mathbf{X}_k,\mathbf{y})-\mathbb{E}\left(\xi_{\mathrm{P,Q}}(\mathbf{X},\mathbf{y})\right)d\mathrm{Q}(\mathbf{y})=0,$$

in view of Lemma 5.2.5 (ii), we can thus exchange $\bar{\mathcal{A}}_{\mathrm{P}}$ by $\mathcal{A}_{\mathrm{P}}$ in (5.21) and the proof of theorem and corollary is completed. □

5.3 Central limit theorem for the solution of the primal problem and Sinkhorn distances

This section covers the weak limit of the quantity

$$\sqrt{\frac{n\,m}{n+m}}\int \eta\,(d\pi^{\epsilon}_{\mathrm{P}_n,\mathrm{Q}_m}-d\pi^{\epsilon}_{\mathrm{P},\mathrm{Q}}),\quad \text{where } \eta\in L^2(\mathrm{P}\times\mathrm{Q}). \tag{5.22}$$

The organization of the section is as follows; first we recall certain results of Section 5.2 contained in the proof of Theorem 5.2.1, then we introduce notations to simplify the description of the limit that allow us to formulate the main result Theorem 5.3.1, which describes the first order linearization, and its main consequence Corollary 5.3.2. To conclude we observe that Corollary 5.3.2, at the same time, implies Corollary 5.3.3, which gives a description of the limit of the Sinkhorn distance introduced in Cuturi (2013) The section concludes with the proof of Theorem 5.3.1.

Before stating this result, for a fixed function $\eta\in L^2(\mathrm{P}\times\mathrm{Q})$ we introduce the notation;

$$\eta^{\epsilon}_{\mathbf{x}}:\mathbf{x}\to\int\eta(\mathbf{x},\mathbf{y})\xi^{\epsilon}_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y})d\mathrm{Q}(\mathbf{y}),\quad\text{and}\qquad \eta^{\epsilon}_{\mathbf{y}}:\mathbf{y}\to\int\eta(\mathbf{x},\mathbf{y})\xi^{\epsilon}_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y})d\mathrm{P}(\mathbf{x}),$$

which substantially simplifies the description of the first-order decomposition of (5.22), described in the following theorem.

Theorem 5.3.1. *Let $\Omega\subset\mathbb{R}^d$ be a compact set, $\mathrm{P},\mathrm{Q}\in\mathcal{P}(\Omega)$ and P_n (resp. Q_m) be the empirical measure of the i.i.d. sample $\mathbf{X}_1,\dots,\mathbf{X}_n$ (resp. $\mathbf{Y}_1,\dots,\mathbf{Y}_m$) distributed as P (resp. Q). Then, if $m=m(n)\to\infty$, $\frac{m}{n+m}\to\lambda\in(0,1)$ and $\eta\in L^2(\mathrm{P}\times\mathrm{Q})$,*

$$\begin{aligned}&\int\eta\,(d\pi^{\epsilon}_{\mathrm{P}_n,\mathrm{Q}_m}-d\pi^{\epsilon}_{\mathrm{P},\mathrm{Q}})\\ &=\frac{1}{n}\sum_{k=1}^{n}(1-\mathcal{A}^{\epsilon}_{\mathrm{Q}}\mathcal{A}^{\epsilon}_{\mathrm{P}})^{-1}\big(\eta^{\epsilon}_{\mathbf{x}}-\mathcal{A}^{\epsilon}_{\mathrm{Q}}\eta^{\epsilon}_{\mathbf{y}}\big)(\mathbf{X}_k)+\frac{1}{m}\sum_{j=1}^{m}(1-\mathcal{A}^{\epsilon}_{\mathrm{P}}\mathcal{A}^{\epsilon}_{\mathrm{Q}})^{-1}\big(\eta^{\epsilon}_{\mathbf{y}}-\mathcal{A}^{\epsilon}_{\mathrm{P}}\eta^{\epsilon}_{\mathbf{x}}\big)(\mathbf{Y}_j)\\ &\qquad+o_P\left(\sqrt{\frac{n+m}{n\,m}}\right).\end{aligned}$$

Theorem 5.3.1 gives the first-order decomposition of the solutions of the regularized optimal transport problem 5.1. We recall that, for a regularization based on the Schrödinger bridge, Harchaoui et al. (2020) arrived at exactly the same thing, conjecturing the truth of Theorem 5.3.1. The techniques of the proofs are completely different; ours is based on the theory of empirical processes, while that of Harchaoui et al. (2020) on a change of measurement and projections in $L^2(\mathrm{P}\times\mathrm{Q})$.

Since $\mathbb{E}\mathcal{A}_{\mathrm{Q}}^{\epsilon}\eta_{\mathbf{y}}^{\epsilon}(\mathbf{X}) = \mathbb{E}\eta_{\mathbf{y}}^{\epsilon}(\mathbf{X})$ and $\mathbb{E}\mathcal{A}_{\mathrm{P}}^{\epsilon}\eta_{\mathbf{x}}^{\epsilon}(\mathbf{Y}) = \mathbb{E}\eta_{\mathbf{x}}^{\epsilon}(\mathbf{Y})$, we have $\mathbb{E}\big(\eta_{\mathbf{x}}^{\epsilon} - \mathcal{A}_{\mathrm{Q}}^{\epsilon}\eta_{\mathbf{y}}^{\epsilon}\big)(\mathbf{X}) = \mathbb{E}\big(\eta_{\mathbf{y}}^{\epsilon} - \mathcal{A}_{\mathrm{P}}^{\epsilon}\eta_{\mathbf{x}}^{\epsilon}\big)(\mathbf{Y}) = 0$. Set $k \in \mathbb{N}$, Lemma 5.2.5 (ii) implies that

$$(\mathbb{E}\mathcal{A}_{\mathrm{Q}}^{\epsilon}\mathcal{A}_{\mathrm{P}}^{\epsilon})^k\big(\eta_{\mathbf{x}}^{\epsilon} - \mathcal{A}_{\mathrm{Q}}^{\epsilon}\eta_{\mathbf{y}}^{\epsilon}\big)(\mathbf{X}) = \mathbb{E}(\mathcal{A}_{\mathrm{P}}^{\epsilon}\mathcal{A}_{\mathrm{Q}}^{\epsilon})^k\big(\eta_{\mathbf{y}}^{\epsilon} - \mathcal{A}_{\mathrm{P}}^{\epsilon}\eta_{\mathbf{x}}^{\epsilon}\big)(\mathbf{Y}) = 0$$

and therefore

$$\mathbb{E}(1 - \mathcal{A}_{\mathrm{Q}}^{\epsilon}\mathcal{A}_{\mathrm{P}}^{\epsilon})^{-1}\big(\eta_{\mathbf{x}}^{\epsilon} - \mathcal{A}_{\mathrm{Q}}^{\epsilon}\eta_{\mathbf{y}}^{\epsilon}\big)(\mathbf{X}) + \mathbb{E}(1 - \mathcal{A}_{\mathrm{P}}^{\epsilon}\mathcal{A}_{\mathrm{Q}}^{\epsilon})^{-1}\big(\eta_{\mathbf{y}}^{\epsilon} - \mathcal{A}_{\mathrm{P}}^{\epsilon}\eta_{\mathbf{x}}^{\epsilon}\big)(\mathbf{Y}) = 0.$$

Then the first-order decomposition of Theorem 5.3.1 is centered and we obtain as an immediate consequence the limit of (5.22).

Corollary 5.3.2. *Let $\Omega \subset \mathbb{R}^d$ be a compact set, $\mathrm{P}, \mathrm{Q} \in \mathcal{P}(\Omega)$ and P_n (resp. Q_m) be the empirical measure of the i.i.d. sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ (resp. $\mathbf{Y}_1, \dots, \mathbf{Y}_m$) distributed as P (resp. Q). Then, if $m = m(n) \to \infty$, $\frac{m}{n+m} \to \lambda \in (0,1)$ and $\eta \in L^2(\mathrm{P} \times \mathrm{Q})$,*

$$\sqrt{\frac{n\,m}{n+m}}\left(\int \eta d\pi_{\mathrm{P}_n,\mathrm{Q}_m}^{\epsilon} - \int \eta d\pi_{\mathrm{P},\mathrm{Q}}^{\epsilon}\right) \longrightarrow N(0, \sigma_{\lambda,\epsilon}^2(\eta)), \ \ \textit{weakly},$$

where the variance $\sigma_{\lambda,\epsilon}^2(\eta)$ is

$$\lambda\,\mathrm{Var}_{\mathbf{X}\sim\mathrm{P}}\left((1 - \mathcal{A}_{\mathrm{Q}}^{\epsilon}\mathcal{A}_{\mathrm{P}}^{\epsilon})^{-1}\big(\eta_{\mathbf{x}}^{\epsilon} - \mathcal{A}_{\mathrm{Q}}^{\epsilon}\eta_{\mathbf{y}}^{\epsilon}\big)(\mathbf{X})\right) + (1-\lambda)\,\mathrm{Var}_{\mathbf{Y}\sim\mathrm{Q}}\left((1 - \mathcal{A}_{\mathrm{P}}^{\epsilon}\mathcal{A}_{\mathrm{Q}}^{\epsilon})^{-1}\big(\eta_{\mathbf{y}}^{\epsilon} - \mathcal{A}_{\mathrm{P}}^{\epsilon}\eta_{\mathbf{x}}^{\epsilon}\big)(\mathbf{Y})\right).$$

Moreover, in the one-sample case we have

$$\sqrt{n}\left(\int \eta d\pi_{\mathrm{P}_n,\mathrm{Q}}^{\epsilon} - \int \eta d\pi_{\mathrm{P}_n,\mathrm{Q}}^{\epsilon}\right) \longrightarrow N(0, \sigma_{\mathrm{P}}^2(\eta)), \ \ \textit{weakly},$$

with

$$\sigma_{\mathrm{P}}^2(\eta) = \mathrm{Var}_{\mathbf{X}\sim\mathrm{P}}\left((1 - \mathcal{A}_{\mathrm{Q}}^{\epsilon}\mathcal{A}_{\mathrm{P}}^{\epsilon})^{-1}\big(\eta_{\mathbf{x}}^{\epsilon} - \mathcal{A}_{\mathrm{Q}}^{\epsilon}\eta_{\mathbf{y}}^{\epsilon}\big)(\mathbf{X})\right).$$

Note that, in the proof of Theorem 5.3.1 two terms are random variables: the difference between densities—which depends on the potentials—and the difference between the empirical processes. Each term is treated separately, giving rise to the linearization described in Theorem 5.3.1.

An immediate corollary of Theorem 5.3.1 is its application to the square norm function, where we obtain the weak limit of the Sinkhorn cost introduced in Cuturi (2013). Formally it is defined as

$$d_S^{\epsilon}(\mathrm{P}, \mathrm{Q}) = \mathbb{E}_{(\mathbf{X},\mathbf{Y})\sim\pi_{\mathrm{P},\mathrm{Q}}^{\epsilon}}\left(\frac{\|\mathbf{X} - \mathbf{Y}\|^2}{2}\right),$$

and represents the cost of *'transporting mass'* from P to Q when using the coupling given by the entropic regularization.

Corollary 5.3.3. *Let $\Omega \subset \mathbb{R}^d$ be a compact set, $\mathrm{P}, \mathrm{Q} \in \mathcal{P}(\Omega)$ and P_n (resp. Q_m) be the empirical measure of the i.i.d. sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ (resp. $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$) distributed as P (resp. Q). Then, if $m = m(n) \to \infty$, $\frac{m}{n+m} \to \lambda \in (0,1)$,*

$$\sqrt{\frac{n\,m}{n+m}}\,(d_S^\epsilon(\mathrm{P}_n, \mathrm{Q}_m) - d_S^\epsilon(\mathrm{P}, \mathrm{Q})) \longrightarrow N\left(0, \sigma^2_{\lambda,\epsilon}\left(\frac{\|\cdot - \cdot\|^2}{2}\right)\right), \quad weakly,$$

where the variance $\sigma^2_{\lambda,\epsilon}(\frac{\|\cdot-\cdot\|^2}{2})$ is defined in Corollary 5.3.3 for the function $(\mathbf{x}, \mathbf{y}) \to \frac{\|\mathbf{x}-\mathbf{y}\|^2}{2}$. Moreover, in the one-sample case we have

$$\sqrt{n}\,(d_S^\epsilon(\mathrm{P}_n, \mathrm{Q}) - d_S^\epsilon(\mathrm{P}, \mathrm{Q})) \longrightarrow N\left(0, \sigma^2_{\mathrm{P},\epsilon}\left(\frac{\|\cdot - \cdot\|^2}{2}\right)\right), \quad weakly,$$

with $\sigma^2_{\mathrm{P},\epsilon}\left(\frac{\|\cdot-\cdot\|^2}{2}\right)$ as in Corollary 5.3.3.

Another interesting application of Corollary 5.3.2 is to the function $(\mathbf{x}, \mathbf{y}) \mapsto \mathbb{1}_{\|\mathbf{x}-\mathbf{y}\|^2 \leq t}$, for $t \geq 0$. Klatt et al. (2020) computed the regularized optimal trasnport problem to match two protein intensity distributions and defined the regularized colocalization measure RCol

$$\mathrm{RCol}(\pi^\epsilon_{\mathrm{P},\mathrm{Q}}, t) = \pi^\epsilon_{\mathrm{P},\mathrm{Q}}(\|\mathbf{x} - \mathbf{y}\|^2 \leq t),$$

which represents the mass of the pixel intensity transported on scales smaller or equal to t in the regularized optimal transport matching the two intensity distributions. Theorem 7.1. in Klatt et al. (2020) gives confidence intervals for the discretized images (finite number of pixels). The following result extend it to general probability distributions representing the pixels.

Corollary 5.3.4. *Let $\Omega \subset \mathbb{R}^d$ be a compact set, $\mathrm{P}, \mathrm{Q} \in \mathcal{P}(\Omega)$ and P_n (resp. Q_m) be the empirical measure of the i.i.d. sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ (resp. $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$) distributed as P (resp. Q). Then, if $m = m(n) \to \infty$, $\frac{m}{n+m} \to \lambda \in (0,1)$,*

$$\sqrt{\frac{n\,m}{n+m}}\,\left(\mathrm{RCol}(\pi^\epsilon_{\mathrm{P}_n,\mathrm{Q}_m}, t) - \mathrm{RCol}(\pi^\epsilon_{\mathrm{P},\mathrm{Q}}, t)\right) \longrightarrow N\left(0, \sigma^2_{\lambda,\epsilon}\left(\mathbb{1}_{\|\cdot-\cdot\|^2 \leq t}\right)\right), \quad weakly,$$

where the variance $\sigma^2_{\lambda,\epsilon}(\mathbb{1}_{\|\cdot-\cdot\|^2\leq t})$ is defined in Corollary 5.3.3 for the function $(\mathbf{x}, \mathbf{y}) \mapsto \mathbb{1}_{\|\mathbf{x}-\mathbf{y}\|^2 \leq t}$. Moreover, in the one-sample case we have

$$\sqrt{n}\,\left(\mathrm{RCol}(\pi^\epsilon_{\mathrm{P}_n,\mathrm{Q}}, t) - \mathrm{RCol}(\pi^\epsilon_{\mathrm{P},\mathrm{Q}}, t)\right) \longrightarrow N\left(0, \sigma^2_{\mathrm{P},\epsilon}\left(\frac{\|\cdot - \cdot\|^2}{2}\right)\right), \quad weakly,$$

with $\sigma^2_{\mathrm{P},\epsilon}\left(\mathbb{1}_{\|\cdot-\cdot\|^2\leq t}\right)$ as in Corollary 5.3.3.

Proof of Theorem 5.3.1. The following relation–which is consequence of Fubini's theorem–will be useful in the proof:

$$\mathbb{E}\left(f(\mathbf{X})(\mathcal{A}^\epsilon_{\mathrm{Q}} g)(\mathbf{X})\right) = \mathbb{E}\left(g(\mathbf{Y})(\mathcal{A}^\epsilon_{\mathrm{P}} f)(\mathbf{Y})\right), \quad \text{for all } f \in L^2(\mathrm{P}), g \in L^2(\mathrm{Q}). \tag{5.23}$$

As has already been done before, we assume that $\epsilon = 1$ and avoid any super/sub-index related with ϵ. The first step is to split (5.22) in three terms: one capturing the difference between the densities

$$A_{n,m} = \sqrt{\frac{n\,m}{n+m}} \int \eta(\mathbf{x},\mathbf{y})\, e^{-\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2} \left(e^{f_{\mathrm{P}_n,\mathrm{Q}}(\mathbf{x})+g_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{x})} - e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{x})+g_{\mathrm{P},\mathrm{Q}}(\mathbf{y})} \right) d\mathrm{Q}_m(\mathbf{y}) d\mathrm{P}_n(\mathbf{x}),$$

one dealing with the empirical process of P

$$B_{n,m} = \sqrt{\frac{n\,m}{n+m}} \int \eta(\mathbf{x},\mathbf{y})\, \xi_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y}) d\mathrm{Q}_m(\mathbf{y})(d\mathrm{P}_n(\mathbf{x}) - d\mathrm{P}(\mathbf{x})).$$

and the last one with that of Q

$$C_m = \sqrt{\frac{n\,m}{n+m}} \int \eta(\mathbf{x},\mathbf{y})\, \xi_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y}) d\mathrm{P}(\mathbf{x})(d\mathrm{Q}_m(\mathbf{y}) - d\mathrm{Q}(\mathbf{y})) = \sqrt{\frac{n\,m}{n+m}} (\mathrm{Q}_m - \mathrm{Q})(\eta_{\mathbf{y}}).$$

Clearly, the third term is already linearized and does not need to be manipulated at this time. For the second one we realise that

$$\int \eta(\mathbf{x},\mathbf{y})\, \xi_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y})(d\mathrm{Q}_m(\mathbf{y}) - d\mathrm{Q}(\mathbf{y}))(d\mathrm{P}_n(\mathbf{x}) - d\mathrm{P}(\mathbf{x})) = O_P(\frac{1}{\sqrt{n\,m}}),$$

which implies that $B_{n,m} = B_n + o_P(1)$, with

$$B_n = \sqrt{\frac{n\,m}{n+m}} \int \eta(\mathbf{x},\mathbf{y})\, \xi_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y}) d\mathrm{Q}(\mathbf{y})(d\mathrm{P}_n(\mathbf{x}) - d\mathrm{P}(\mathbf{x})).$$

The first one, however, requires an additional effort. We split it again in two different terms $A_n = A^1_{n,m} + A^2_{n,m}$, with

$$A^1_{n,m} = \sqrt{\frac{n\,m}{n+m}} \int \eta(\mathbf{x},\mathbf{y})\ (\xi_{\mathrm{P}_n,\mathrm{Q}_m} - \xi_{\mathrm{P},\mathrm{Q}})\, d\mathrm{Q}_m(\mathbf{y})(d\mathrm{P}_n(\mathbf{x}) - d\mathrm{P}(\mathbf{x}))$$

and

$$A^2_{n,m} = \sqrt{\frac{n\,m}{n+m}} \int \eta(\mathbf{x},\mathbf{y})\ (\xi_{\mathrm{P}_n,\mathrm{Q}_m} - \xi_{\mathrm{P},\mathrm{Q}})\, d\mathrm{Q}_m(\mathbf{y}) d\mathrm{P}(\mathbf{x}).$$

We show first that $A^1_{n,m} \longrightarrow 0$ in probability. The operator

$$h \mapsto \Xi_{n,m} h = \int \eta(\mathbf{x},\mathbf{y})\, e^{-\frac{1}{\epsilon}\|\mathbf{x}-\mathbf{y}\|^2} h(\mathbf{x},\mathbf{y}) d\mathrm{Q}_m(\mathbf{y})(d\mathrm{P}_n(\mathbf{x}) - d\mathrm{P}(\mathbf{x}))$$

belongs to $(\mathcal{C}^s(\Omega \times \Omega))'$ and, moreover, $\|\Xi_{n,m}\|_{(\mathcal{C}^s(\Omega\times\Omega))'} \to 0$ a.s. This is because for any $h_1, h_2 \in \mathcal{C}^s(\Omega \times \Omega)$, we have

$$\left| \eta(\mathbf{x},\mathbf{y})\, e^{-\frac{1}{\epsilon}\|\mathbf{x}-\mathbf{y}\|^2} (h_1(\mathbf{x},\mathbf{y}) - h_2(\mathbf{x},\mathbf{y})) \right| \le |\eta(\mathbf{x},\mathbf{y})\, e^{-\frac{1}{\epsilon}\|\mathbf{x}-\mathbf{y}\|^2}| \|h_1 - h_2\|_{\mathcal{C}^s(\Omega\times\Omega)},$$

thus the covering numbers for the infinity norm of the class

$$\mathcal{F}_\eta = \{(\mathbf{x},\mathbf{y}) \mapsto \eta(\mathbf{x},\mathbf{y})\, e^{-\frac{1}{\epsilon}\|\mathbf{x}-\mathbf{y}\|^2} h(\mathbf{x},\mathbf{y}),\ \|h\|_{\mathcal{C}^s(\Omega\times\Omega)} \le 1\}$$

can be bounded by the ones of the unit ball of $\mathcal{C}^s(\Omega\times\Omega)$ (Theorem 2.7.11 in Vaart and Wellner (1996)) which is Glivenko-Cantelli (Theorem 2.7.1 in Vaart and Wellner (1996)). Lemma 5.6.1 gives

$$\|e^{h_{\mathrm{P}_n,\mathrm{Q}_m}} - e^{h_{\mathrm{P},\mathrm{Q}}}\|^2_{\mathcal{C}^s(\Omega\times\Omega)} = O_P\left(\sqrt{\frac{n+m}{n\,m}}\right),$$

$$A^1_{n,m} \le \sqrt{\frac{n\,m}{n+m}} \sup_{f\in\mathcal{F}} \int \eta(\mathbf{x},\mathbf{y})\ (\xi_{\mathrm{P}_n,\mathrm{Q}_m} - \xi_{\mathrm{P},\mathrm{Q}})\, d\mathrm{Q}_m(\mathbf{y})(d\mathrm{P}_n(\mathbf{x}) - d\mathrm{P}(\mathbf{x})),$$

which proves the claim.

It only remains to derive the weak limit of $A^2_{n,m}$. First note that, by repeating the previous arguments, we obtain the convergence in probability towards 0 of

$$\sqrt{\frac{n\,m}{n+m}} \int \eta(\mathbf{x},\mathbf{y})\ (\xi_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{x},\mathbf{y}) - \xi_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y}))\,(d\mathrm{Q}_m(\mathbf{y}) - d\mathrm{Q}(\mathbf{y}))d\mathrm{P}(\mathbf{x}),$$

which allows us to write $A^2_{n,m}$, up to additive $o_P(1)$ terms, as

$$\sqrt{\frac{n\,m}{n+m}} \int \eta(\mathbf{x},\mathbf{y})\ (\xi_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{x},\mathbf{y}) - \xi_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y}))\, d\mathrm{Q}(\mathbf{y})d\mathrm{P}(\mathbf{x}).$$

We apply Taylor's theorem to the exponential, giving rise to

$$\begin{aligned}
e^{f_{\mathrm{P}_n,\mathrm{Q}_m}+g_{\mathrm{P}_n,\mathrm{Q}_m}} - e^{f_{\mathrm{P},\mathrm{Q}}+g_{\mathrm{P},\mathrm{Q}}}
&= e^{f_{\mathrm{P},\mathrm{Q}}+g_{\mathrm{P},\mathrm{Q}}}(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}+g_{\mathrm{P}_n,\mathrm{Q}_m}-f_{\mathrm{P},\mathrm{Q}}-g_{\mathrm{P},\mathrm{Q}}} - 1)\\
&= e^{f_{\mathrm{P},\mathrm{Q}}+g_{\mathrm{P},\mathrm{Q}}}(f_{\mathrm{P}_n,\mathrm{Q}_m} + g_{\mathrm{P}_n,\mathrm{Q}_m} - f_{\mathrm{P},\mathrm{Q}} - g_{\mathrm{P},\mathrm{Q}})\\
&\qquad + O_P(\|f_{\mathrm{P}_n,\mathrm{Q}_m} + g_{\mathrm{P}_n,\mathrm{Q}_m} - f_{\mathrm{P},\mathrm{Q}} - g_{\mathrm{P},\mathrm{Q}}\|^2_\infty).
\end{aligned}$$

Then, applying Lemma 5.6.3, we obtain

$$e^{f_{\mathrm{P}_n,\mathrm{Q}_m}+g_{\mathrm{P}_n,\mathrm{Q}_m}} - e^{f_{\mathrm{P},\mathrm{Q}}+g_{\mathrm{P},\mathrm{Q}}} = e^{f_{\mathrm{P},\mathrm{Q}}+g_{\mathrm{P},\mathrm{Q}}}(f_{\mathrm{P}_n,\mathrm{Q}_m} + g_{\mathrm{P}_n,\mathrm{Q}_m} - f_{\mathrm{P},\mathrm{Q}} - g_{\mathrm{P},\mathrm{Q}}) + O_P(\frac{1}{n}).$$

As a consequence, we have

$$A^2_{n,m} = \sqrt{\frac{n\,m}{n+m}} \int \eta\ (f_{\mathrm{P}_n,\mathrm{Q}_m} + g_{\mathrm{P}_n,\mathrm{Q}_m} - f_{\mathrm{P},\mathrm{Q}} - g_{\mathrm{P},\mathrm{Q}})\, d\pi_{\mathrm{P},\mathrm{Q}} + o_P(1),$$

and, in view of Theorem 5.2.1, also

$$\int \eta\left(f_{\mathrm{P}_n,\mathrm{Q}_m} + g_{\mathrm{P}_n,\mathrm{Q}_m} - f_{\mathrm{P},\mathrm{Q}} - g_{\mathrm{P},\mathrm{Q}}\right) d\pi_{\mathrm{P},\mathrm{Q}} =$$
$$\int \eta\left(-(1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathbb{G}^m_{\mathrm{Q},s} + (1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathcal{A}_{\mathrm{Q}}\mathbb{G}^n_{\mathrm{P},s}\right) d\pi_{\mathrm{P},\mathrm{Q}}$$
$$+ \int \eta\left(\mathcal{A}_{\mathrm{P}}(1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathbb{G}^m_{\mathrm{Q},s} - (1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\mathbb{G}^n_{\mathrm{P},s}\right) d\pi_{\mathrm{P},\mathrm{Q}} \tag{5.24}$$

with $\mathbb{G}^n_{\mathrm{P},s}$ and $\mathbb{G}^m_{\mathrm{Q},s}$ as in Theorem 5.2.1.

In order to introduce the term η inside the operator, we compute

$$\begin{aligned}
&\int \eta\left((1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathbb{G}^m_{\mathrm{Q},s}\right) d\pi_{\mathrm{P},\mathrm{Q}}\\
&= \int \left(\int \eta(\mathbf{x},\mathbf{y}) e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{x})+g_{\mathrm{P},\mathrm{Q}}(\mathbf{y})-\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2} d\mathrm{Q}(\mathbf{y})\right) ((1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathbb{G}^m_{\mathrm{Q},s})(\mathbf{x}) d\mathrm{P}(\mathbf{x})\\
&= \int \eta_{\mathbf{x}}(\mathbf{x})((1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathbb{G}^m_{\mathrm{Q},s})(\mathbf{x}) d\mathrm{P}(\mathbf{x})\\
&= \int \mathbb{G}^m_{\mathrm{Q},s}(\mathbf{x})((1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\eta_{\mathbf{x}})(\mathbf{x}) d\mathrm{P}(\mathbf{x}),
\end{aligned}$$

where the last step is consequence of (5.23), so $(1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}$ is self-adjoin. Each term of (5.24) can be treated in the same way, which gives the relations;

$$\begin{aligned}
\int \eta\mathcal{A}_{\mathrm{P}}(1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathbb{G}^m_{\mathrm{Q},s} d\pi_{\mathrm{P},\mathrm{Q}} &= \int \mathbb{G}^m_{\mathrm{Q},s}(\mathbf{x})((1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathcal{A}_{\mathrm{Q}}\eta_{\mathbf{y}})(\mathbf{x}) d\mathrm{P}(\mathbf{x}),\\
\int \eta(1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathbb{G}^n_{\mathrm{P},s} d\pi_{\mathrm{P},\mathrm{Q}} &= \int \mathbb{G}^n_{\mathrm{P},s}(\mathbf{y})((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\eta_{\mathbf{y}})(\mathbf{y}) d\mathrm{Q}(\mathbf{y}),\\
\int \eta(1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathcal{A}_{\mathrm{Q}}\mathbb{G}^n_{\mathrm{P},s} d\pi_{\mathrm{P},\mathrm{Q}} &= \int \mathbb{G}^n_{\mathrm{P},s}(\mathbf{y})((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\mathcal{A}_{\mathrm{P}}\eta_{\mathbf{x}})(\mathbf{x}) d\mathrm{Q}(\mathbf{x}).
\end{aligned}$$

Therefore we obtain

$$\begin{aligned}
A^2_{n,m} = \sqrt{\frac{n\,m}{n+m}}\frac{1}{m}\Bigg(&\sum_{k=1}^m \int \left(((1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\big(-\eta_{\mathbf{x}}+\mathcal{A}_{\mathrm{Q}}\eta_{\mathbf{y}}\big))(\mathbf{x})\right)\xi_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{Y}_k) d\mathrm{P}(\mathbf{x})\\
&- \int \left(((1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\big(-\eta_{\mathbf{x}}+\mathcal{A}_{\mathrm{Q}}\eta_{\mathbf{y}}\big))(\mathbf{x})\right) d\pi_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y})\Bigg)\\
\sqrt{\frac{n\,m}{n+m}}\frac{1}{n}\Bigg(&\sum_{j=1}^n \int \left(((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\big(-\eta_{\mathbf{y}}+\mathcal{A}_{\mathrm{P}}\eta_{\mathbf{x}}\big))(\mathbf{y})\right)\xi_{\mathrm{P},\mathrm{Q}}(\mathbf{X}_j,\mathbf{y}) d\mathrm{Q}(\mathbf{y})\\
&- \int \left(((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\big(-\eta_{\mathbf{y}}+\mathcal{A}_{\mathrm{P}}\eta_{\mathbf{x}}\big))(\mathbf{y})\right) d\pi_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y})\Bigg) + o_P(1),
\end{aligned}$$

Now we recover those forgotten B_n and C_m to write (5.22) with respect to the empirical processes;

$$\sqrt{\frac{n\,m}{n+m}}\Big(\int \eta d\pi^{\epsilon}_{\mathrm{P}_n,\mathrm{Q}_m} - \int \eta d\pi^{\epsilon}_{\mathrm{P},\mathrm{Q}}\Big) = \sqrt{\frac{n\,m}{n+m}} U_n^{\mathrm{P}} + \sqrt{\frac{n\,m}{n+m}} U_m^{\mathrm{Q}} + o_p(1)$$

where

$$\begin{aligned} &U_m^{\mathrm{Q}} = \\ &\frac{1}{m}\Big(\sum_{k=1}^{m}\int \big(((1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\big(-\eta_{\mathbf{x}}+\mathcal{A}_{\mathrm{Q}}\eta_{\mathbf{y}}\big))(\mathbf{x})+\eta(\mathbf{x},\mathbf{Y}_k)\big)\xi_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{Y}_k)d\mathrm{P}(\mathbf{x})\Big) \\ &\qquad - \int \big(((1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\big(-\eta_{\mathbf{x}}+\mathcal{A}_{\mathrm{Q}}\eta_{\mathbf{y}}\big))(\mathbf{x})+\eta(\mathbf{x},\mathbf{y})\big)d\pi_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y}) \end{aligned}$$

and

$$\begin{aligned} &U_n^{\mathrm{P}} = \\ &\frac{1}{n}\Big(\sum_{k=1}^{n}\int \big(((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\big(-\eta_{\mathbf{y}}+\mathcal{A}_{\mathrm{P}}\eta_{\mathbf{x}}\big))(\mathbf{y})+\eta(\mathbf{X}_j,\mathbf{y})\big)\xi_{\mathrm{P},\mathrm{Q}}(\mathbf{X}_j,\mathbf{y})d\mathrm{Q}(\mathbf{y})\Big) \\ &\qquad - \int \big(((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\big(-\eta_{\mathbf{y}}+\mathcal{A}_{\mathrm{Q}}\eta_{\mathbf{x}}\big))(\mathbf{y})+\eta(\mathbf{x},\mathbf{y})\big)d\pi_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y}). \end{aligned}$$

We now simplify the term

$$\begin{aligned} &\frac{1}{m}\Big(\sum_{k=1}^{m}\int \big(((1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\big(-\eta_{\mathbf{x}}+\mathcal{A}_{\mathrm{Q}}\eta_{\mathbf{y}}\big))(\mathbf{x})+\eta(\mathbf{x},\mathbf{Y}_k)\big)\xi_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{Y}_k)d\mathrm{P}(\mathbf{x})\Big) \\ &\qquad = \frac{1}{m}\sum_{k=1}^{m}\mathcal{A}_{\mathrm{P}}((1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\big(-\eta_{\mathbf{x}}+\mathcal{A}_{\mathrm{Q}}\eta_{\mathbf{y}}\big))(\mathbf{Y}_k)+\eta_{\mathbf{y}}(\mathbf{Y}_k) \end{aligned}$$

by realising that

$$\begin{aligned} \mathcal{A}_{\mathrm{P}} = (1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})\mathcal{A}_{\mathrm{P}} &= (1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}(\mathcal{A}_{\mathrm{P}}-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}}) \\ &= (1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\mathcal{A}_{\mathrm{P}}(1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}}) \end{aligned}$$

implies $\mathcal{A}_{\mathrm{P}}(1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1} = (1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\mathcal{A}_{\mathrm{P}}$. This gives

$$\begin{aligned}
&\frac{1}{m}\sum_{k=1}^{m}\mathcal{A}_{\mathrm{P}}((1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\big(-\eta_{\mathbf{x}}+\mathcal{A}_{\mathrm{Q}}\eta_{\mathbf{y}}\big))(\mathbf{Y}_k)+\eta_{\mathbf{y}}(\mathbf{Y}_k)\\
&=\frac{1}{m}\sum_{k=1}^{m}\big((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\big(-\mathcal{A}_{\mathrm{P}}\eta_{\mathbf{x}}+\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}}\eta_{\mathbf{y}}-\eta_{\mathbf{y}}+\eta_{\mathbf{y}}\big)+\eta_{\mathbf{y}}\big)(\mathbf{Y}_k)\\
&=\frac{1}{m}\sum_{k=1}^{m}\big((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\big(-\mathcal{A}_{\mathrm{P}}\eta_{\mathbf{x}}-(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})\eta_{\mathbf{y}}+\eta_{\mathbf{y}}\big)+\eta_{\mathbf{y}}\big)(\mathbf{Y}_k)\\
&=\frac{1}{m}\sum_{k=1}^{m}(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\big(\eta_{\mathbf{y}}-\mathcal{A}_{\mathrm{P}}\eta_{\mathbf{x}}\big)(\mathbf{Y}_k)
\end{aligned}$$

and, following the same arguments, also

$$\begin{aligned}
&\int\big(((1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\big(-\eta_{\mathbf{x}}+\mathcal{A}_{\mathrm{Q}}\eta_{\mathbf{y}}\big))(\mathbf{x})+\eta(\mathbf{x},\mathbf{y})\big)d\pi_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y})\\
&\qquad\qquad=\int(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\big(\eta_{\mathbf{y}}-\mathcal{A}_{\mathrm{P}}\eta_{\mathbf{x}}\big)d\mathrm{Q}. \qquad (5.25)
\end{aligned}$$

Since $\mathrm{Q}\mathcal{A}_{\mathrm{P}} = \mathrm{P}$, we have $\mathrm{Q}(\eta_{\mathbf{x}}) = \mathrm{P}(\eta_{\mathbf{y}}) = \pi_{\mathrm{P},\mathrm{Q}}(\eta)$ and therefore

$$U_m^{\mathrm{Q}} = \frac{1}{m}\sum_{k=1}^{m}(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\big(\eta_{\mathbf{y}}-\mathcal{A}_{\mathrm{P}}\eta_{\mathbf{x}}\big)(\mathbf{Y}_k).$$

The same argument yields

$$U_n^{\mathrm{P}} = \frac{1}{n}\sum_{k=1}^{n}(1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\big(\eta_{\mathbf{x}}-\mathcal{A}_{\mathrm{Q}}\eta_{\mathbf{y}}\big)(\mathbf{X}_k),$$

which finishes the proof. □

5.4 Weak limit of the Divergences

Recall that, for probabilities $\mathrm{P},\mathrm{Q} \in \mathcal{P}(\Omega)$, its quadratic Sinkhorn's divergence (Genevay et al. (2018)) is defined as

$$D_\epsilon(\mathrm{P},\mathrm{Q}) = S_\epsilon(\mathrm{P},\mathrm{Q}) - \frac{1}{2}\left(S_\epsilon(\mathrm{P},\mathrm{P}) + S_\epsilon(\mathrm{Q},\mathrm{Q})\right).$$

The main result of this section, Theorem 5.4.3, gives the limits of the quantity $a_n\left(D_\epsilon(\mathrm{P}_n,\mathrm{Q}) - D_\epsilon(\mathrm{P}_n,\mathrm{Q})\right)$, where the sequence $\{a_n\}_{n\in\mathbb{N}}$ depends on the hypothesis $H_0:\ \mathrm{P}=\mathrm{Q}$ or $H_1:\ \mathrm{P}=\mathrm{Q}$. In this last case, the limit can be established by means of the del Barrio and Loubes (2019)'s technique based on Efron-Stein inequality—see also Mena and Niles-Weed (2019); del

Barrio et al. (2021a) for reproduction and improvement of this argument—with common rate $a_n = \sqrt{n}$. The case $\mathrm{P} = \mathrm{Q}$, however has the faster rate $a_n = n$ and the limit depends on the P-Brownian bridge in $(\mathcal{C}^s(\Omega))'$, call it $\mathbb{G}_\mathrm{P}$ which is (p.82 in Vaart and Wellner (1996)) the centered Gaussian process with covariance function

$$(f, g) \mapsto \int f(\mathbf{x})g(\mathbf{x})d\mathrm{P}(\mathbf{x}) - \int f(\mathbf{x})\mathrm{P}(\mathbf{x}) \int g(\mathbf{x})\mathrm{P}(\mathbf{x}).$$

More precisely, the limit in the one sample case will be the action of the operator $\mathbb{G}_\mathrm{P}$ to the function $((1 - \mathcal{A}^\epsilon_\mathrm{P}\mathcal{A}^\epsilon_\mathrm{P})^{-1}(1 + 2\mathcal{A}^\epsilon_\mathrm{P}))\xi_{\mathrm{P},\mathrm{P}}\mathbb{G}_\mathrm{P}$, where $\xi_{\mathrm{P},\mathrm{P}}\mathbb{G}_\mathrm{P}$ is the function mapping $\mathbf{x} \to \mathbb{G}_\mathrm{P}\xi_{\mathrm{P},\mathrm{P}}(\cdot, \mathbf{x})$.

Theorem 5.4.1. *Let $\Omega \subset \mathbb{R}^d$ be a compact set, $\mathrm{P}, \mathrm{Q} \in \mathcal{P}(\Omega)$ and P_n (resp. Q_m) be the empirical measure of the i.i.d. sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ (resp. $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$) distributed as P (resp. Q). Then, if $m = m(n) \to \infty$ and $\frac{m}{n+m} \to \lambda \in (0, 1)$, we have the following limits.*

- *Under H_0 : $\mathrm{P} = \mathrm{Q}$,*

$$\frac{n\,m}{n+m} D_1(\mathrm{P}_n, \mathrm{P}'_m) \xrightarrow{w} \frac{1}{4}(\sqrt{\lambda}\mathbb{G}_\mathrm{P} - \sqrt{1-\lambda}\mathbb{G}_{\mathrm{P}'}) \left(((1 - \mathcal{A}^\epsilon_\mathrm{P}\mathcal{A}^\epsilon_\mathrm{P})^{-1}(1 + 2\mathcal{A}^\epsilon_\mathrm{P}))(\sqrt{\lambda}\xi_{\mathrm{P},\mathrm{P}}\mathbb{G}_\mathrm{P} - \sqrt{1-\lambda}\xi_{\mathrm{P},\mathrm{P}}\mathbb{G}_{\mathrm{P}'}) \right),$$

  and

$$n\, D_1(\mathrm{P}_n, \mathrm{P}) \xrightarrow{w} \frac{1}{4}\mathbb{G}_\mathrm{P} \left(((1 - \mathcal{A}^\epsilon_\mathrm{P}\mathcal{A}^\epsilon_\mathrm{P})^{-1}(1 + 2\mathcal{A}^\epsilon_\mathrm{P}))(\xi_{\mathrm{P},\mathrm{P}}\mathbb{G}_\mathrm{P}) \right),$$

- *Under H_1 : $\mathrm{P} \neq \mathrm{Q}$,*

$$\sqrt{\frac{n\,m}{n+m}}(D_\epsilon(\mathrm{P}_n, \mathrm{Q}_m) - D_\epsilon(\mathrm{P}, \mathrm{Q})) \xrightarrow{w} N(0, \lambda\,\mathrm{Var}_\mathrm{P}(\psi^\epsilon_{\mathrm{P},\mathrm{Q}}) + (1-\lambda)\,\mathrm{Var}_\mathrm{Q}(\psi^\epsilon_{\mathrm{Q},\mathrm{P}})),$$

  and

$$\sqrt{n}(D_\epsilon(\mathrm{P}_n, \mathrm{Q}) - D_\epsilon(\mathrm{P}, \mathrm{Q})) \xrightarrow{w} N(0, \mathrm{Var}_\mathrm{P}(\psi^\epsilon_{\mathrm{P},\mathrm{Q}})),$$

  where $\psi^\epsilon_{\mathrm{P},\mathrm{Q}} = f^\epsilon_{\mathrm{P},\mathrm{Q}} - \frac{1}{2}(f^\epsilon_{\mathrm{P},\mathrm{P}} + g^\epsilon_{\mathrm{P},\mathrm{P}})$ and $\psi^\epsilon_{\mathrm{Q},\mathrm{P}} = f^\epsilon_{\mathrm{Q},\mathrm{P}} - \frac{1}{2}(f^\epsilon_{\mathrm{Q},\mathrm{Q}} + g^\epsilon_{\mathrm{Q},\mathrm{Q}})$.

The limit of Theorem 5.4.1 is difficult to express in terms of known or common random variables. One way to proceed is by an inner product in the Sobolev space $W^{2s,2}(\Omega)$. Since Ω is compact there exists a ball $R\mathbb{B}$ centered in $\mathbf{0}$ with radius R so that the probabilities P, Q are supported on it. Therefore, for the following augmentations we can assume without loosing generality that $\Omega = R\mathbb{B}$. In this regular domain, the Sobolev embedding theorem (Theorem 6.3. in Adams and Fournier (2003)) states the existence of a continuous linear mapping $\tau : W^{2s,2}(\Omega) \to \mathcal{C}^s(\Omega)$, such that $\tau(f) = f\text{-}\ell_d$ a.s. We can define the adjoint operator $\tau^* : \mathrm{Dom}(\tau^*) \to (W^{2s,2}(\Omega))' = W^{2s,2}(\Omega)$, determined by the property $v(\tau f) = \langle \tau^* v, f \rangle_W$, for all $f \in W^{2s,2}(\Omega)$ and $v \in \mathrm{Dom}(\tau^*)$. Since τ is bounded, $\mathrm{Dom}(\tau^*) = (\mathcal{C}^s(\Omega))'$ (chapter 2, Remark 16. in Brezis (2010)), both the empirical process

$\sqrt{n}(\mathrm{P}_n - \mathrm{P})$ and P-Brownian bridge belong to $\mathrm{Dom}(\tau^*)$.

Since $\tau^*\mathbb{G}_{\mathrm{P}}$ and $\tau^*\mathbb{G}_{\mathrm{P}'}$ are Gaussian—for all $f \in W^{2s,2}(\Omega)$, we have $\langle \tau^*\mathbb{G}_{\mathrm{P}}, f\rangle_W = \mathbb{G}_{\mathrm{P}}(\tau f)$—independent and equally distributed, we can write (cf. Proposition 3.6. in Ledoux and Talagrand (1991))

$$\tau^*\mathbb{G}_{\mathrm{P}} = \sum_{i=1}^{\infty} N_i f_i, \quad \tau^*\mathbb{G}_{\mathrm{P}'} = \sum_{i=1}^{\infty} N_i' f_i \tag{5.26}$$

where $\{N_i\}_{i\in\mathbb{N}}$ and $\{N_i'\}_{i\in\mathbb{N}}$ are mutually i.i.d. sequences with $N_i, N_i' \sim N(0,1)$ and $\{f_i\}_{i\in\mathbb{N}} \subset W^{2s,2}(\Omega)$. The convergence in (5.26) is a.s. in $W^{2s,2}(\Omega)$. The following result is consequence of the fact that the operator

$$\Theta = \mathcal{D}\tau^*(1 + 2\mathcal{A}_{\mathrm{P}}^*)(1 - \mathcal{A}_{\mathrm{P}}^*\mathcal{A}_{\mathrm{P}}^*)^{-1},$$

where

$$W^{2s,2}(\Omega) \ni f \mapsto \mathcal{D}f = \sum_{|\alpha|\le 2s} \int D_\alpha f(\mathbf{y}) D_\alpha^{\mathbf{y}} \xi_{\mathrm{P},\mathrm{P}}(\cdot, \mathbf{y}) d\ell_d(\mathbf{y}) \in W^{2s,2}(\Omega),$$

is self-adjoint.

Lemma 5.4.2. *The operator Θ is self-adjoint and definite positive. As a consequence there exist sequences $\{x_{i,j}\}_{i,j\in\mathbb{N}} \subset \mathbb{R}$ and $\{e_j\}_{j\in\mathbb{N}} \subset W^{2s,2}(\Omega)$ such that $f_i = \sum_{j=1}^{n} x_{i,j} e_j$, for each $i \in \mathbb{N}$ and $\{e_j\}_{j\in\mathbb{N}}$ is an orthonormal basis of $W^{2s,2}(\Omega)$ such that $\Theta e_j = \lambda_j e_j$, were $\lambda_j \ge 0$, for all $j \in \mathbb{N}$.*

With this notation we can rewrite the limits under the null hypothesis of Theorem 5.4.1 in terms of independent Gaussian random variables.

Theorem 5.4.3. *Let $\Omega \subset \mathbb{R}^d$ be a compact set, $\mathrm{P} \in \mathcal{P}(\Omega)$, P_n and P_m' be independent empirical measures of P. Then, if $m = m(n) \to \infty$ and $\frac{m}{n+m} \to \lambda \in (0,1)$, we have the following limits:*

$$n\, D_1(\mathrm{P}_n, \mathrm{P}) \xrightarrow{w} \frac{1}{4}\sum_{j=1}^{\infty} \lambda_j \left(\sum_{i=1}^{\infty} x_{i,j} N_i\right)\left(\sum_{i=1}^{\infty} x_{i,j} N_i\right)$$

and

$$\frac{n\,m}{n+m}(D_\epsilon(\mathrm{P}_n, \mathrm{P}_m') - D_\epsilon(\mathrm{P}, \mathrm{Q})) \xrightarrow{w} \frac{1}{4}\sum_{j=1}^{\infty} \lambda_j \left(\sum_{i=1}^{\infty} x_{i,j}(\sqrt{\lambda} N_i + \sqrt{1-\lambda} N_i')\right)\left(\sum_{i=1}^{\infty} x_{i,j}(\sqrt{\lambda} N_i + \sqrt{1-\lambda} N_i')\right),$$

where $\{N_i\}_{i\in\mathbb{N}}$ and $\{N_i'\}_{i\in\mathbb{N}}$ are mutually i.i.d. with $N_i, N_i' \sim N(0,1)$ and the sequences $\{x_{i,j}\}_{i,j\in\mathbb{N}} \subset \mathbb{R}$ and $\{\lambda_j\}_{j\in\mathbb{N}} \subset [0,\infty)$ as in Lemma 5.4.2.

Proof of Theorem 5.4.1 if $\mathrm{P}=\mathrm{Q}$. As usually, we prove it for $\epsilon=1$ and the two-sample case. We denote $\mathrm{P}'_m=\mathrm{Q}_m=\frac{1}{m}\sum_{k=1}^m \delta_{\mathbf{X}'_k}$ and we want to derive the limit of

$$\begin{aligned} D_1(\mathrm{P}_n,\mathrm{P}'_m) &= S_1(\mathrm{P}_n,\mathrm{P}'_m)-\tfrac{1}{2}\left(S_1(\mathrm{P}_n,\mathrm{P}_n)+S_1(\mathrm{P}'_m,\mathrm{P}'_m)\right)\\ &= \tfrac{1}{2}\left(S_1(\mathrm{P}_n,\mathrm{P}'_m)-S_1(\mathrm{P}_n,\mathrm{P}_n)\right)+\tfrac{1}{2}\left(S_1(\mathrm{P}_n,\mathrm{P}'_m)-S_1(\mathrm{P}'_m,\mathrm{P}'_m)\right). \end{aligned}$$

First note that, since

$$\int e^{h_{\mathrm{P}'_m,\mathrm{P}'_m}(\mathbf{x},\mathbf{y})-\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2}d\mathrm{P}_n d\mathrm{P}'_m=1,$$

Lemma 5.6.2 yields

$$\begin{aligned} |S_1(\mathrm{P}_n,\mathrm{P}'_m)-\int h_{\mathrm{P}'_m,\mathrm{P}'_m}d\mathrm{P}_n d\mathrm{P}'_m-\int \frac{1}{2}(h_{\mathrm{P}_n,\mathrm{P}'_m}-h_{\mathrm{P}'_m,\mathrm{P}'_m})^2 d\pi_{\mathrm{P}_n,\mathrm{P}'_m}|&\\ \leq C\|h_{\mathrm{P}_n,\mathrm{P}'_m}-h_{\mathrm{P}'_m,\mathrm{P}'_m}\|_s^3,& \end{aligned}$$

and therefore

$$\begin{aligned} |S_1(\mathrm{P}_n,\mathrm{P}'_m)-S_1(\mathrm{P}'_m,\mathrm{P}'_m)-\int h_{\mathrm{P}'_m,\mathrm{P}'_m}(d\mathrm{P}_n d\mathrm{P}'_m-d\mathrm{P}'_m d\mathrm{P}'_m)&\\ -\int \frac{1}{2}(h_{\mathrm{P}_n,\mathrm{P}'_m}-h_{\mathrm{P}'_m,\mathrm{P}'_m})^2 d\pi_{\mathrm{P}_n,\mathrm{P}'_m}|&\\ \leq C\|h_{\mathrm{P}_n,\mathrm{P}'_m}-h_{\mathrm{P}'_m,\mathrm{P}'_m}\|_s^3.& \end{aligned}$$

By the same argument we also have

$$\begin{aligned} |S_1(\mathrm{P}_n,\mathrm{P}'_m)-S_1(\mathrm{P}_n,\mathrm{P}_n)-\int h_{\mathrm{P}_n,\mathrm{P}_n}(d\mathrm{P}_n d\mathrm{P}'_m-d\mathrm{P}_n d\mathrm{P}_n)&\\ -\int \frac{1}{2}(h_{\mathrm{P}_n,\mathrm{P}'_m}-h_{\mathrm{P}_n,\mathrm{P}_n})^2 d\pi_{\mathrm{P}_n,\mathrm{P}'_m}|&\\ \leq C\|h_{\mathrm{P}_n,\mathrm{P}'_m}-h_{\mathrm{P}'_m,\mathrm{P}'_m}\|_s^3& \end{aligned}$$

and, as a consequence,

$$\begin{aligned} |D_1(\mathrm{P}_n,\mathrm{P}'_m)-\frac{1}{2}(\int h_{\mathrm{P}_n,\mathrm{P}_n}(d\mathrm{P}_n d\mathrm{P}'_m-d\mathrm{P}_n d\mathrm{P}_n)-\int h_{\mathrm{P}'_m,\mathrm{P}'_m}(d\mathrm{P}_n d\mathrm{P}'_m-d\mathrm{P}'_m d\mathrm{P}'_m))&\\ -\frac{1}{4}\int((h_{\mathrm{P}_n,\mathrm{P}'_m}-h_{\mathrm{P}'_m,\mathrm{P}'_m})^2+(h_{\mathrm{P}_n,\mathrm{P}'_m}-h_{\mathrm{P}_n,\mathrm{P}_n})^2)d\pi_{\mathrm{P}_n,\mathrm{P}'_m}|&\\ \leq C\|h_{\mathrm{P}_n,\mathrm{P}'_m}-h_{\mathrm{P}'_m,\mathrm{P}'_m}\|_s^3.& \end{aligned}$$

The relations

$$\int h_{\mathrm{P}_n,\mathrm{P}_n}(d\mathrm{P}_n d\mathrm{P}'_m-d\mathrm{P}'_m d\mathrm{P}'_m)=\int f_{\mathrm{P}_n,\mathrm{P}_n}(d\mathrm{P}_n-d\mathrm{P}'_m),$$

$$\int h_{\mathrm{P}'_m,\mathrm{P}'_m}(d\mathrm{P}'_m d\mathrm{P}_n - d\mathrm{P}'_m d\mathrm{P}_n) = \int g_{\mathrm{P}'_m,\mathrm{P}'_m}(d\mathrm{P}_n - d\mathrm{P}'_m)$$

and the fact that, in this symmetric case $f_{\mathrm{P}_n,\mathrm{P}_n} = g_{\mathrm{P}_n,\mathrm{P}_n} + c$ give

$$\begin{aligned}|D_1(\mathrm{P}_n, \mathrm{P}'_m) - \frac{1}{2}\int (g_{\mathrm{P}_n,\mathrm{P}_n} - g_{\mathrm{P}'_m,\mathrm{P}'_m})(d\mathrm{P}'_m - d\mathrm{P}'_n) \\ - \frac{1}{4}\int ((h_{\mathrm{P}_n,\mathrm{P}'_m} - h_{\mathrm{P}'_m,\mathrm{P}'_m})^2 + (h_{\mathrm{P}_n,\mathrm{P}'_m} - h_{\mathrm{P}_n,\mathrm{P}_n})^2) d\pi_{\mathrm{P}_n,\mathrm{P}'_m}| \\ \leq C\|h_{\mathrm{P}_n,\mathrm{P}'_m} - h_{\mathrm{P}'_m,\mathrm{P}'_m}\|_s^3.\end{aligned}$$

By Theorem 4.5 in del Barrio et al. (2022), we have

$$\begin{aligned}|D_1(\mathrm{P}_n, \mathrm{P}'_m) - \frac{1}{2}\int (g_{\mathrm{P}_n,\mathrm{P}_n} - g_{\mathrm{P}'_m,\mathrm{P}'_m})(d\mathrm{P}'_m - d\mathrm{P}'_n) \\ - \frac{1}{4}\int ((h_{\mathrm{P}_n,\mathrm{P}'_m} - h_{\mathrm{P}'_m,\mathrm{P}'_m})^2 + (h_{\mathrm{P}_n,\mathrm{P}'_m} - h_{\mathrm{P}_n,\mathrm{P}_n})^2) d\pi_{\mathrm{P}_n,\mathrm{P}'_m}| = o_P\left(\frac{n+m}{n\,m}\right).\end{aligned} \tag{5.27}$$

The inequality

$$\begin{aligned}|\int (h_{\mathrm{P}_n,\mathrm{P}'_m} - h_{\mathrm{P}'_m,\mathrm{P}'_m})^2 (d\pi_{\mathrm{P}_n,\mathrm{P}'_m} - d\pi_{\mathrm{P}'_m,\mathrm{P}'_m})| \\ \leq \|(h_{\mathrm{P}_n,\mathrm{P}'_m} - h_{\mathrm{P}'_m,\mathrm{P}'_m})^2\|_s \|\pi_{\mathrm{P}_n,\mathrm{P}'_m} - \pi_{\mathrm{P}'_m,\mathrm{P}'_m}\|_{\mathcal{C}^s(\Omega\times\Omega)'}.\end{aligned}$$

the relation

$$\|\left(h_{\mathrm{P}_n,\mathrm{P}'_m} - h_{\mathrm{P}'_m,\mathrm{P}'_m}\right)^2\|_s = O_P\left(\frac{n+m}{n\,m}\right),$$

which is consequence of of Lemma 5.6.6 and Theorem 4.5 in del Barrio et al. (2022), and the fact that $\|\pi_{\mathrm{P}_n,\mathrm{P}'_m} - \pi_{\mathrm{P}'_m,\mathrm{P}'_m}\|_{\mathcal{C}(\Omega\times\Omega)'} \to 0$ a.s. implies that we can exchange $d\pi_{\mathrm{P}_n,\mathrm{P}'_m}$ by $d\pi_{\mathrm{P},\mathrm{P}}$ in (5.27), obtaining that the limit of $\frac{n\,m}{n+m} D_1(\mathrm{P}_n, \mathrm{P}'_m)$ is the same of

$$\begin{aligned}\frac{n\,m}{n+m}\bigg(\frac{1}{2}\int (g_{\mathrm{P}_n,\mathrm{P}_n} - g_{\mathrm{P}'_m,\mathrm{P}'_m})(d\mathrm{P}'_m - d\mathrm{P}'_n) \\ + \frac{1}{4}\int ((h_{\mathrm{P}_n,\mathrm{P}'_m} - h_{\mathrm{P}'_m,\mathrm{P}'_m})^2 + (h_{\mathrm{P}_n,\mathrm{P}'_m} - h_{\mathrm{P}_n,\mathrm{P}_n})^2) d\pi_{\mathrm{P},\mathrm{P}}\bigg).\end{aligned} \tag{5.28}$$

Since, up $o_P\left(\sqrt{\frac{n+m}{n\,m}}\right)$ terms, we have

$$\begin{aligned}h_{\mathrm{P}_n,\mathrm{P}'_m}(\mathbf{x},\mathbf{y}) - h_{\mathrm{P},\mathrm{P}}(\mathbf{x},\mathbf{y}) = \\ \left(-(1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathbb{G}^m_{\mathrm{P}',s} + (1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathcal{A}_{\mathrm{Q}}\mathbb{G}^n_{\mathrm{P},s}\right)\bigg)(\mathbf{x}) \\ + \left(\mathcal{A}_{\mathrm{P}}(1-\mathcal{A}_{\mathrm{Q}}\mathcal{A}_{\mathrm{P}})^{-1}\mathbb{G}^m_{\mathrm{P}',s} - (1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{Q}})^{-1}\mathbb{G}^n_{\mathrm{P},s}\right)(\mathbf{y})\end{aligned} \tag{5.29}$$

and

$$h_{\mathrm{P}'_m,\mathrm{P}'_m}(\mathbf{x},\mathbf{y}) - h_{\mathrm{P},\mathrm{P}}(\mathbf{x},\mathbf{y}) = \Big(-(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}\mathbb{G}^m_{\mathrm{P}',s} + (1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}\mathcal{A}_{\mathrm{P}}\mathbb{G}^m_{\mathrm{P}',s}\Big)(\mathbf{x}) + \left(\mathcal{A}_{\mathrm{P}}(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}\mathbb{G}^m_{\mathrm{P}',s} - (1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}\mathbb{G}^m_{\mathrm{P}',s}\right)(\mathbf{y}),$$

then

$$h_{\mathrm{P}_n,\mathrm{P}'_m}(\mathbf{x},\mathbf{y}) - h_{\mathrm{P}'_m,\mathrm{P}'_m}(\mathbf{x},\mathbf{y}) = \Big((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}\mathcal{A}_{\mathrm{P}}(\mathbb{G}^n_{\mathrm{P},s} - \mathbb{G}^m_{\mathrm{P}',s})\Big)(\mathbf{x}) - \left((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(\mathbb{G}^n_{\mathrm{P},s} - \mathbb{G}^m_{\mathrm{P}',s})\right)(\mathbf{y}),$$

where we claim that taking squares we obtain

$$\begin{aligned}(h_{\mathrm{P}_n,\mathrm{P}'_m}(\mathbf{x},\mathbf{y}) - h_{\mathrm{P}'_m,\mathrm{P}'_m}(\mathbf{x},\mathbf{y}))^2 = {}& \Big((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}\mathcal{A}_{\mathrm{P}}(\mathbb{G}^n_{\mathrm{P},s} - \mathbb{G}^m_{\mathrm{P}',s})\Big)^2(\mathbf{x}) + \left((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(\mathbb{G}^n_{\mathrm{P},s} - \mathbb{G}^m_{\mathrm{P}',s})\right)^2(\mathbf{y}) \\ & - 2\Big((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}\mathcal{A}_{\mathrm{P}}(\mathbb{G}^n_{\mathrm{P},s} - \mathbb{G}^m_{\mathrm{P}',s})\Big)(\mathbf{x})\left((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(\mathbb{G}^n_{\mathrm{P},s} - \mathbb{G}^m_{\mathrm{P}',s})\right)(\mathbf{y}) \\ & + o_P\left(\frac{n+m}{n\,m}\right). \end{aligned} \tag{5.30}$$

We justify now the claim, denote as $\delta^L_{n,m}$ and $\delta^R_{n,m}$ the left and right hand sides of (5.29). We can upper bound

$$|(\delta^L_{n,m})^2 - (\delta^R_{n,m})^2| \le |\delta^L_{n,m} - \delta^R_{n,m}||\delta^L_{n,m} + \delta^R_{n,m}.|$$

Theorem 5.2.1 implies that $|\delta^L_{n,m} - \delta^R_{n,m}| = o_P(\sqrt{\frac{n+m}{n\,m}})$, while Theorem 4.5 in del Barrio et al. (2022) that $|\delta^L_{n,m}| = O_P(\sqrt{\frac{n+m}{n\,m}})$, then proving that $|\delta^R_{n,m}| = O_P(\sqrt{\frac{n+m}{n\,m}})$ we prove also the claim. But this is immediate due to $|\delta^R_{n,m}| \le |\delta^L_{n,m} - \delta^R_{n,m}| + |\delta^L_{n,m}|$ and the previous bounds.

Integrating (5.30) w.r.t. $\pi_{\mathrm{P},\mathrm{P}}$ yields

$$\begin{aligned}&\|(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}\mathcal{A}_{\mathrm{P}}(\mathbb{G}^n_{\mathrm{P},s} - \mathbb{G}^m_{\mathrm{P}',s})\|^2_{L(\mathrm{P})} + \|(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(\mathbb{G}^n_{\mathrm{P},s} - \mathbb{G}^m_{\mathrm{P}',s})\|^2_{L(\mathrm{P})} \\ &- 2\langle(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}\mathcal{A}_{\mathrm{P}}(\mathbb{G}^n_{\mathrm{P},s} - \mathbb{G}^m_{\mathrm{P}',s}), \mathcal{A}_{\mathrm{P}}(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(\mathbb{G}^n_{\mathrm{P},s} - \mathbb{G}^m_{\mathrm{P}',s})\rangle_{L(\mathrm{P})}.\end{aligned} \tag{5.31}$$

Since every one of the processes are centered $\mathcal{A}_{\mathrm{P}} = \mathcal{A}_{\mathrm{P}}$, $\mathcal{A}^*_{\mathrm{P}} = \mathcal{A}_{\mathrm{P}}$ in $L^2(\mathrm{P})$ and all of the operators commute, we have (5.31) equals to

$$\begin{aligned}\langle(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s}),&((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-2}+\mathcal{A}_{\mathrm{P}}^2(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-2}-2\mathcal{A}_{\mathrm{P}}^2(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-2})(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s})\rangle_{L(\mathrm{P})}\\ &=\langle(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s}),((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-2}-\mathcal{A}_{\mathrm{P}}^2(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-2})(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s})\rangle_{L(\mathrm{P})}\\ &=\langle(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s}),(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s})\rangle_{L(\mathrm{P})}\\ &=(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s})(\mathcal{A}_{\mathrm{P}}(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}\xi_{\mathrm{P},\mathrm{P}}(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s}))\end{aligned} \tag{5.32}$$

By the same means, we can prove (up to additive $o_P\left(\frac{nm}{n+m}\right)$ terms) the equality

$$\int(h_{\mathrm{P}_n,\mathrm{P}'_m}-h_{\mathrm{P}_n,\mathrm{P}_n})^2\pi_{\mathrm{P},\mathrm{P}}=(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s})(\mathcal{A}_{\mathrm{P}}(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}\xi_{\mathrm{P},\mathrm{P}}(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s})). \tag{5.33}$$

Now we focus on the limit of $\frac{1}{2}\frac{nm}{n+m}\int\left(g_{\mathrm{P}'_m,\mathrm{P}'_m}-g_{\mathrm{P}_n,\mathrm{P}_n}\right)(d\mathrm{P}_n-d\mathrm{P}'_m)$. On the one hand, in view of Lemma 5.2.5, $(g_{\mathrm{P}_n,\mathrm{P}_n}-g_{\mathrm{P}'_m,\mathrm{P}'_m})$ can be expressed, up to additive $o_P\left(\sqrt{\frac{n+m}{nm}}\right)$ terms, in $\mathcal{C}^s(\Omega)$ as

$$\begin{aligned}\big((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s})&+(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}\mathcal{A}_{\mathrm{P}}(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s})\big)\\ &=((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}+(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}\mathcal{A}_{\mathrm{P}})(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s})\\ &=((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+\mathcal{A}_{\mathrm{P}}))(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s}).\end{aligned}$$

then

$$\begin{aligned}&\int\left(g_{\mathrm{P}'_m,\mathrm{P}'_m}-g_{\mathrm{P}_n,\mathrm{P}_n}\right)(d\mathrm{P}_n-d\mathrm{P}'_m)\\ &=(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s})(((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+\mathcal{A}_{\mathrm{P}}))\xi_{\mathrm{P},\mathrm{P}}(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s}))+o_P\left(\frac{n+m}{nm}\right).\end{aligned} \tag{5.34}$$

The relations (5.28), (5.32) , (5.33) and (5.34) give

$$\begin{aligned}&D_1(\mathrm{P}_n,\mathrm{P}'_m)\\ &\quad=\frac{1}{2}(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s})(((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+\mathcal{A}_{\mathrm{P}}))\xi_{\mathrm{P},\mathrm{P}}(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s}))\\ &\quad+\frac{1}{2}(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s})(\mathcal{A}_{\mathrm{P}}(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}\xi_{\mathrm{P},\mathrm{P}}(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s}))+o_P\left(\frac{n+m}{nm}\right).\end{aligned}$$

which, using that $(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}$ and $\mathcal{A}_{\mathrm{P}}$ commute, yields that the weak limit of $\frac{nm}{n+m}D_1(\mathrm{P}_n,\mathrm{P}'_m)$ is the same of

$$\frac{1}{2}\frac{nm}{n+m}(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s})(((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+2\mathcal{A}_{\mathrm{P}}))\xi_{\mathrm{P},\mathrm{P}}(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s})).$$

The limit

$$\sqrt{\frac{n\,m}{n+m}}\xi_{\mathrm{P},\mathrm{P}}(\mathbb{G}^n_{\mathrm{P},s}-\mathbb{G}^m_{\mathrm{P}',s})\to\xi_{\mathrm{P},\mathrm{P}}(\mathbb{G}_{\mathrm{P},s}-\mathbb{G}_{\mathrm{P}',s}) \tag{5.35}$$

holds in $\mathcal{C}^\alpha(\Omega)\times\mathcal{C}^\alpha(\Omega)$, for any $\alpha\in\mathbb{N}$ (see (5.19)) and, on the other hand, for $\alpha>s$, the unit ball in $\mathcal{C}^\alpha(\Omega)$ is Donsker (cf. Theorems 2.5.2 and 2.7.1 in Vaart and Wellner (1996)), so $\sqrt{\frac{n\,m}{n+m}}(\mathrm{P}_n-\mathrm{P}_m)\to\sqrt{\lambda}\mathbb{G}_{\mathrm{P}}-\sqrt{1-\lambda}\mathbb{G}_{\mathrm{P}'}$, weakly in $(\mathcal{C}^\alpha(\Omega))'$, for any $\alpha\geq s$.

To conclude we realise that, if E is a Banach space, $E\ni f_n\to f\in E$ and $E'\ni F_n\to F\in E'$, then

$$|F_n(f_n)-F(f)|\leq|F_n(f_n)-F(f_n)|+|F(f_n)-F(f_n)|\leq\|F_n-F\|_{E'}\|f_n\|_E+\|f_n-f\|_E\|F\|_{E'}. \tag{5.36}$$

Since $\sup_{n\in\mathbb{N}}\|f_n\|_E\in\mathbb{R}$, both of the terms of (5.36) tends to 0. As a consequence we obtain the weak limit

$$\frac{n\,m}{n+m}D_1(\mathrm{P}_n,\mathrm{P}'_m)\xrightarrow{w}$$
$$\frac{1}{4}(\sqrt{\lambda}\mathbb{G}_{\mathrm{P}}-\sqrt{1-\lambda}\mathbb{G}_{\mathrm{P}'})\left(((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+2\mathcal{A}_{\mathrm{P}}))(\sqrt{\lambda}\xi_{\mathrm{P},\mathrm{P}}\mathbb{G}_{\mathrm{P}}-\sqrt{1-\lambda}\xi_{\mathrm{P},\mathrm{P}}\mathbb{G}_{\mathrm{P}'})\right),$$

which, finishes the proof of the two-sample case. The one-sample case follows by the same means. □

Proof of Theorem 5.4.1 if $\mathrm{P}\neq\mathrm{Q}$,. Here we prove the one-sample case. Let $\mathbf{X}'_1\sim\mathrm{P}$ be a r.v. independent of $\mathbf{X}_1,\dots,\mathbf{X}_n$ and denote $\mathrm{P}'_n=\frac{1}{n}\delta_{\mathbf{X}'_1}+\frac{1}{n}\sum_{i=2}^n\delta_{\mathbf{X}_i}$. Note that

$$D_1(\mathrm{P}_n,Q)-D_1(\mathrm{P}'_n,Q)=S_\epsilon(\mathrm{P}_n,Q)-S_\epsilon(\mathrm{P}'_n,Q)-\tfrac{1}{2}\left(S_\epsilon(\mathrm{P}_n,\mathrm{P}_n)-S_\epsilon(\mathrm{P}'_n,\mathrm{P}'_n)\right)$$

together with

$$f_{\mathrm{P}_n,Q}(\mathbf{X}_1)-f_{\mathrm{P}_n,\mathrm{P}}(\mathbf{X}'_1)\leq n(S_\epsilon(\mathrm{P}_n,Q)-S_\epsilon(\mathrm{P}'_n,Q))\leq f_{\mathrm{P}'_n,Q}(\mathbf{X}_1)-f_{\mathrm{P}'_n,Q}(\mathbf{X}'_1)$$

and

$$n(S_\epsilon(\mathrm{P}_n,\mathrm{P}_n)-S_\epsilon(\mathrm{P}'_n,\mathrm{P}'_n))\geq g_{\mathrm{P}_n,\mathrm{P}_n}(\mathbf{X}_1)+f_{\mathrm{P}_n,\mathrm{P}_n}(\mathbf{X}_1)-g_{\mathrm{P}_n,\mathrm{P}_n}(\mathbf{X}'_1)-f_{\mathrm{P}_n,\mathrm{P}_n}(\mathbf{X}'_1),$$
$$n(S_\epsilon(\mathrm{P}_n,\mathrm{P}_n)-S_\epsilon(\mathrm{P}'_n,\mathrm{P}'_n))\leq g_{\mathrm{P}'_n,\mathrm{P}'_n}(\mathbf{X}_1)+f_{\mathrm{P}'_n,\mathrm{P}'_n}(\mathbf{X}_1)-g_{\mathrm{P}'_n,\mathrm{P}'_n}(\mathbf{X}'_1)-f_{\mathrm{P}'_n,\mathrm{P}'_n}(\mathbf{X}'_1),$$

implies the bound

$$\begin{aligned}&n(D_1(\mathrm{P}_n,Q)-D_1(\mathrm{P}'_n,Q))\\&\leq f_{\mathrm{P}'_n,Q}(\mathbf{X}_1)-f_{\mathrm{P}'_n,Q}(\mathbf{X}'_1)-\tfrac{1}{2}\left(g_{\mathrm{P}_n,\mathrm{P}_n}(\mathbf{X}_1)+f_{\mathrm{P}_n,\mathrm{P}_n}(\mathbf{X}_1)-g_{\mathrm{P}_n,\mathrm{P}_n}(\mathbf{X}'_1)-f_{\mathrm{P}_n,\mathrm{P}_n}(\mathbf{X}'_1)\right).\end{aligned} \tag{5.37}$$

We observe that

$$\begin{aligned}&\tfrac{1}{n}\psi_{\mathrm{P},Q}(\mathbf{X}'_1)+\frac{1}{n}\sum_{i=2}^n\psi_{\mathrm{P},Q}(\mathbf{X}_i)-\frac{1}{n}\sum_{i=1}^n\psi_{\mathrm{P},Q}(\mathbf{X}_i)\\&=\frac{1}{n}(f_{\mathrm{P},Q}(\mathbf{X}'_1)-\tfrac{1}{2}(f_{\mathrm{P},\mathrm{P}}(\mathbf{X}'_1)-g_{\mathrm{P},\mathrm{P}}(\mathbf{X}'_1)-f_{\mathrm{P},Q}(\mathbf{X}_1)+\tfrac{1}{2}(f_{\mathrm{P},\mathrm{P}}(\mathbf{X}_1)+g_{\mathrm{P},\mathrm{P}}(\mathbf{X}_1))).\end{aligned}$$

Therefore, the Efron-Stein inequality yields

$$\operatorname{Var}\left(D_1(\mathrm{P}_n,Q)-\frac{1}{n}\sum_{i=2}^{n}\psi_{\mathrm{P},Q}(\mathbf{X}_i)\right) \\ \leq n\,\mathbb{E}\left(D_1(\mathrm{P}_n,Q)-\psi_{\mathrm{P},Q}(\mathbf{X}_1)-D_1(\mathrm{P}'_n,Q)+\psi_{\mathrm{P},Q}(\mathbf{X}'_1)\right)_+^2$$

and, from (5.37), we obtain

$$n\ \operatorname{Var}\left(D_1(\mathrm{P}_n,Q)-\frac{1}{n}\sum_{i=2}^{n}\psi_{\mathrm{P},Q}(\mathbf{X}_i)\right) \\ \leq \mathbb{E}(f_{\mathrm{P}'_n,Q}(\mathbf{X}_1)-f_{\mathrm{P}'_n,Q}(\mathbf{X}'_1)-\tfrac{1}{2}\left(g_{\mathrm{P}_n,\mathrm{P}_n}(\mathbf{X}_1)+f_{\mathrm{P}_n,\mathrm{P}_n}(\mathbf{X}_1)-g_{\mathrm{P}_n,\mathrm{P}_n}(\mathbf{X}'_1)-f_{\mathrm{P}_n,\mathrm{P}_n}(\mathbf{X}'_1)\right) \\ -\psi_{\mathrm{P},Q}(\mathbf{X}_1)+\psi_{\mathrm{P},Q}(\mathbf{X}'_1))^2. \tag{5.38}$$

The convergence of the the right hand side of (5.38) towards 0 is consequence of the convergence of the potentials, see the last part of the proof of Lemma 3.2 in del Barrio et al. (2022). As a consequence the following central limit theorem holds.

Lemma 5.4.4. *Let* $\mathrm{P},Q\in\mathrm{P}(\mathbb{R}^d)$ *be* σ*-subgaussian probabilities, then*

$$\sqrt{n}(D_\epsilon(\mathrm{P}_n,Q)-\mathbb{E}D_\epsilon(\mathrm{P}_n,Q))\xrightarrow{w}N(0,\operatorname{Var}_{\mathrm{P}}(\psi^\epsilon_{\mathrm{P},Q})),$$

where $\psi^\epsilon_{\mathrm{P},Q}=f^\epsilon_{\mathrm{P},Q}-\frac{1}{2}(f^\epsilon_{\mathrm{P},\mathrm{P}}-g^\epsilon_{\mathrm{P},\mathrm{P}})$.

The fact that the central limit theorem of Lemma 5.4.4 is centered around its expectation and not around its population counterpart is trivially solved by Lemma 3.2 in del Barrio et al. (2022), giving the convergence

$$\sqrt{n}\left(\mathbb{E}D_\epsilon(\mathrm{P}_n,\mathrm{Q}_m)-D_\epsilon(\mathrm{P},\mathrm{Q})\right)\longrightarrow 0,$$

and the proof is thus completed. □

***Proof of Theorem 5.4.3*.** Theorem 5.4.1 states that $((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+2\mathcal{A}_{\mathrm{P}}))\xi_{\mathrm{P},\mathrm{P}}\mathbb{G}_{\mathrm{P}}$ belongs to $\mathcal{C}^\alpha(\Omega)$, for any $\alpha\in\mathbb{N}$ and taking $\alpha=2s$ we have $((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+2\mathcal{A}_{\mathrm{P}}))\xi_{\mathrm{P},\mathrm{P}}\mathbb{G}_{\mathrm{P}}\in W^{2s,2}(\Omega)$. By definition of adjoint, we obtain

$$\mathbb{G}_{\mathrm{P}}\left((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+2\mathcal{A}_{\mathrm{P}})\xi_{\mathrm{P},\mathrm{P}}\mathbb{G}_{\mathrm{P}}\right)=\mathbb{G}_{\mathrm{P}}\left(\tau(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+2\mathcal{A}_{\mathrm{P}})\xi_{\mathrm{P},\mathrm{P}}\mathbb{G}_{\mathrm{P}}\right) \\ =\langle\tau^*\mathbb{G}_{\mathrm{P}},(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+2\mathcal{A}_{\mathrm{P}})\xi_{\mathrm{P},\mathrm{P}}\mathbb{G}_{\mathrm{P}})\rangle_W,$$

and, since $(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+2\mathcal{A}_{\mathrm{P}})\xi_{\mathrm{P},\mathrm{P}}=(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+2\mathcal{A}_{\mathrm{P}})\tau\xi_{\mathrm{P},\mathrm{P}}$, also

$$\langle\tau^*\mathbb{G}_{\mathrm{P}},(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+2\mathcal{A}_{\mathrm{P}})\xi_{\mathrm{P},\mathrm{P}}\mathbb{G}_{\mathrm{P}})\rangle_W \\ =\langle\tau^*((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+2\mathcal{A}_{\mathrm{P}}))^*\tau^*\mathbb{G}_{\mathrm{P}},\xi_{\mathrm{P},\mathrm{P}}\mathbb{G}_{\mathrm{P}})\rangle_W.$$

Set $f \in W^{2s,2}(\Omega)$ and note that (passing if necessary by a.s. representations) Fubini's theorem yields

$$\begin{aligned}\langle f, \xi_{\mathrm{P},\mathrm{P}}\mathbb{G}_{\mathrm{P}}\rangle_W &= \lim_{n\to\infty}\sqrt{n}\sum_{|\alpha|\le 2s}\int D_\alpha f(\mathbf{y})D^{\mathbf{y}}_\alpha\xi_{\mathrm{P},\mathrm{P}}(\mathbf{x},\mathbf{y})d\ell_d(\mathbf{y})(d\mathrm{P}_n(\mathbf{x})-d\mathrm{P}(\mathbf{x}))\\ &= \mathbb{G}_{\mathrm{P}}\left(\sum_{|\alpha|\le 2s}\int D_\alpha f(\mathbf{y})D^{\mathbf{y}}_\alpha\xi_{\mathrm{P},\mathrm{P}}(\cdot,\mathbf{y})d\ell_d(\mathbf{y})\right) = \mathbb{G}_{\mathrm{P}}\left(\tau\mathcal{D}f\right),\end{aligned}$$

where the operator

$$W^{2s,2}(\Omega)\ni f\mapsto \mathcal{D}f = \sum_{|\alpha|\le 2s}\int D_\alpha f(\mathbf{y})D^{\mathbf{y}}_\alpha\xi_{\mathrm{P},\mathrm{P}}(\cdot,\mathbf{y})d\ell_d(\mathbf{y})\in W^{2s,2}(\Omega)$$

is bounded. Then we have

$$\begin{aligned}\mathbb{G}_{\mathrm{P}}\left((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+2\mathcal{A}_{\mathrm{P}})\xi_{\mathrm{P},\mathrm{P}}\mathbb{G}_{\mathrm{P}}\right) =&\mathbb{G}_{\mathrm{P}}\left(\tau\mathcal{D}\tau^*((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+2\mathcal{A}_{\mathrm{P}}))^*\tau^*\mathbb{G}_{\mathrm{P}}\right)\\ =&\langle\mathcal{D}\tau^*((1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+2\mathcal{A}_{\mathrm{P}}))^*\tau^*\mathbb{G}_{\mathrm{P}},\tau^*\mathbb{G}_{\mathrm{P}}\rangle_W\end{aligned}$$

and, therefore the limits can be stated as follows:

- in the one-sample case;
$$\frac{1}{4}\langle\Theta\tau^*\mathbb{G}_{\mathrm{P}},\tau^*\mathbb{G}_{\mathrm{P}}\rangle_W \tag{5.39}$$
- in the two-sample case;
$$\frac{1}{4}\langle\Theta\tau^*(\sqrt{\lambda}\mathbb{G}_{\mathrm{P}'}+\sqrt{1-\lambda}\mathbb{G}_{\mathrm{P}'}),\tau^*((\sqrt{\lambda}\mathbb{G}_{\mathrm{P}'}+\sqrt{1-\lambda}\mathbb{G}_{\mathrm{P}'}))\rangle_W, \tag{5.40}$$
where $\mathbb{G}_{\mathrm{P}'}$ is an independent copy of $\mathbb{G}_{\mathrm{P}}$.

We use the representation (5.26) to obtain

$$\tau^*(\sqrt{\lambda}\mathbb{G}_{\mathrm{P}'}+\sqrt{1-\lambda}\mathbb{G}_{\mathrm{P}'}) = \sum_{i=1}^\infty f_i(\sqrt{\lambda}N_i+\sqrt{1-\lambda}N_i').$$

Lemma 5.4.2 implies the equality

$$\tau^*(\sqrt{\lambda}\mathbb{G}_{\mathrm{P}'}+\sqrt{1-\lambda}\mathbb{G}_{\mathrm{P}'}) = \sum_{i=1}^\infty(\sqrt{\lambda}N_i+\sqrt{1-\lambda}N_i')\sum_{j=1}^\infty x_{i,j}e_i,$$

where we can apply the operator Θ, yielding

$$\Theta\tau^*(\sqrt{\lambda}\mathbb{G}_{\mathrm{P}'}+\sqrt{1-\lambda}\mathbb{G}_{\mathrm{P}'}) = \sum_{i=1}^\infty(\sqrt{\lambda}N_i+\sqrt{1-\lambda}N_i')\sum_{j=1}^\infty x_{i,j}\lambda_i e_i.$$

As a consequence, (5.40) is equal to

$$\frac{1}{4}\sum_{j=1}^\infty\lambda_i e_i\left(\sum_{i=1}^\infty x_{i,j}(\sqrt{\lambda}N_i+\sqrt{1-\lambda}N_i')\right)\left(\sum_{i=1}^\infty x_{i,j}(\sqrt{\lambda}N_i+\sqrt{1-\lambda}N_i')\right)$$

and we conclude. The same vervatim applies to (5.39) □

Acknowledgements

The authors would like to thank Eustasio del Barrio for commenting that his method based on the Efron-Stein inequality could work for the Sinkhorn Divergence under the alternative hypothesis.

5.5 Proofs of the Lemmas

Proof of Lemma 5.2.3. Note that it is enough to check out that the limits

$$\limsup_{\|h_1\|_s \to 0} \frac{\| \log \left(\mathbb{E} \left(e^{(f_{\mathrm{P,Q}}+h_1)(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2} \right) \right) - \log \left(\mathbb{E} \left(e^{(f_{\mathrm{P,Q}})(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2} \right) \right) - \mathcal{A}_{\mathrm{P}}(h_1) \|_s}{\|h_1\|_s},$$

and

$$\limsup_{\|h_2\|_s \to 0} \frac{\| \log \left(\mathbb{E} \left(e^{(g_{\mathrm{P,Q}}+h_2)(\mathbf{Y})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2} \right) \right) - \log \left(\mathbb{E} \left(e^{(f_{\mathrm{P,Q}})(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2} \right) \right) - \mathcal{A}_{\mathrm{Q}}(h_2) \|_s}{\|h_2\|_s}$$

are 0 . We focus only o the first one, the second follows by the same arguments. Lemma 5.6.5 yields $\|e^{f_{\mathrm{P,Q}}+h_1} - e^{f_{\mathrm{P,Q}}} - h_1 e^{f_{\mathrm{P,Q}}}\|_s = o(\|h_1\|_s)$. The linearity and continuity of the operator $h \to \mathbb{E}\left(h(\mathbf{X})e^{-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)$ implies its Fréchet differentiability. By composition, the operator $f \to \mathbb{E}\left(e^{f(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)$ is also Fréchet differentiable. Since $\mathbb{E}\left(e^{f(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right) > 0$, Lemma 5.6.5 applied to the logarithm concludes the proof. □

Proof of Lemma 5.2.4. Recall the notation $\mathrm{P}^{\mathbf{X}}(f_{\mathrm{P,Q}}(\mathbf{X},\cdot)) = \mathbb{E}_{\mathbf{X}}(f_{\mathrm{P,Q}}(\mathbf{X},\cdot))$ and similar with the empirical version $\mathrm{P}_n^{\mathbf{X}}(f_{\mathrm{P,Q}}(\mathbf{X},\cdot)) = \frac{1}{n}\sum_{k=1}^n (f(\mathbf{X}_k,\cdot))$, for any integrable $f : \Omega \times \Omega \to \mathbb{R}$. Note that, the second component of the difference

$$(\Psi(f_{\mathrm{P,Q}}, g_{\mathrm{P,Q}}) - \Psi(f_{\mathrm{P}_n,\mathrm{Q}_m}, g_{\mathrm{P}_n,\mathrm{Q}_m})) - (\Psi_{n,m}(f_{\mathrm{P,Q}}, g_{\mathrm{P,Q}}) - \Psi(f_{\mathrm{P,Q}}, g_{\mathrm{P,Q}}))$$

can be written as

$$\begin{aligned} &\log \mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right) - \log \mathrm{P}_n^{\mathbf{X}}\left(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right) \\ &- \log \mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P,Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right) - \log \mathrm{P}_n^{\mathbf{X}}\left(e^{f_{\mathrm{P,Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right). \end{aligned} \tag{5.41}$$

In accordance with Lemma 5.6.5—applied to the natural logarithm—we obtain the relation

$$\begin{aligned} &\log \mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right) = \log \mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P,Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right) \\ &+ \frac{\mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)}{\mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P,Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)} - 1 + o_P(\|\mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2} - e^{f_{\mathrm{P,Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)\|_s) \end{aligned}$$

in $\mathcal{C}^s(\Omega)$, and, in view of Lemma 5.6.3, also

$$\log \mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right) = \log \mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right) + \frac{\mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)}{\mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)} - 1 + o_P\left(\sqrt{\frac{n+m}{n\,m}}\right). \quad (5.42)$$

On the other hand

$$\log \mathrm{P}_n^{\mathbf{X}}\left(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right) = \log \mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right) + \frac{\mathrm{P}_n^{\mathbf{X}}\left(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)}{\mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)} - 1 + o_P(\|\mathrm{P}_n^{\mathbf{X}}\left(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right) - \mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)\|_s)$$

in $\mathcal{C}^s(\Omega)$. Since

$$\begin{aligned}
&\|\mathrm{P}_n^{\mathbf{X}}\left(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right) - \mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)\|_s \\
&\leq \|\mathrm{P}_n^{\mathbf{X}}\left(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2} - e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)\|_s + \|(\mathrm{P}_n^{\mathbf{X}} - \mathrm{P}^{\mathbf{X}})\left(e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)\|_s \\
&\leq \|e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X})-\frac{1}{2}\|\cdot\|^2} - e^{f_{\mathrm{P},\mathrm{Q}}-\frac{1}{2}\|\cdot\|^2}\|_s + \|(\mathrm{P}_n^{\mathbf{X}} - \mathrm{P}^{\mathbf{X}})\left(e^{f_{\mathrm{P},\mathrm{Q}}-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)\|_s,
\end{aligned}$$

applying Lemma 5.6.3, we have

$$\log \mathrm{P}_n^{\mathbf{X}}\left(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right) = \log \mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right) + \frac{\mathrm{P}_n^{\mathbf{X}}\left(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)}{\mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)} - 1 + o_P\left(\sqrt{\frac{n+m}{n\,m}}\right). \quad (5.43)$$

Finally, the same argument shows that

$$\log \mathrm{P}_n^{\mathbf{X}}\left(e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right) = \log \mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right) + \frac{\mathrm{P}_n^{\mathbf{X}}\left(e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)}{\mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)} - 1 + o_P\left(\sqrt{\frac{n+m}{n\,m}}\right). \quad (5.44)$$

The relations (5.42),(5.43) and (5.44) yield the following rewriting of (5.41);

$$\frac{(\mathrm{P}_n^{\mathbf{X}} - \mathrm{P}^{\mathbf{X}})\left(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2} - e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)}{\mathrm{P}^{\mathbf{X}}\left(e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)} + o_P\left(\sqrt{\frac{n+m}{n\,m}}\right). \quad (5.45)$$

Then, since $0 < \|P^{\mathbf{X}}\left(e^{f_{P,Q}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)\|_\infty < \infty$, we have to prove that

$$\frac{n\,m}{n+m}\,\mathbb{E}\|(P^{\mathbf{X}} - P_n^{\mathbf{X}})\left(e^{f_{P_n,Q_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2} - e^{f_{P,Q}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)\|_\infty \le C, \qquad (5.46)$$

where $C > 0$ is a deterministic constant. Setting $\mathbf{y} \in \Omega$ we have

$$\begin{aligned}
&|(P^{\mathbf{X}} - P_n^{\mathbf{X}})\left(e^{f_{P_n,Q_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\mathbf{y}\|^2} - e^{f_{P,Q}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\mathbf{y}\|^2}\right)| \\
&\quad \le \|e^{f_{P_n,Q_m}} - e^{f_{P,Q}}\|_s \sup_{\|f\|_s\le 1} (P^{\mathbf{X}} - P_n^{\mathbf{X}})(e^{\langle \mathbf{X},\mathbf{y}\rangle} f(\mathbf{X})) \\
&\quad \le \|e^{f_{P_n,Q_m}} - e^{f_{P,Q})}\|_s \sup_{\mathbf{y}\in\Omega}\sup_{\|f\|_s\le 1} (P^{\mathbf{X}} - P_n^{\mathbf{X}})(e^{\langle \mathbf{X},\mathbf{y}\rangle} f(\mathbf{X})).
\end{aligned}$$

At this point, we proceed as in the proof of Lemma 3.2. in del Barrio et al. (2022). Note that (5.53) gives the bound

$$\mathbb{E}\|(P^{\mathbf{X}} - P_n^{\mathbf{X}})\left(e^{f_{P_n,Q_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2} - e^{f_{P,Q}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\right)\|_s \le C\sqrt{\mathbb{E}\|e^{f_{P_n,Q_m}} - e^{f_{P,Q}}\|_s^2\,\mathbb{E}\|P - P_n\|_s'^2},$$

which, in view of Lemma 5.6.3, can be majorized by

$$\frac{C(\Omega,d,s)}{\sqrt{n}}\sqrt{\mathbb{E}\|P - P_n\|_s'^2} \le \frac{C(\Omega,d,s)}{n},$$

where the second inequality is consequence of Lemma 5.6.4. Therefore (5.46) holds, and, repeating the same arguments for the second component of Ψ, thus a *fortiori*

$$\begin{aligned}
&\|\Psi_{n,m}(f_{P_n,Q_m}, g_{P_n,Q_m}) - \Psi(f_{P_n,Q_m}, g_{P_n,Q_m}) - \\
&\qquad (\Psi_{n,m}(f_{P,Q}, g_{P,Q}) - \Psi(f_{P,Q}, g_{P,Q})\|_{\mathcal{C}^s(\Omega)\times\mathcal{C}^s(\Omega)} = o_P\left(\sqrt{\frac{n+m}{n\,m}}\right). \qquad (5.47)
\end{aligned}$$

Therefore, we can write the following equality in $\mathcal{C}^s(\Omega)\times\mathcal{C}^s(\Omega)$;

$$\begin{aligned}
&\sqrt{\frac{n\,m}{n+m}}(\Psi_{n,m}(f_{P_n,Q_m}, g_{P_n,Q_m}) - \Psi(f_{P_n,Q_m}, g_{P_n,Q_m})) \\
&\qquad = \sqrt{\frac{n\,m}{n+m}}(\Psi_{n,m}(f_{P,Q}, g_{P,Q}) - \Psi(f_{P,Q}, g_{P,Q})) + o_P(1). \qquad (5.48)
\end{aligned}$$

Using $\Psi(f_{P,Q}, g_{P,Q}) = \Psi_{n,m}(f_{P_n,Q_m}, g_{P_n,Q_m}) = 0$ (see (5.16)) and (5.48), we have

$$\begin{aligned}
&\sqrt{\frac{n\,m}{n+m}}(\Psi(f_{P,Q}, g_{P,Q}) - \Psi(f_{P_n,Q_m}, g_{P_n,Q_m})) \\
&\quad = \sqrt{\frac{n\,m}{n+m}}(\Psi_{n,m}(f_{P_n,Q_m}, g_{P_n,Q_m}) - \Psi(f_{P_n,Q_m}, g_{P_n,Q_m})) \\
&\qquad = \sqrt{\frac{n\,m}{n+m}}(\Psi_{n,m}(f_{P,Q}, g_{P,Q}) - \Psi(f_{P,Q}, g_{P,Q})) + o_P(1).
\end{aligned}$$

□

Proof of Lemma 5.4.2. It is easy to see that $\mathcal{D}$ is compact (the proof of Lemma 5.2.5 for $\mathcal{A}_{\mathrm{P}}$ applies verbatim to $\mathcal{D}$), and self-adjoint; for every $f, g \in W^{2s,2}(\Omega)$,

$$\langle \mathcal{D}f, g\rangle_W = \sum_{|\alpha|\leq 2s, |\beta|\leq 2s} \int D_\alpha f(\mathbf{y}) D^{\mathbf{y}}_{\alpha+\beta} \xi_{\mathrm{P,P}}(\mathbf{x},\mathbf{y}) D_\beta g(\mathbf{y}) d\ell_d(\mathbf{y}) d\ell_d(\mathbf{x}),$$

which is symmetric in f and g. Then we have

$$\begin{aligned}\langle \mathcal{D}\tau^*(1+2\mathcal{A}_{\mathrm{P}}^*)(1-\mathcal{A}_{\mathrm{P}}^*\mathcal{A}_{\mathrm{P}}^*)^{-1} f, g\rangle_W &= \langle f, (1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+2\mathcal{A}_{\mathrm{P}})\tau\mathcal{D}^* g\rangle_W \\ &= \langle f, (1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+2\mathcal{A}_{\mathrm{P}})\tau\mathcal{D} g\rangle_W,\end{aligned}$$

for all $f, g \in W^{2s,2}(\Omega)$. Therefore, we should prove that $(1-\mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}(1+2\mathcal{A}_{\mathrm{P}})\tau\mathcal{D}$ is self-adjoint. As an intermediate step, we prove that $\mathcal{A}_{\mathrm{P}}\tau\mathcal{D}$ is also self-adjoint. We compute

$$\begin{aligned}\langle f, \mathcal{A}_{\mathrm{P}}\tau\mathcal{D}g\rangle_W &= \sum_{|\alpha|\leq 2s} \int D_\alpha f(\mathbf{x}) D_\alpha \mathcal{A}_{\mathrm{P}} \sum_{|\beta|\leq 2s} \int D_\beta g(\mathbf{y}) D^{\mathbf{y}}_\beta \xi_{\mathrm{P,P}}(\mathbf{x},\mathbf{y}) d\ell_d(\mathbf{y}) d\ell_d(\mathbf{x}) \\ &= \sum_{|\alpha|\leq 2s}\sum_{|\beta|\leq 2s} \int D_\alpha f(\mathbf{x}) D_\alpha \mathcal{A}_{\mathrm{P}} D_\beta g(\mathbf{y}) D^{\mathbf{y}}_\beta \xi_{\mathrm{P,P}}(\mathbf{x},\mathbf{y}) d\ell_d(\mathbf{y}) d\ell_d(\mathbf{x}) \\ &= \sum_{|\alpha|\leq 2s}\sum_{|\beta|\leq 2s} \int D_\alpha f(\mathbf{x}) D_\beta g(\mathbf{y}) D_\alpha \mathcal{A}_{\mathrm{P}} D^{\mathbf{y}}_\beta \xi_{\mathrm{P,P}}(\mathbf{x},\mathbf{y}) d\ell_d(\mathbf{y}) d\ell_d(\mathbf{x}),\end{aligned}$$

and then, realising that the symmetry $\xi_{\mathrm{P,P}}(\mathbf{x},\mathbf{y}) = \xi_{\mathrm{P,P}}(\mathbf{y},\mathbf{x})$ gives, for each $\mathbf{y} \in \Omega$, the relation

$$D_\alpha \mathcal{A}_{\mathrm{P}} D^{\mathbf{y}}_\beta \xi_{\mathrm{P,P}}(\cdot,\mathbf{y}) = \int D^{\mathbf{y}}_\alpha(\xi_{\mathrm{P,P}}(\mathbf{x}',\cdot)) D^{\mathbf{y}}_\beta \xi_{\mathrm{P,P}}(\mathbf{x}',\mathbf{y}) d\mathrm{P}(\mathbf{x}'),$$

and *a fortitori*

$$\begin{aligned}&\langle f, \mathcal{A}_{\mathrm{P}}\tau\mathcal{D}g\rangle_W \\ &= \sum_{|\alpha|\leq 2s}\sum_{|\beta|\leq 2s} \int \Big(D_\alpha f(\mathbf{x}) D_\beta g(\mathbf{y}) D^{\mathbf{y}}_\alpha(\xi_{\mathrm{P,P}}(\mathbf{x}',\mathbf{x})) D^{\mathbf{y}}_\beta \xi_{\mathrm{P,P}}(\mathbf{x}',\mathbf{y}) \Big) d\mathrm{P}(\mathbf{x}') d\ell_d(\mathbf{y}) d\ell_d(\mathbf{x}),\end{aligned}$$

which proves that $\mathcal{A}_{\mathrm{P}}\tau\mathcal{D}$ is self-adjoint and, moreover, compact (see the proof of Lemma 5.2.5). Theorem 6.8. in Brezis (2010) yields the existence of a sequence of eigenvalues $\{\nu_j\}_{j\in\mathbb{N}}$ of $\mathcal{A}_{\mathrm{P}}$, all of them in the interval $(-1,1)$, for certain $\epsilon > 0$—otherwise, for any centered $f \in \mathcal{C}^s(\Omega)$,

$$\begin{aligned}\mathcal{A}_{\mathrm{P}} f &= \nu f, \\ \int \left(\int \xi_{\mathrm{P,P}}(\mathbf{x},\mathbf{y}) f(\mathbf{x}) d\mathrm{P}(\mathbf{x}) \right)^2 d\mathrm{P}(\mathbf{y}) &= \nu^2 \int f(\mathbf{x})^2 d\mathrm{P}(\mathbf{x}),\end{aligned}$$

which, using Jensen's inequality and the optimality conditions;

$$\nu^2 \int f(\mathbf{x})^2 d\mathrm{P}(\mathbf{x}) = \int \left(\int \xi_{\mathrm{P},\mathrm{P}}(\mathbf{x},\mathbf{y}) f(\mathbf{x}) d\mathrm{P}(\mathbf{x}) \right)^2 d\mathrm{P}(\mathbf{y}) < \int f(\mathbf{x})^2 d\mathrm{P}(\mathbf{x}),$$

yields to contradiction. Moreover, since far from 0, the sequence $\{\nu_j\}_{j\in\mathbb{N}}$ has not acumulative points (Theorem 6.8. in Brezis (2010)), $\{\nu_j\}_{j\in\mathbb{N}} \subset [-1+\epsilon, 1+\epsilon]$, for some $\epsilon > 0$. This implies that the Neumann series $(1-\mathcal{A}_{\mathrm{P}}^2)^{-1}\tau = \sum_{k=1}^{\infty} (\mathcal{A}_{\mathrm{P}})^{2k}\tau$ converges as operators and we can express

$$(1 - \mathcal{A}_{\mathrm{P}}\mathcal{A}_{\mathrm{P}})^{-1}\tau\mathcal{D} = \sum_{k=1}^{\infty} (\mathcal{A}_{\mathrm{P}})^{2k}\tau\mathcal{D} = \sum_{k=1}^{\infty} (\mathcal{A}_{\mathrm{P}}\tau)^{2k}\mathcal{D}, \tag{5.49}$$

where the last equality comes from the fact that $\mathcal{A}_{\mathrm{P}} f \in \mathcal{C}^{\infty}(\Omega)$ for all $f \in \mathcal{C}(\Omega)$. Since, for any arbitrary $k \in \mathbb{N}$

$$\begin{aligned}((\mathcal{A}_{\mathrm{P}}\tau)^k\mathcal{D})^* &= \mathcal{D}^*(\mathcal{A}_{\mathrm{P}}\tau)^*((\mathcal{A}_{\mathrm{P}}\tau)^{k-1})^* = \mathcal{A}_{\mathrm{P}}\tau\mathcal{D}((\mathcal{A}_{\mathrm{P}}\tau)^{k-1})^* = \mathcal{A}_{\mathrm{P}}\tau\mathcal{D}^*((\mathcal{A}_{\mathrm{P}}\tau)^{k-1})^* \\ &= \mathcal{A}_{\mathrm{P}}\tau\mathcal{D}^*(\mathcal{A}_{\mathrm{P}}\tau)^*((\mathcal{A}_{\mathrm{P}}\tau)^{k-2})^* = (\mathcal{A}_{\mathrm{P}}\tau)^2\mathcal{D}((\mathcal{A}_{\mathrm{P}}\tau)^{k-2})^* \\ &= \ldots = (\mathcal{A}_{\mathrm{P}}\tau)^k\mathcal{D},\end{aligned}$$

(5.49) is an infinite sum of self-adjoint operators. As limits of self-adjoint operators are still self-adjoint—this is consequence of the inequality $\|T_n f - Tf\|_W \leq \|T_n - T\|\|f\|_W$—$\Theta$ is self-adjoint.

To prove that is positive, it is enough to see that the values $\{\nu_j\}_{j\in\mathbb{N}}$ are positive. Suppose that the contrary holds, i.e. $\mathcal{A}_{\mathrm{P}} f = -\nu f$, for $\nu > 0$, then

$$\int \xi_{\mathrm{P},\mathrm{P}}(\mathbf{x},\mathbf{y}) f(\mathbf{x}) d\mathrm{P}(\mathbf{x}) f(\mathbf{y}) d\mathrm{P}(\mathbf{y}) = \int e^{f_{\mathrm{P},\mathrm{P}}(\mathbf{x})} e^{-\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2} f(\mathbf{x}) d\mathrm{P}(\mathbf{x}) f(\mathbf{y}) e^{f_{\mathrm{P},\mathrm{P}}(\mathbf{y})} d\mathrm{P}(\mathbf{y}),$$

which is the Gaussian kernel $e^{-\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2}$ acting on the measure $f(\mathbf{y})e^{f_{\mathrm{P},\mathrm{P}}(\mathbf{y})}d\mathrm{P}(\mathbf{y})$. As a consequence

$$\int\int \xi_{\mathrm{P},\mathrm{P}}(\mathbf{x},\mathbf{y}) f(\mathbf{x}) d\mathrm{P}(\mathbf{x}) f(\mathbf{y}) d\mathrm{P}(\mathbf{y}) \geq 0,$$

which contradicts $\mathcal{A}_{\mathrm{P}} f = -\nu f$, for $\nu > 0$. □

5.6 Auxiliary results

Lemma 5.6.1. *Let $\Omega \subset \mathbb{R}^d$ be a compact set,* $\mathrm{P}, \mathrm{Q} \in \mathcal{P}(\Omega)$*, and its empirical versions* $\mathrm{P}_n, \mathrm{Q}_m$*, with* $\frac{n}{m+n} \to \lambda \in (0,1)$*. Then*

$$\|e^{h_{\mathrm{P}_n,\mathrm{Q}_m}} - e^{h_{\mathrm{P},\mathrm{Q}}} - e^{h_{\mathrm{P},\mathrm{Q}}}(h_{\mathrm{P}_n,\mathrm{Q}_m} - h_{\mathrm{P},\mathrm{Q}})\|_{\mathcal{C}^s(\Omega\times\Omega)} = o_P\left(\sqrt{\frac{n+m}{nm}}\right), \quad \textit{and} \tag{5.50}$$

$$\|e^{h_{\mathrm{P}_n,\mathrm{Q}_m}} - e^{h_{\mathrm{P},\mathrm{Q}}}\|_{\mathcal{C}^s(\Omega\times\Omega)} = O_P\left(\sqrt{\frac{n+m}{nm}}\right). \tag{5.51}$$

Proof. Since the functions $h_{\mathrm{P}_n,\mathrm{Q}_m}$ and $h_{\mathrm{P}_n,\mathrm{Q}_m}$ are uniformly bounded, for all $n, m \in \mathbb{N}$, Lemma 5.6.5 applied to the exponential gives

$$\|e^{h_{\mathrm{P}_n,\mathrm{Q}_m}} - e^{h_{\mathrm{P},\mathrm{Q}}} - e^{h_{\mathrm{P},\mathrm{Q}}}(h_{\mathrm{P}_n,\mathrm{Q}_m} - h_{\mathrm{P},\mathrm{Q}})\|_s = o_P(\|h_{\mathrm{P},\mathrm{Q}} - h_{\mathrm{P}_n,\mathrm{Q}_m}\|_s)$$

and, using Theorem 4.5 in del Barrio et al. (2022), we obtain (5.50). To prove (5.51) we apply the inverse triangle inequality to (5.50);

$$\begin{aligned}&\|e^{h_{\mathrm{P}_n,\mathrm{Q}_m}} - e^{h_{\mathrm{P},\mathrm{Q}}} - e^{h_{\mathrm{P},\mathrm{Q}}}(h_{\mathrm{P}_n,\mathrm{Q}_m} - h_{\mathrm{P},\mathrm{Q}})\|_{\mathcal{C}^s(\Omega\times\Omega)}\\ &\qquad\geq \|e^{h_{\mathrm{P}_n,\mathrm{Q}_m}} - e^{h_{\mathrm{P},\mathrm{Q}}}\|_{\mathcal{C}^s(\Omega\times\Omega)} - \|e^{h_{\mathrm{P},\mathrm{Q}}}(h_{\mathrm{P}_n,\mathrm{Q}_m} - h_{\mathrm{P},\mathrm{Q}})\|_{\mathcal{C}^s(\Omega\times\Omega)}.\end{aligned}$$

Then we apply Lemma 5.6.6 and (5.50) to obtain

$$\|e^{h_{\mathrm{P}_n,\mathrm{Q}_m}} - e^{h_{\mathrm{P},\mathrm{Q}}}\|_{\mathcal{C}^s(\Omega\times\Omega)} \leq C\|(h_{\mathrm{P}_n,\mathrm{Q}_m} - h_{\mathrm{P},\mathrm{Q}})\|_{\mathcal{C}^s(\Omega\times\Omega)} + o_P\left(\sqrt{\frac{n+m}{nm}}\right).$$

Theorem 4.5 in del Barrio et al. (2022) concludes. □

Lemma 5.6.2. *Let $\Omega \subset \mathbb{R}^d$ be a compact set and $\mathrm{P}, \mathrm{Q} \in \mathcal{P}(\Omega)$, then*

$$|S_1(\mathrm{P},\mathrm{Q}) - \int h d\mathrm{P} d\mathrm{Q} - \frac{1}{2}\int (h - h_{\mathrm{P},\mathrm{Q}})^2 d\pi_{\mathrm{P},\mathrm{Q}}| \leq \frac{1}{6}\|h - h_{\mathrm{P},\mathrm{Q}}\|_\infty^3 e^{\|h - h_{\mathrm{P},\mathrm{Q}}\|_\infty},$$

for all $h(\mathbf{x},\mathbf{y}) = f(\mathbf{x}) + g(\mathbf{y})$, with $f, g \in \mathcal{C}(\Omega)$ and $\int e^{h(\mathbf{x},\mathbf{y}) - \frac{1}{2}||\mathbf{x}-\mathbf{y}||^2} d\mathrm{P}(\mathbf{x}) d\mathrm{Q}(\mathbf{y}) = 1$.

Proof. The application of Taylor's theorem to the exponential gives

$$|e^x - 1 - x - \frac{1}{2}x^2| \leq \frac{1}{6}|x^3|e^{|x|}. \tag{5.52}$$

Since also $\int e^{h_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y}) - \frac{1}{2}||\mathbf{x}-\mathbf{y}||^2} d\mathrm{P}(\mathbf{x}) d\mathrm{Q}(\mathbf{y}) = 1$, (5.5) yields

$$\begin{aligned}S_1(\mathrm{P},\mathrm{Q}) = &\int h_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y}) d\mathrm{P}(\mathbf{x})\mathrm{Q}(\mathbf{y})\\ &\qquad - \int \left(e^{h_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y}) - \frac{1}{2}||\mathbf{x}-\mathbf{y}||^2} - e^{h(\mathbf{x},\mathbf{y}) - \frac{1}{2}||\mathbf{x}-\mathbf{y}||^2}\right) d\mathrm{P}(\mathbf{x}) d\mathrm{Q}(\mathbf{y}).\end{aligned}$$

We can decompose

$$\begin{aligned}&-\int \left(e^{h_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y}) - \frac{1}{2}||\mathbf{x}-\mathbf{y}||^2} - e^{h(\mathbf{x},\mathbf{y}) - \frac{1}{2}||\mathbf{x}-\mathbf{y}||^2}\right) d\mathrm{P}(\mathbf{x}) d\mathrm{Q}(\mathbf{y})\\ &\qquad = \int e^{h_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y}) \frac{1}{2}||\mathbf{x}-\mathbf{y}||^2}(e^{h(\mathbf{x},\mathbf{y}) - h_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y})} - 1) d\mathrm{P}(\mathbf{x}) d\mathrm{Q}(\mathbf{y})\end{aligned}$$

and, finally, (5.52) implies

$$
\begin{aligned}
|S_1(\mathrm{P},\mathrm{Q}) - & \int h_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y}) d\mathrm{P}(\mathbf{x})\mathrm{Q}(\mathbf{y}) \\
& - \int \left(h(\mathbf{x},\mathbf{y}) - h_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y}) + \frac{1}{2}(h(\mathbf{x},\mathbf{y}) - h_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y}))^2 \right) d\pi_{\mathrm{P},\mathrm{Q}}(\mathbf{x},\mathbf{y})| \\
& \leq \frac{1}{6}\|h - h_{\mathrm{P},\mathrm{Q}}\|_\infty^3 e^{\|h - h_{\mathrm{P},\mathrm{Q}}\|_\infty}.
\end{aligned}
$$

Since (5.5) cancels the linear terms, the proof is completed. □

Lemma 5.6.3. *Let $\Omega \subset \mathbb{R}^d$ be a compact set, $\mathrm{P},\mathrm{Q} \in \mathcal{P}(\Omega)$ and the associate empirical measures P_n and Q_m. Then there exists a constant $C(\Omega, d, s)$ such that*

$$
\|e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X}) - \frac{1}{2}\|\mathbf{X} - \cdot\|^2} - e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X}) - \frac{1}{2}\|\mathbf{X} - \cdot\|^2}\|_s \leq C(\Omega, d, s)\|f_{\mathrm{P},\mathrm{Q}} - f_{\mathrm{P}_n,\mathrm{Q}_m}\|_s
$$

and

$$
\mathbb{E}\left(\|e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X}) - \frac{1}{2}\|\mathbf{X} - \cdot\|^2} - e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X}) - \frac{1}{2}\|\mathbf{X} - \cdot\|^2}\|_s\right)^2 \leq C(\Omega, d, s)\left(\frac{1}{n} + \frac{1}{m}\right).
$$

Proof. The proof follows by Faà di Bruno's formula. We illustrate first the first case–which corresponds with the uniform norm. The mean value theorem yields,

$$
|e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{x}) - \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2} - e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{x}) - \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2}| \leq e^{\mathrm{diam}(\Omega)^2 + \|f_{\mathrm{P},\mathrm{Q}}\|_\infty + \|f_{\mathrm{P}_n,\mathrm{Q}_m}\|_\infty}\|f_{\mathrm{P}_n,\mathrm{Q}_m} - f_{\mathrm{P},\mathrm{Q}}\|_\infty, \quad \mathbf{x},\mathbf{y} \in \Omega.
$$

For higher order analysis, we set $a \leq s$, and apply Faà di Bruno's formula, which yields the existence of some positive constants $\{\lambda_{a,b}\}_{|b|=a}$ such that

$$
\begin{aligned}
&|D_a(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{x}) - \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2} - e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{x}) - \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2})| \\
&\leq \sum_{|b|=a} \lambda_{a,b} \left| e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{x}) - \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2} \prod_{j=1}^d D_b f_{\mathrm{P},\mathrm{Q}}(\mathbf{x}) - e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{x}) - \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2} \prod_{j=1}^d D_b f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{x}) \right|.
\end{aligned}
$$

Lemma 5.6.1 yields

$$
\begin{aligned}
&|D_a(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{x}) - \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2} - e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{x}) - \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2})| \\
&\leq \sum_{|b|=a} \lambda_{a,b} e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{x}) - \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2} \left| \prod_{j=1}^d D_{b_j} f_{\mathrm{P},\mathrm{Q}}(\mathbf{x}) - \prod_{j=1}^d D_{b_j} f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{x}) \right| + C(a,\Omega)\|f_{\mathrm{P},\mathrm{Q}} - f_{\mathrm{P}_n,\mathrm{Q}_n}\|_s.
\end{aligned}
$$

Writing $\prod_{j=1}^d D_{b_j} f_{\mathrm{P},\mathrm{Q}}(\mathbf{x}) - \prod_{j=1}^d D_{b_j} f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{x})$ as the telescopic sum

$$
\begin{aligned}
&\sum_{j=1}^n \left(\prod_{i \leq j} D_{b_i} f_{\mathrm{P},\mathrm{Q}}(\mathbf{x}) \prod_{i > j} D_{b_i} f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{x}) - \prod_{i < j} D_{b_i} f_{\mathrm{P},\mathrm{Q}}(\mathbf{x}) \prod_{i \geq j} D_{b_i} f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{x}) \right) \\
&\qquad = \sum_{j=1}^n \prod_{i<j} D_{b_i} f_{\mathrm{P},\mathrm{Q}}(\mathbf{x}) \prod_{i>j} D_{b_i} f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{x}) \left(D_{b_j} f_{\mathrm{P},\mathrm{Q}}(\mathbf{x}) - D_{b_j} f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{x}) \right),
\end{aligned}
$$

we obtain the bound

$$|D_a(e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{x})-\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2} - e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{x})-\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2})| \leq C(\Omega, d, s)\|f_{\mathrm{P},\mathrm{Q}} - f_{\mathrm{P}_n,\mathrm{Q}_m}\|_s.$$

Since this holds for any $\mathbf{x}, \mathbf{y} \in \Omega$, we obtain

$$\|e^{f_{\mathrm{P}_n,\mathrm{Q}_m}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2} - e^{f_{\mathrm{P},\mathrm{Q}}(\mathbf{X})-\frac{1}{2}\|\mathbf{X}-\cdot\|^2}\|_s \leq C(\Omega, d, s)\|f_{\mathrm{P},\mathrm{Q}} - f_{\mathrm{P}_n,\mathrm{Q}_m}\|_s.$$

The proof is thus completed by applying Theorem 4.5 in del Barrio et al. (2022). □

Lemma 5.6.4. *Let $\Omega \subset \mathbb{R}^d$ be a compact set, $\mathrm{P}, \in \mathcal{P}(\Omega)$ and P_n be the empirical measure of the i.i.d. sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ distributed as P. Then*

$$n\,\mathbb{E}\left(\sup_{\mathbf{y}\in\Omega, f\in\mathcal{C}^s(\Omega), \|f\|_s\leq 1}\left|\int f(\mathbf{x})e^{-\|\mathbf{x}-\mathbf{y}\|^2}d(\mathrm{P}_n - \mathrm{P})(\mathbf{x})\right|\right)^2 = O(1).$$

Moreover, the class $\{\mathbf{x} \mapsto g(\mathbf{y})e^{\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2}f(\mathbf{x}), \quad \mathbf{y} \in \Omega,\ \|f\|_s \leq 1,\ \|g\|_s \leq 1\}$ is P-Donsker.

Proof. Since

$$\{\mathbf{x} \mapsto g(\mathbf{y})e^{\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2}f(\mathbf{x}), \quad \mathbf{y} \in \Omega,\ \|f\|_s \leq 1,\ \|g|_s \leq 1\} \subset \{f \in \mathcal{C}^s(\Omega), \quad \|f\|_s \leq C\} \tag{5.53}$$

holds for certain constant $C > 0$ and 2.7.2 Corollary Vaart and Wellner (1996) and Exercise 2.3.1 in Giné and Nickl (2015) give

$$\mathbb{E}\left(\sup_{\|f\|_s\leq 1}(\mathrm{P}_n - \mathrm{P})(f)\right)^2 \leq \frac{C}{n}.$$

Therefore the first statement holds. The last one is consequence of (5.53), Theorems 2.5.2 and 2.7.1 in Vaart and Wellner (1996) □

Lemma 5.6.5. *Let $\alpha \in \mathbb{N}$, $I \subset \mathbb{R}$ be a compact interval, $F \in \mathcal{C}^\infty(I)$ and $g \in \mathcal{C}^\alpha(\Omega)$, with $g(\Omega) \subset I$. Then the operator*

$$\begin{aligned} \delta_F : \mathcal{C}^\alpha(\Omega) &\longrightarrow \mathcal{C}^\alpha(\Omega) \\ g &\longmapsto F(g), \end{aligned}$$

is Fréchet differentiable in g with derivative $D\delta_F(g)h = F'(g)h$. Moreover, setting $f \in \mathcal{C}^\alpha(\Omega)$, the operator

$$\begin{aligned} \mathcal{C}^\alpha(\Omega) &\longrightarrow \mathcal{C}^\alpha(\Omega) \\ g &\longmapsto fg, \end{aligned}$$

is Fréchet differentiable in g with derivative $h \to fh$.

Proof. To check out the first claim, i.e.

$$\|F(g+h) - F(g) - F'(g)h\|_\alpha = o(\|h\|_\alpha),$$

we prove the bound for each derivarive in a direction a by a recursive argument on $|a|$. Note that the first case, the uniform norm, is trivial; set $\mathbf{x} \in \Omega$ and apply Taylor's theorem to obtain

$$|F(g(\mathbf{x}) + h(\mathbf{x})) - F(g(\mathbf{x})) - F'(g(\mathbf{x}))h(\mathbf{x})| \leq |F''(g(\mathbf{x}))h(\mathbf{x})^2| \leq \sup_I |F| \|h\|_s.$$

Suppose that the result holds for $a \in \mathbb{N}$ and derivative in a direction b with $|b| = a+1$, then there exists a decomposition $D_b = D_c D_d$ with $|d| = 1$ and $|c| = a$. Therefore, the chain rule yields

$$\begin{aligned} &D_b(F(g+h) - F(g) - F'(g)h) \\ &= D_c(F'(g+h)(D_d g + D_d h) - F'(g))D_d g - F''(g)D_d g h - F'(g)D_g h \\ &= D_c\left((F'(g+h) - F'(g) - F''(g)h)D_d g + (F'(g+h) - F'(g))D_d h\right). \end{aligned}$$

Since the function F' still satisfies the assumptions of the theorem we obtain, by induction hypothesis and Lemma 5.6.6, the limits

$$\begin{aligned} &\|D_c((F'(g+h) - F'(g) - F''(g)h)D_d g)\|_\infty \\ &\qquad \leq C\|D_c(F'(g+h)(D_d g + D_d h) - F'(g))\|_\infty \|g\|_s = o(\|h\|_s) \end{aligned}$$

and

$$\|D_c((F'(g+h) - F'(g))D_d h)\|_\infty \leq C\|D_c(F'(g+h) - F'(g))\|_\infty \|h\|_s = o(\|h\|_s),$$

which finish the proof of the first claim. For the second one the relation $f(g+h) - fg - fh = 0$ and Lemma 5.6.6 conclude. □

Lemma 5.6.6. *Let $f, g \in \mathcal{C}^s(\Omega)$ then there exists a constant C depending on Ω, s and d such that*

$$\|f\,g\|_s \leq C\|f\|_s \|g\|_s.$$

Proof. The proof is direct consequence of the multivariate Leibniz rule, i.e.

$$\|D_a(f\,g)\|_\infty \leq \sum_{|\alpha| \leq a} \binom{a}{\alpha} \|D_\alpha f\|_\infty \|D_{a-\alpha} g\|_\infty \leq C\|f\|_s \|g\|_s.$$

□

Bibliography

Robert A Adams and John JF Fournier. *Sobolev spaces*. Elsevier, 2003.

Jérémie Bigot, Elsa Cazelles, and Nicolas Papadakis. Central limit theorems for entropy-regularized optimal transport on finite spaces and statistical applications. *Electronic Journal of Statistics*, 13(2):5120 – 5150, 2019. doi: 10.1214/19-EJS1637. URL https://doi.org/10.1214/19-EJS1637.

Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. New York: Springer., 2010.

Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39:1853–1865, 2017.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pages 2292–2300, 2013. URL http://papers.nips.cc/paper/4927-sinkhorn-distances-lightspeed-computation-of-optimal-transport.

Lucas de Lara, Alberto González-Sanz, Nicholas Asher, and Jean-Michel Loubes. Transport-based counterfactual models. *Pre-Print*, 2021.

Lucas De Lara, Alberto González-Sanz, and Jean-Michel Loubes. Diffeomorphic registration using sinkhorn divergences. *arXiv preprint arXiv:2206.13948*, 2022.

Eustasio del Barrio and Jean-Michel Loubes. Central limit theorems for empirical transportation cost in general dimension. *The Annals of Probability*, 47(2):926 – 951, 2019. doi: 10.1214/18-AOP1275. URL https://doi.org/10.1214/18-AOP1275.

Eustasio del Barrio, Evarist Giné, and Carlos Matrán. Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *Ann. Probab.*, 27 (2):1009–1071, 1999. ISSN 0091-1798. doi: 10.1214/aop/1022677394. URL https://doi.org/10.1214/aop/1022677394.

Eustasio del Barrio, Evarist Giné, and Frederic Utzet. Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli*, 11(1):131–189, 2005. ISSN 1350-7265. doi: 10.3150/bj/1110228245. URL https://doi.org/10.3150/bj/1110228245.

Eustasio Del Barrio, Paula Gordaliza, and Jean-Michel Loubes. A central limit theorem for lp transportation cost on the real line with application to fairness assessment in machine learning. *Information and Inference: A Journal of the IMA*, 8(4):817–849, 2019.

Eustasio del Barrio, Alberto González-Sanz, and Jean-Michel Loubes. Central limit theorems for general transportation costs. *Pre-Print*, 2021a.

Eustasio del Barrio, Alberto González-Sanz, and Jean-Michel Loubes. A central limit theorem for semidiscrete wasserstein distances. *Pre-Print*, 2021b.

Eustasio del Barrio, Alberto Gonzalez-Sanz, Jean-Michel Loubes, and Jonathan Niles-Weed. An improved central limit theorem and fast convergence rates for entropic transportation costs. *arXiv preprint arXiv:2204.09105*, 2022.

Jean Feydy, Benjamin Charlier, François-Xavier Vialard, and Gabriel Peyré. Optimal transport for diffeomorphic registration. In *MICCAI*, 2017.

Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trouvé, and Gabriel Peyré. Interpolating between optimal transport and mmd using sinkhorn divergences. In *AISTATS*, 2019.

Rémi Flamary, Cédric Févotte, Nicolas Courty, and Valentin Emiya. Optimal spectral transportation with application to music transcription. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper/2016/file/352fe25daf686bdb4edca223c921acea-Paper.pdf.

N. Fournier and A. Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162:707–738, 2013.

Aude Genevay, Gabriel Peyré, and Marco Cuturi. Learning generative models with sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain*, pages 1608–1617, 2018. URL http://proceedings.mlr.press/v84/genevay18a.html.

Aude Genevay, Lénaïc Chizat, Francis Bach, Marco Cuturi, and Gabriel Peyré. Sample complexity of sinkhorn divergences. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pages 1574–1583. PMLR, 16–18 Apr 2019. URL http://proceedings.mlr.press/v89/genevay19a.html.

Eva Giné and Richard Nickl. Mathematical foundations of infinite-dimensional statistical models. In *Cambridge Series in Statistical and Probabilistic Mathematics, [40]. Cambridge University Press, New York.*, 2015.

Ziv Goldfeld, Kengo Kato, Gabriel Rioux, and Ritwik Sadhu. Statistical inference with regularized optimal transport. *arXiv: Probability*, 2022.

Javier González-Delgado, Alberto González-Sanz, Juan Cortés, and Pierre Neuvial. Two-sample goodness-of-fit tests on the flat torus based on wasserstein distance and their relevance to structural biology. 2021.

Alberto González-Sanz, Jean-Michel Loubes, and Jonathan Niles-Weed. Weak limits of entropy regularized optimal transport; potentials, plans and divergences. 2022.

Florian F. Gunsilius and Yuliang Xu. Matching for causal effects via multimarginal optimal transport. *arXiv: Probability*, 2021.

Zaïd Harchaoui, Lang Liu, and Soumik Pal. Asymptotics of entropy-regularized optimal transport via chaos decomposition. *arXiv: Probability*, 2020.

Shayan Hundrieser, Marcel Klatt, Thomas Staudt, and Axel Munk. A unifying approach to distributional limits for empirical optimal transport. *Pre-Print*, 02 2022. URL https://arxiv.org/pdf/2202.12790.pdf.

Marcel Klatt, Carla Tameling, and Axel Munk. Empirical regularized optimal transport: statistical theory and applications. *SIAM J. Math. Data Sci.*, 2(2):419–443, 2020. doi: 10.1137/19M1278788. URL https://doi.org/10.1137/19M1278788.

Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. Number 1 in Graduate studies in mathematics. Springer Berlin, Heidelberg, 1991.

Gonzalo Mena and Jonathan Niles-Weed. Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/5acdc9ca5d99ae66afdfe1eea0e3b26b-Paper.pdf.

Aram-Alexandre Pooladian and Jonathan Niles-Weed. Entropic estimation of optimal transport maps. 2021.

Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. In *SSVM*, 2011.

Alexander Shapiro. On concepts of directional differentiability. *Journal of Optimization Theory and Applications*, 66:477–487, 1990.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998. doi: 10.1017/CBO9780511802256.

Aad W. Van Der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer, New York, NY, 1996.

Cédric Villani. *Topics in optimal transportation*. Number 58 in Graduate studies in mathematics. American Mathematical Soc., 2003.

Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A):2620 – 2648, 2019. doi: 10.3150/18-BEJ1065. URL https://doi.org/10.3150/18-BEJ1065.

Part II

Center-outward distribution; regularity and quantile regression

CHAPTER 6

A note on the regularity of optimal-transport-based center-outward distribution and quantile functions

---

The content of this chapter is published in Journal of Multivariate analysis (2020) as del Barrio et al. (2020).

Contents

We provide sufficient conditions under which the center-outward distribution and quantile functions introduced in Chernozhukov et al. (2017) and Hallin (2017) are homeomorphisms, thereby extending a recent result by Figalli (2018). Our approach relies on Cafarelli's classical regularity theory for the solutions of the Monge-Ampère equation, but has to deal with difficulties related with the unboundedness at the origin of the density of the spherical uniform reference measure. Our conditions are satisfied by probabilities on Euclidean space with a general (bounded or unbounded) convex support which are not covered by Figalli. We also provide some additional results about center-outward distribution and quantile functions, including the fact that quantile sets exhibit a limiting form of the so-called "lighthouse convexity" property.

6.1 Introduction: center-outward distribution and quantile functions

Univariate distribution and quantiles functions, together with their empirical counterparts and the closely related concepts of ranks and order statistics, count among the most fundamental and useful tools in mathematical statistics. Ranks indeed are not just distribution-free: in models driven by noise with unspecified density, they generate the sub-σ-field of all distribution-free events (more precisely, the ranks generate an "essentially maximal ancillary σ-field": see Hallin et al. (2020a) for details), which is also the largest sub-σ-field independent, irrespective of the underlying distribution, of the minimal sufficient σ-field generated by the order statistic; suitable rank-based procedures achieve optimality in several senses in nonparametric testing as well (semi)parametric efficiency (see, e.g. Hájek et al. (1999); Hallin (1994); Hallin and Tribel (2000); Hallin and Werker (2003)). A major limitation of the classical concepts of ranks and quantiles, however, is that, due to the absence of a canonical ordering of $\mathbb{R}^d$ for $d \geq 2$, they do not readily extend to the multivariate context.

The problem is not new, and numerous attempts have been made to fill that gap by defining multivariate versions of distribution and quantiles functions, with the ultimate goal of constructing suitable mutivariate versions of classical rank- and quantile-based inference procedures. The traditional definition of a multivariate distribution function is somewhat helpless in that respect, and does not produce any satisfactory concept of quantiles—let alone a satisfactory concept of ranks (see Genest and Rivest (2001)). The componentwise approach, closely related with copula transforms, has been studied intensively (see Puri and Sen (1971)), but does not even enjoy distribution-freeness. Nor do the so-called "spatial ranks" (Oja (1999, 2010)) inspired by the L_1 characterization of univariate quantiles. The whole theory of statistical depth (see Serfling (2002, 2019) for authoritative surveys), in a sense, is motivated by the same objective of providing a (data-driven) ordering of $\mathbb{R}^d$ and adequate concepts of multivariate ranks (Zuo and He (2006)) and quantiles (He and Wang (1997)); here again, the resulting notions fail to be distribution-free. As for the "Mahalanobis ranks and signs" considered, e.g., in Hallin and Paindaveine (2002a,b, 2004); Hallin et al. (2010), they do enjoy distribution-freeness and all the desired properties expected from ranks—under the restrictive assumption, however, of elliptical symmetry.

This fundamental shortcoming of all available solutions has motivated the introduction in Chernozhukov et al. (2017) of the measure transportation-based concepts of Monge-Kantorovich depth, vector ranks, and quantiles (in a different context, one should mention the precursory role of the "comonotonic measures of multivariate risks" developed by Ekeland, Galichon, and Henry in 2012 Ekeland et al. (2012)) and the development, in Hallin (2017) and del Barrio et al. (2018), of center-outward distribution and quantile functions, ranks, and signs considered here. These center-outward concepts, unlike all previous ones, are shown (see del Barrio et al. (2018); Hallin et al. (2020a), where we refer to for more references and further discussion) to enjoy all the properties that make their univariate counterpart a fundamental and successful tool for statistical inference.

Let P be a Borel probability measure on the real line with finite second moment and continuous distribution function F and denote by $\mathrm{U}_{[0,1]}$ the uniform distribution on $(0,1)$: then F is a solution to "Monge's quadratic transportation problem", that is,

$$\int_{\mathbb{R}} |x - F(x)|^2 d\mathrm{P}(x) = \min_{T:\, T\sharp\mathrm{P}=\mathrm{U}_{[0,1]}} \int_{\mathbb{R}} |x - T(x)|^2 d\mathrm{P}(x)$$

(see, e.g., Villani (2003)), where $T\sharp P$ denotes the "push forward of P by T"—namely, the distribution of $T(X)$ under $X \sim \mathrm{P}$, with T a measurable map from $\mathbb{R}$ to $(0,1)$. With generalization to higher dimension in mind, however, del Barrio et al. (2018); Hallin (2017); Hallin et al. (2020a) rather consider $F_{\pm}(x) = 2F(x) - 1$, the so-called "center-outward distribution function" of P, satisfying the transportation problem

$$\int_{\mathbb{R}} |x - F_{\pm}(x)|^2 d\mathrm{P}(x) = \min_{T:\, T\sharp\mathrm{P}=\mathrm{U}_1} \int_{\mathbb{R}} |x - T(x)|^2 d\mathrm{P}(x),$$

where U_1 is the uniform distribution over $(-1, 1)$, the one-dimensional unit ball $\mathbb{B}_1$. Clearly, $F_{\pm}$ and F carry the same information about P.

The latter definition, indeed, readily extends to arbitrary dimensions. Let P denote a Borel probability measure on $\mathbb{R}^d$ with finite second-order moments and Lebesgue density p. Measure transportation theory (see, e.g., Theorem 2.12 in Villani (2003)) tells us that there exists a P-a.s. unique map $\mathbf{F}_{\pm}$ such that

$$\int_{\mathbb{R}^d} |\mathbf{x} - \mathbf{F}_{\pm}(\mathbf{x})|^2 d\mathrm{P}(\mathbf{x}) = \min_{\mathbf{T}:\, \mathbf{T}\sharp\mathrm{P}=\mathrm{U}_d} \int_{\mathbb{R}^d} |\mathbf{x} - \mathbf{T}(\mathbf{x})|^2 d\mathrm{P}(\mathbf{x}), \tag{6.1}$$

where $|\mathbf{x}|$ stands for the Euclidean norm of $\mathbf{x}$ and U_d denotes the "uniform" distribution over the open d-dimensional unit ball $\mathbb{B}_d$. The center-outward distribution function is defined as a solution $\mathbf{F}_{\pm}$ of this optimal transportation problem.

By uniform over $\mathbb{B}_d$ we refer to "spherical" uniformity, that is, U_d here corresponds to the uniform choice of a direction on the unit sphere $\mathbb{S}_{d-1} := \bar{\mathbb{B}}_d - \mathbb{B}_d$ in $\mathbb{R}^d$ combined with an independent uniform choice in $(0, 1)$ of a distance to the origin ($\bar{\mathbb{B}}_d$ denotes the closed unit ball in $\mathbb{R}^d$) . A simple change of variable shows that U_d has density

$$u_d(\mathbf{x}) = \frac{1}{a_d|\mathbf{x}|^{d-1}} I\big[\mathbf{x} \in \mathbb{B}_d \setminus \{\mathbf{0}\}\big], \tag{6.2}$$

where $a_d = 2\pi^{d/2}/\Gamma(d/2)$ denotes the area (the $(d-1)$-dimensional Hausdorff measure, see, e.g., Evans and Gariepy (1992)) of the sphere $\mathbb{S}_{d-1}$. Note the singularity at the origin since $\mathbf{x} \mapsto 1/|\mathbf{x}|^{d-1}$ is infinite at $\mathbf{x} = \mathbf{0}$; while we safely can neglect $\mathbf{0}$ itself, which has measure zero, by putting $u_d(\mathbf{0}) = 0$, u_d nevertheless remains unbounded in the vicinity of $\mathbf{0}$.

This definition of the center-outward distribution function as the solution of a quadratic transportation problem suffers from two major limitations. First, finite second-order moments are needed in order for the optimization problem (6.1) to make sense. Second, the distribution function $\mathbf{F}_{\pm}$ based on (6.1) is only defined P-a.s.; this means, for instance, that $\mathbf{F}_{\pm}$ is not well defined outside the support of P.

The first of these two limitations has been relaxed in Hallin (2017) thanks to a celebrated theorem by McCann McCann (1995). Under the assumption that P has finite second-order moments, Brenier in 1991 had shown that optimal transportation maps (hence, all versions of the P-a.s. unique solution $\mathbf{F}_{\pm}$ of Monge's problem (6.1)) coincide P-a.s. with the Lebesgue-a.e. gradient $\nabla\varphi$ of a convex function φ, which has the interpretation of a "potential" (the notation $\nabla\varphi$ here is used for the Lebesgue-a.e. gradient of φ, that is, $\nabla\varphi(\mathbf{x})$ is defined as the gradient at $\mathbf{x}$ of φ whenever φ is differentiable at $\mathbf{x}$—which, for a convex φ, holds

Lebesgue-a.e.; note that, contrary to $\nabla\varphi$, which is a.e. unique, φ is not—unless we impose, without loss of generality, that $\varphi(\mathbf{0}) = 0$, see, e.g., Lemma 2.1 in del Barrio and Loubes (2019)). More precisely, $\mathbf{F}_\pm$ a.s. is of the form $\nabla\varphi$ where φ (i) is lower semicontinous (lsc in the sequel), (ii) is convex, and (iii) is such that $\nabla\varphi\sharp\mathrm{P} = \mathrm{U}_d$. McCann McCann (1995) further showed that these last three conditions uniquely determine $\nabla\varphi$, even in the absence of second moment assumptions, while under finite second-order moments, $\nabla\varphi$ is a solution of Monge's problem (6.1). Thus, putting

$$\mathbf{F}_\pm(\mathbf{x}) := \nabla\varphi(\mathbf{x}) \quad \mathbf{x}\text{-a.e. in } \mathbb{R}^d, \tag{6.3}$$

the center-outward distribution function $\mathbf{F}_\pm$ is no longer characterized as the almost surely unique solution of an optimization problem (6.1) requiring finite moments of order two but as the unique a.e. gradient $\nabla\varphi$ of a convex function pushing P forward to U_d. We nevertheless conform to the common usage of improperly calling $\nabla\varphi$ the optimal transport pushing P forward to U_d.

While taking care of the moment assumption—existence of second-order moments indeed is an embarrassing assumption when distribution and quantile functions are to be defined—the second limitation still remains. The non-uniqueness of $\mathbf{F}_\pm := \nabla\varphi$, however, disappears if P is such that φ is everywhere differentiable. That this is indeed the case was shown by Figalli in 2018 Figalli (2018) for P in the so-called class of distributions with "nonvanishing densities" (precisely, the distributions P with densities p and support $\mathcal{X} = \mathbb{R}^d$ satisfying Assumption A below). For any P in that class of distributions, Figalli actually establishes that $\nabla\varphi(\mathbf{x})$ is a gradient for all $\mathbf{x}$ and, when restricted to

$$\mathbb{R}^d_{(\mathbf{0})} := \mathbb{R}^d \setminus \{\mathbf{x} : \nabla\varphi(\mathbf{x}) \neq \mathbf{0}\},$$

a homeomorphism between $\mathbb{R}^d_{(\mathbf{0})}$ and the punctured ball $\mathbb{B}_d \setminus \{\mathbf{0}\}$. The latter property is quite essential if sensible quantile regions and contours—namely, closed, continuous, connected, and nested—based on an inverse $\mathbf{Q}_\pm$ of $\mathbf{F}_\pm$ (see Section 2.1 for a precise definition), are to be defined: see Hallin et al. (2020a) for details.

The introduction by Hallin Hallin (2017) of center-outward ranks and quantiles rapidly attracted the attention of the nonparametric community. It has triggered, among others, Faugeras and Rüschendorf (Faugeras and Rüschendorf (2017); Faugeras and Rüschendorf (2019)) and de Valk and Segers de Valk and Segers (2018). Applications to the long-standing open problem of constructing distribution-free tests for the hypothesis of independence between vectors with unspecified densities have been proposed by Deb and Sen Deb and Sen (2019), Shi, Drton, and Han Shi et al. (2019), Shi, Hallin, Drton, and Han Shi et al. (2020), and Ghosal and Sen Ghosal and Sen (2019). Fully distribution-free center-outward rank tests for multiple-output regression and MANOVA are constructed in Hallin et al. (2020b). Optimal center-outward R-estimators also have been derived (Hallin, La Vecchia, and Liu Hallin et al. (2019)) for VARMA models, while center-outward quantile-based methods for the measurement of multivariate risk are proposed in del Barrio, Beirlant, Buitendag, and Hallin del Barrio et al. (2019).

The goal of the present paper is to provide simple sufficient conditions for Figalli's result to hold beyond the assumption of nonvanishing densities; we more particularly consider distributions with convex (bounded or unbounded) supports. Beyond other theoretical considerations, the homeomorphic nature of $\mathbf{F}_\pm$ and $\mathbf{Q}_\pm$ are the key properties used to prove a.s. convergence of the empirical center-outward distribution functions to their theoretical counterparts in del Barrio et al. (2018) and Hallin et al. (2020a). Hence, the results of the present paper also are extending the validity of the center-outward Glivenko-Cantelli theorem in these references—a theorem that also serves as a basis for the Hàjek-type asymptotic representation results in Hallin et al. (2020b).

From a technical point of view, our main result is Theorem 6.2.5 below, which relies on the classical regularity theory for solutions of Monge-Ampère equations associated with the name of Caffarelli (see Caffarelli (1990, 1991, 1992)), as discussed in Section 6.2. The use of that theory to investigate the regularity of optimal transportation maps between two probabilities typically requires that both probabilities have densities that are bounded and bounded away from zero over their respective supports. Recently, under local versions of this condition, very general regularity results of this kind have been given in Cordero-Erausquin and Figalli (2019) and Ghosal and Sen (2019). However, the spherical uniform reference measure U_d considered here, in dimension $d \geq 2$, yields unbounded densities at the origin, so that the results in Cordero-Erausquin and Figalli (2019); Ghosal and Sen (2019) do not apply. Note that the choice of the spherical uniform reference is not a whimsical one. It preserves the independence between $\|\mathbf{F}_\pm\|$ and $\mathbf{F}_\pm/\|\mathbf{F}_\pm\|$ (extending the independence, for $d = 1$, between $|F_\pm|$ and $\text{sign}(F_\pm)$) and produces simple and easily interpretable quantile contours with prescribed probability content (we refer to Hallin et al. (2020a) for details; see also the remarks after Theorem 6.2.5 below). To our knowledge, the only reference dealing with this kind of unbounded density is Figalli (2018) which, however, requires P to be supported on the whole space $\mathbb{R}^d$. Here we extend the result in Figalli (2018) to cover the case of P with (bounded or unbounded) convex supports.

The sequel of this paper is organized as follows. Our main regularity result is established in Section 2, along with a succinct account of the main elements of Cafarelli's theory and some auxiliary results. We conclude with Section 3, which presents some new results on center-outward distribution and quantile functions. These include an asymptotic invariance property extending a well-known feature of classical univariate distribution functions and the ability of quantile contours to capture the shape of the bounded support of a probability measure by converging (in Hausdorff distance) to the boundary of the support. Finally, we include a result on the geometry of quantile sets, showing that they turn out to exhibit a limiting form of "lighthouse convexity".

6.2 Regularity of center-outward distribution and quantile functions

6.2.1 Center-outward quantile functions

The Introduction was focused on the distribution functions $\mathbf{F}_\pm$. Exchanging the roles of P and U_d, we could have emphasized transportation from the unit ball to the support of P, leading to the definition of the center-outward quantile function $\mathbf{Q}_\pm$ with, mutatis mutandis, the same comments.

Let P denote a Borel probability measure over $\mathbb{R}^d$ with Lebesgue density p. While the center-outward distribution function is defined as the optimal (in the McCann sense) transport pushing P forward to U_d, the center-outward quantile map or quantile function $\mathbf{Q}_\pm$ of P is defined as the optimal transport pushing U_d forward to P. Namely,

$$\mathbf{Q}_\pm(\mathbf{u}) := \nabla\psi(\mathbf{u}) \quad \mathbf{u}\text{-a.e. in } \mathbb{B}_d \tag{6.4}$$

where $\nabla\psi$ is, in agreement with McCann's Theorem, the unique a.e. gradient of a convex function ψ with domain containing $\mathbb{B}_d$ such that $\nabla\psi\sharp\mathrm{U}_d = \mathrm{P}$ (we adhere to the usual convention of considering that a function defined on a subset $A \subset \mathbb{R}^d$ is convex if it can be extended to a convex function on $\mathbb{R}^d$ with values in $\mathbb{R} \cup \{\infty\}$; the domain of the convex function is then redefined as the set where it takes finite values). Again, imposing, without loss of generality, that $\psi(\mathbf{0}) = 0$, the convex potential ψ is uniquely defined (indeed, two convex functions with a.e. equal gradients on an open convex set are equal up to an additive constant, see, e.g., Lemma 2.1 in del Barrio and Loubes (2019)) and a.e. differentiable over $\mathbb{B}_d$. We extend ψ to a lsc convex function on $\mathbb{R}^d$ with the standard procedure of setting $\psi(\mathbf{u}) := \liminf_{\mathbf{z}\to\mathbf{u},|\psi(\mathbf{z})|<1} \psi(\mathbf{z})$ if $|\mathbf{u}| = 1$ and $\psi(\mathbf{u}) := +\infty$ for $\mathbf{u} \notin \bar{\mathbb{B}}_d$ (see, e.g. (A.18) in Figalli (2017)). With this extension, φ is the Legendre transform of ψ, that is,

$$\varphi(\mathbf{x}) = \psi^*(\mathbf{x}) := \sup_{\mathbf{u}\in\mathbb{B}_d} (\langle \mathbf{u}, \mathbf{x}\rangle - \psi(\mathbf{u})), \quad \mathbf{x} \in \mathbb{R}^d. \tag{6.5}$$

We observe that the domain of φ is $\mathbb{R}^d$ and that φ, being the sup of a 1-Lipschitz function, is also 1-Lipschitz. In particular, for almost every $\mathbf{x} \in \mathbb{R}^d$, φ is differentiable with $|\nabla\varphi(\mathbf{x})| \leq 1$ and, as a consequence (see, e.g., Corollary A.27 in Figalli (2017)),

$$\partial\varphi(\mathbb{R}^d) \subseteq \bar{\mathbb{B}}_d; \tag{6.6}$$

here, and throughout this paper, $\bar{B}$ stands for the closure of a set B, $\partial\varphi(\mathbf{x})$ for the subdifferential of the convex function φ at $\mathbf{x}$ (recall that the subdifferential of φ at $\mathbf{x}$ is the set of all $\mathbf{z} \in \mathbb{R}^d$ such that $\varphi(\mathbf{y}) - \varphi(\mathbf{x}) \geq \langle \mathbf{z}, \mathbf{y} - \mathbf{x}\rangle$ for all $\mathbf{y}$), and $\partial\varphi(A) := \bigcup_{\mathbf{x}\in A} \partial\varphi(\mathbf{x})$. Furthermore, Proposition 10 in McCann (1995) (see also Remark 16) shows that, since P has a density, necessarily $\nabla\psi(\nabla\varphi(\mathbf{x})) = \mathbf{x}$ for almost every $\mathbf{x}$ in the support of P and $\nabla\varphi(\nabla\psi(\mathbf{y})) = \mathbf{y}$ for almost every $\mathbf{y} \in \mathbb{B}_d$. In that sense, $\mathbf{Q}_\pm$ and $\mathbf{F}_\pm$ are the inverse of each other. In this way, we have defined $\mathbf{F}_\pm(\mathbf{x})$ for almost every $\mathbf{x} \in \mathbb{R}^d$ (through equation (6.3)) and $\mathbf{Q}_\pm(\mathbf{u})$ for almost every $\mathbf{u} \in \mathbb{B}_d$; the definitions coincide with those in Chernozhukov et al. (2017); Hallin (2017) for $\mathbf{x}$ in the support of P.

As an example of a distribution with support strictly smaller than $\mathbb{R}^d$, consider P with constant density on the unit ball $\mathbb{B}_d$. A simple computation shows that the convex map

$$\psi(\mathbf{u}) = \frac{d}{d+1}|\mathbf{u}|^{\frac{d+1}{d}} \text{ for } |\mathbf{u}| \leq 1, \qquad \psi(\mathbf{u}) = +\infty \text{ for } |\mathbf{u}| > 1$$

has gradient $\nabla\psi(\mathbf{u}) = |\mathbf{u}|^{\frac{1}{d}-1}\mathbf{u}$, $0 < |\mathbf{u}| < 1$, which maps U_d to P. Hence,

$$\mathbf{Q}_\pm(\mathbf{u}) = |\mathbf{u}|^{\frac{1}{d}-1}\mathbf{u}, \quad 0 < |\mathbf{u}| < 1.$$

Elementary computations yield

$$\varphi(\mathbf{x}) = \frac{1}{d+1}|\mathbf{x}|^{d+1}, \ |\mathbf{x}| \leq 1, \qquad \varphi(\mathbf{x}) = |\mathbf{x}| - \frac{d}{d+1}, \ |\mathbf{x}| > 1,$$

as the Legendre transform of ψ, with gradient

$$\mathbf{F}_\pm(\mathbf{x}) = |\mathbf{x}|^{d-1}, \ |\mathbf{x}| \leq 1, \qquad \mathbf{F}_\pm(\mathbf{x}) = \frac{\mathbf{x}}{|\mathbf{x}|}, \ |\mathbf{x}| < 1.$$

Of course, $\mathbf{F}_\pm$ maps P to U_d and this, together with the fact that $\mathbf{F}_\pm$ is the gradient of a convex function, uniquely determines $\mathbf{F}_\pm$ in $\mathbb{B}_d$. Our extension guarantees that $\mathbf{F}_\pm(\mathbf{x}) \in \bar{\mathbb{B}}_d$ for every $\mathbf{x} \in \mathbb{R}^d$ and coincides with its univariate analogue, also for probabilities with bounded support.

6.2.2 Some regularity results for Monge-Ampère equations

As announced in the Introduction, our approach to the regularity of the center-outward distribution and quantile functions is based on the classical regularity theory for Monge-Ampère equations. We refer to Figalli (2017) for a comprehensive account of this theory, of which we present here a minimal account.

Given an open set $\mathcal{X} \subseteq \mathbb{R}^d$ and a convex function $\varphi : \mathcal{X} \to \mathbb{R}$, denoting by ℓ_d he Lebesgue measure on $\mathbb{R}^d$, the Monge-Ampère measure associated with φ is defined by

$$\mu_\varphi(E) := \ell_d\Big(\partial\varphi(E)\Big)$$

for every Borel set $E \subseteq \mathbb{R}^d$. It can be checked that μ_φ is indeed a locally finite Borel measure on $\mathcal{X}$. The crucial link between center-outward distribution functions and Monge-Ampère measures can be summarized as follows. Assume P is a probability on $\mathcal{X}$ with Lebesgue density p and let φ be a convex function from $\mathcal{X}$ to $\mathbb{R}$. Then, for every Borel set A,

$$\mathrm{Q}(A) := (\nabla\varphi\sharp\mathrm{P})(A) = \mathrm{P}\big(\partial\varphi^*(A)\big)$$

where φ^* is the Legendre transform of φ. We recall that convexity of φ implies that it is differentiable at almost every point in $\mathcal{X}$ (see, e.g., Theorem 25.4 in Rockafellar (1970)) and, therefore,

$$(\nabla\varphi\sharp\mathrm{P})(A) = P(\{\mathbf{x} : \nabla\varphi(\mathbf{x}) \in A\}) = P(\{\mathbf{x} : \partial\varphi(\mathbf{x}) \subseteq A\}).$$

This and the fact that $\mathbf{y} \in \partial\varphi(\mathbf{x})$ if and only if $\mathbf{x} \in \partial\varphi^*(\mathbf{y})$ yield the last equality above. Hence, if Q has a density q, for every Borel set A,

$$\int_{\partial\varphi(A)} q(\mathbf{y})d\mathbf{y} = \int_A p(\mathbf{x})d\mathbf{x}$$

(see Lemma 4.6 in Villani (2003)); if, moreover, $\mathrm{Q} = \mathrm{U}_d$,

$$\int_{\partial\varphi(A)} u_d(\mathbf{y})d\mathbf{y} = \int_A p(\mathbf{x})d\mathbf{x}. \tag{6.7}$$

Observing that

$$\mu_\varphi(A) = \ell_d(\partial\varphi(A)) = \ell_d(\partial\varphi(A) \cap \mathbb{B}_d),$$

where the second equality follows from (6.6), we obtain from (6.7) that, for A such that $\ell_d(A) = 0$,

$$\mu_\varphi(A) \le a_d \int_{\partial\varphi(A)} u_d(\mathbf{y})d\mathbf{y} = a_d \int_A p(\mathbf{x})d\mathbf{x} = 0$$

with a_d as in (6.2). Thus, the Monge-Ampère measure μ_φ is Lebesgue-absolutely continuous. Since the density of the absolutely continuous part of the Monge-Ampère measure μ_φ is given by $\big(p(\mathbf{x})/u_d(\nabla\varphi(\mathbf{x}))\big)$ (see McCann McCann (1997) or Theorem 4.8 in Villani (2003)), we conclude that, for every Borel set $A \subseteq \mathbb{R}^d$,

$$\mu_\varphi(A) = \int_A \frac{p(\mathbf{x})}{u_d(\nabla\varphi(\mathbf{x}))}d\mathbf{x} = a_d \int_A p(\mathbf{x})|\nabla\varphi(\mathbf{x})|^{d-1}d\mathbf{x}. \tag{6.8}$$

Let us focus now on the Monge-Ampère measure μ_ψ associated (see Section 6.2.1) with $\mathbf{Q}_\pm$ and ψ (both defined over $\mathbb{B}_d$). Since $\nabla\psi$ pushes U_d forward to P, we have that $\nabla\psi(\mathbf{y}) \in \mathcal{X}$ $\mathbf{y}$-a.e. in $\mathbb{B}_d$. By continuity (see Theorem 25.5 in Rockafellar (1970)), $\nabla\psi(\mathbf{y}) \in \bar{\mathcal{X}}$ for every point $\mathbf{y}$ of differentiability of ψ. Using again Corollary A.27 in Figalli (2017), we conclude that $\partial\psi(\mathbb{B}_d)$ is included in the convex hull $\overline{\mathrm{conv}}(\mathcal{X})$ of $\mathcal{X}$. Hence, if $\mathcal{X}$ itself is convex, we obtain that

$$\partial\psi(\mathbb{B}_d) \subseteq \overline{\mathcal{X}}. \tag{6.9}$$

Analogous to (6.7), we have that

$$\int_{\partial\psi(B)} p(\mathbf{x})d\mathbf{x} = \int_B u_d(\mathbf{y})d\mathbf{y} \ \text{ for every Borel set } B \subseteq \mathbb{R}^d. \tag{6.10}$$

Now, denoting by $r\,\mathbb{B}_d$ the open ball with radius r centered at the origin, let us assume that the Borel set $B \subseteq r\,\mathbb{B}_d$, with $0 < r < 1$, has Lebesgue measure zero. Since $\bar{B} \subseteq r\,\mathbb{B}_d$ is compact, $\partial\psi(\bar{B})$ also is compact (see, e.g. Lemma A.22 in Figalli (2017)). Hence, there exists $R > 0$ such that

$$\partial\psi(B) \subseteq \partial\psi(\bar{B}) \subseteq R\,\mathbb{B}_d.$$

The following assumption, which requires the density p of P to be bounded and bounded away from 0 on compact subsets of the support, is absolutely essential (the same assumption is also made by Figalli in Figalli (2018)).

Assumption A. For every $R > 0$, there exist constants $0 < \lambda_R \leq \Lambda_R$ such that

$$\lambda_R \leq p(\mathbf{x}) \leq \Lambda_R \quad \text{for all } \mathbf{x} \in \mathcal{X} \cap R\,\mathbb{B}_d. \tag{6.11}$$

Since $\mathcal{X}$ is convex (hence $\ell_d(\bar{\mathcal{X}} - \mathcal{X}) = 0$), Assumption A entails

$$\mu_\psi(B) \leq \frac{1}{\lambda_R} \int_{\partial\psi(B)} p(\mathbf{x})d\mathbf{x} = \frac{1}{\lambda_R} \int_B u_d(\mathbf{y})d\mathbf{y} = 0.$$

Assuming convexity of $\mathcal{X}$ and (6.11), we conclude that μ_ψ is absolutely continuous with respect to ℓ_d and, using Theorem 4.8 in Villani (2003) again, that, for every Borel set $B \subset \mathbb{B}_d$,

$$\mu_\psi(B) = \int_B \frac{u_d(\mathbf{y})}{p(\nabla\psi(\mathbf{y}))} d\mathbf{y} = \frac{1}{a_d} \int_B \frac{1}{p(\nabla\psi(\mathbf{y}))|\mathbf{y}|^{d-1}} d\mathbf{y}. \tag{6.12}$$

We summarize this discussion in the next proposition.

Proposition 6.2.1. *Let* P *be a probability measure with density* p *supported on the open set* $\mathcal{X} \subseteq \mathbb{R}^d$*. Denoting by* ψ *the convex, lower semicontinuous function from* $\mathbb{B}_d$ *to* $\mathbb{R}$ *satisfying* $\psi(\mathbf{0}) = 0$ *and* $\nabla\psi\sharp\mathrm{U}_d = \mathrm{P}$*, let* $\varphi : \mathbb{R}^d \to \mathbb{R}$ *be defined as in* (6.5)*. Then,*

(i) μ_φ *is absolutely continuous with respect to* ℓ_d *and, for every Borel* $A \subseteq \mathbb{R}^d$*,*

$$\mu_\varphi(A) = a_d \int_A p(\mathbf{x})|\nabla\varphi(\mathbf{x})|^{d-1} d\mathbf{x};$$

(ii) *if, moreover,* $\mathcal{X}$ *is convex and* p *satisfies Assumption A, then* μ_ψ *is absolutely continuous with respect to* ℓ_d *and, for every Borel set* $B \subseteq \mathbb{B}_d$*,*

$$\mu_\psi(B) = \frac{1}{a_d} \int_B \frac{1}{p(\nabla\psi(\mathbf{y}))|\mathbf{y}|^{d-1}} d\mathbf{y}.$$

Next, let us show that, for well-behaved probability measures P (those with convex support and density p satisfying Assumption A), the center-outward distribution function $\mathbf{F}_\pm$ cannot map points in the interior of the support of P to extremal points of the unit ball.

Lemma 6.2.2. *Let* P *be a probability measure with density* p *supported on the convex open set* $\mathcal{X} \subseteq \mathbb{R}^d$ *and such that Assumption A holds. Then* $(\partial\varphi)(\mathcal{X}) \cap \mathbb{S}_{d-1} = \emptyset$*, where* $\mathbb{S}_{d-1} = \bar{\mathbb{B}}_d \setminus \mathbb{B}_d$*.*

Proof. Assume that there exists $x \in \mathcal{X}$ such that $|\mathbf{y}| = 1$ for some $\mathbf{y} \in \partial\varphi(\mathbf{x})$. Without loss of generality, we can assume $\mathbf{x} = \mathbf{0}$. Since $\mathcal{X}$ is open, there exists $\epsilon > 0$ such that $\epsilon\overline{\mathbb{B}}_d \subset \mathcal{X}$. For small $\theta > 0$, consider the sets

$$\mathcal{C}_{\epsilon,\vartheta} := \Big\{\mathbf{x} \in \mathbb{R}^d : \Big|\frac{\mathbf{x}}{|\mathbf{x}|} - \mathbf{y}\Big| \leq \sin\vartheta, |\mathbf{x}| \leq \epsilon\Big\}$$
$$\mathcal{D}_{\vartheta} := \{\mathbf{b} \in \mathbb{B}_d : \ \langle \mathbf{y} - \mathbf{b}, \mathbf{y}\rangle \leq 2\vartheta|\mathbf{y} - \mathbf{b}|\}.$$

Now, if $\mathbf{a} \in \mathcal{C}_{\epsilon,\vartheta}$ and $\mathbf{b} \in \partial\varphi(\mathbf{a})$, the monotonicity of $\partial\varphi$ implies that $\langle \mathbf{y} - \mathbf{b}, \mathbf{a}\rangle \leq 0$. Hence,

$$\langle \mathbf{y} - \mathbf{b}, \mathbf{y}\rangle = \langle \mathbf{y} - \mathbf{b}, \mathbf{y} - \frac{\mathbf{a}}{|\mathbf{a}|}\rangle + \langle \mathbf{y} - \mathbf{b}, \frac{\mathbf{a}}{|\mathbf{a}|}\rangle \leq |\mathbf{y} - \mathbf{b}|\sin(\vartheta) \leq |\mathbf{y} - \mathbf{b}|2\vartheta.$$

This shows that $\partial\varphi(\mathcal{C}_{\epsilon,\vartheta}) \subseteq \mathcal{D}_{\vartheta}$. But the density p, inside $\mathcal{C}_{\epsilon,\vartheta}$, is bounded from below by λ_ϵ and the density u_d is bounded from above by $2/a_d$ inside $\mathcal{D}_\vartheta$ for $\vartheta \ll 1$: then, in view of the transport equation (6.7), we have

$$\frac{2}{a_d}\ell_d\big(\mathcal{D}_\vartheta\big) \geq \int_{\mathcal{D}_\vartheta} u_d(\mathbf{b})d\mathbf{b} \geq \int_{\partial\varphi(\mathcal{C}_{\epsilon,\vartheta})} u_d(\mathbf{b})d\mathbf{b} = \int_{\mathcal{C}_{\epsilon,\vartheta}} p(\mathbf{x})d\mathbf{x} \geq \lambda_\epsilon \ell_d\big(\mathcal{C}_{\epsilon,\vartheta}\big).$$

This, however, cannot hold true since $\ell_d\big(\mathcal{C}_{\epsilon,\vartheta}\big) \approx \epsilon^d\vartheta^{d-1}$ and $\ell_d\big(\mathcal{D}_\vartheta\big) \approx \vartheta^{d+1}$ as $\theta \to 0$. The claim follows. □

We now proceed to provide sufficient conditions under which the center-outward quantile function $\mathbf{Q}_\pm$ is continuous at every point in the open unit ball (except, possibly, at the origin). It is well known that differentiability of a lower semicontinuous convex function ψ (which entails continuity of its gradient) is equivalent to strict convexity of its convex conjugate (see Theorem 26.3 in Rockafellar (1970)). As announced, the techniques we are using here are in the spirit of those developed by Caffarelli in Caffarelli (1990), Caffarelli (1991) or Figalli in Figalli (2017), Figalli (2018), which in turn largely rely on the fact that, under some control for the Monge-Ampère measure, the intersection between the graph and supporting hyperplanes of ψ either consists of a single point or has an extreme point (see Theorem 4.10 in Figalli (2017)). A central result in Caffarelli's regularity theory (see Corollary 4.21 in Figalli (2017)) is that a strictly convex function ψ on an open set Ω for which there exist constants $0 < \lambda < \Lambda$ such that

$$\lambda\ell_d(A) \leq \mu_\psi(A) \leq \Lambda\ell_d(A) \tag{6.13}$$

for every Borel set $A \subseteq \Omega$ is automatically of class $\mathcal{C}^{1,\alpha}_{\text{loc}}$ for some $\alpha > 0$ that depends only on λ, Λ, and d (condition (6.13) in the sequel is summarized, with a slight abuse of notation, as $\lambda d\mathbf{x} \leq \mu_\psi \leq \Lambda d\mathbf{x}$). The fact that U_d for $d \geq 2$ has an unbounded density adds some complication to the particular problem here, though. On the other hand, the density u_d is bounded away from 0, which allows us to control the growth of the Monge-Ampère measure, as we show next.

Lemma 6.2.3. *Let* P *satisfy the assumptions in Proposition 6.2.1(ii). Denoting by* M *a compact subset of* $\mathbb{B}_d$*, there exist constants* α_M *and* A_M *such that, for every Borel set* $A \subseteq M$*,*

$$\alpha_M \ell_d(A) \leq \mu_\psi(A) \leq A_M(\ell_d(A))^{1/d}. \tag{6.14}$$

Proof. The compactness of M entails that of $\partial\psi(M)$; in particular, $\partial\psi(M) \subseteq R\,\mathbb{B}_d$ for some $R > 0$. Hence, using Proposition 6.2.1 (ii) and taking $\lambda_R, \Lambda_R \in \mathbb{R}$ as in Assumption A, we obtain

$$\mu_\psi(A) = a_d \int_A \frac{1}{p(\nabla\psi(\mathbf{y}))|\mathbf{y}|^{d-1}} d\mathbf{y} \geq \frac{a_d}{\Lambda_R}\ell_d(A).$$

For the upper bound in (6.14), note that the ball $(\ell_d(A)/c_d)^{1/d}B_d$ (where $c_d = \pi^{d/2}/\Gamma(1 + d/2)$ denotes the volume of the d-dimensional unit ball) maximizes $\int_B |\mathbf{y}|^{1-d}d\mathbf{y}$ among all subsets of $\mathbb{B}_d$ with Lebesgue measure $\ell_d(A)$. On the other hand, by the co-area formula (see, e.g., Proposition 1, page 118 in Evans and Gariepy (1992)),

$$\int_{r\mathbb{B}_d} |\mathbf{y}|^{1-d}d\mathbf{y} = \int_0^r \left[\int_{\partial s\mathbb{B}_d} |\mathbf{y}|^{1-d}d\mathcal{H}^{d-1}(\mathbf{y})\right] ds = \int_0^r a_d\, ds = a_d\, r \tag{6.15}$$

where $\mathcal{H}^{d-1}$ denotes the $(d-1)$-dimensional Hausdorff measure. Combining (6.15) with Proposition 6.2.1 (ii), we conclude that

$$\mu_\psi(A) \leq \frac{1}{\lambda_R\, a_d}\int_A |\mathbf{y}|^{1-d}d\mathbf{y} \leq \frac{1}{\lambda_R\, a_d^{1/d}}(\ell_d(A))^{1/d}.$$

□

Note that the lower bound in Lemma 6.2.3 remains valid for a compact subset M of $\bar{\mathbb{B}}_d$ provided that $\partial\psi(M)$ is bounded: indeed, that lower bound only requires the upper bound from Assumption A. A similar conclusion holds for the upper bound. Additionally, if the density p is uniformly bounded, the lower bound holds for any subset of $\bar{\mathbb{B}}_d$.

6.2.3 Main result

We are ready now for the main result of this note. Our proof follows the lines of Figalli (2017, 2018); Cordero-Erausquin and Figalli (2019), and the related Proposition 3.3 in Ghosal and Sen (2019) but, unlike Figalli (2018), we cover cases in which the support of the probability measure P is is not the whole space $\mathbb{R}^d$. Similar to Cordero-Erausquin and Figalli (2019) and Ghosal and Sen (2019), we have to handle carefully the fact that $\mathcal{X}$ is not necessarily bounded and use a "minimal" extension of the quantile function potential, namely,

$$\widetilde{\psi}(\mathbf{z}) := \sup_{\mathbf{b}\in\mathbb{B}_d,\ \mathbf{y}\in\partial\psi(\mathbf{b})} \{\langle \mathbf{y}, \mathbf{z}-\mathbf{b}\rangle + \psi(\mathbf{b})\},\ \ \mathbf{z}\in\mathbb{R}^d. \tag{6.16}$$

Obviously, $\widetilde{\psi}$ is still a lower semicontinuous convex function and $\widetilde{\psi}(\mathbf{z})$ coincides with $\psi(\mathbf{z})$ for $\mathbf{z}\in\mathbb{B}_d$. Since we have that $\mathbf{Q}_{\pm}(\mathbf{z}) := \nabla\psi(\mathbf{z}) \in \mathcal{X}$ for every differentiability point $\mathbf{z}$ of ψ in $\mathbb{B}_d$, we see (using, once more, Corollary A.27 in Figalli (2017)) that, provided that $\mathcal{X}$ is convex, $\partial\psi(\mathbb{B}_d) \subseteq \bar{\mathcal{X}}$. The "minimality" of the extension (6.16) refers to the fact that $\partial\widetilde{\psi}(\mathbb{R}^d) \subset \bar{\mathcal{X}}$, as can be checked from a simple application of the Hahn-Banach separation theorem. Of course, the values of $\widetilde{\psi}$ outside $\mathbb{B}_d$ are not relevant for the study of its differentiability inside $\mathbb{B}_d$, but the use of $\widetilde{\psi}$ will be useful in the next proof. We note also that the discussion leading to Proposition 6.2.1 can be reproduced with $\widetilde{\psi}$ substituted for ψ to conclude that $\mu_{\widetilde{\psi}}$ is absolutely continuous with respect to the Lebesgue measure and that, for every Borel set $B\subseteq\mathbb{R}^d$,

$$\mu_{\widetilde{\psi}}(B) = \int_{B\cap\mathcal{X}} \frac{u_d(\mathbf{y})}{p(\nabla\widetilde{\psi}(\mathbf{y}))} d\mathbf{y}. \tag{6.17}$$

Finally, observe that $\mu_{\widetilde{\psi}}$ in concentrated on $\mathbb{B}_d$, that is, if $B\subseteq\mathbb{R}^d\setminus\mathbb{B}_d$, then $\mu_{\widetilde{\psi}}(B)=0$, see Theorem 4.8 in Villani (2003) or Cordero-Erausquin and Figalli (2019) for further details.

The main result of this note follows from the following crucial lemma. Its proof relies on the use of normalizing maps for convex sets. A convex set $\Omega\subset\mathbb{R}^d$ is said to be normalized if $\mathbb{B}_d\subseteq\Omega\subseteq d\,\mathbb{B}_d$. For each open bounded convex set Ω there exists a unique invertible affine transformation L normalizing Ω (this is John's celebrated Lemma of convex analysis, see Lemma A.13 in Figalli (2017)). We refer to L as the normalizing map and to $L(\Omega)$ as the normalized version of Ω.

Lemma 6.2.4. *Let* P *be a probability measure with density* p *supported on the open convex set* $\mathcal{X}\subseteq\mathbb{R}^d$*. Assume that* p *satisfies (6.11), denote by* $\psi:\mathbb{B}_d\to\mathbb{R}^d$ *the optimal transportation potential from* U_d *to* P *with* $\psi(\mathbf{0})=0$*. Then* ψ *is strictly convex on* $\mathbb{B}_d$*.*

Proof. From the comments after (6.16) we see that we can equivalently work with $\widetilde{\psi}$. To prove the result, assume that the contrary holds true. Then, there exist $\mathbf{y}\in\mathbb{B}_d$ and $\mathbf{t}\in\partial\widetilde{\psi}(\mathbf{y})$ such that, putting $l(\mathbf{z}) := \widetilde{\psi}(\mathbf{y}) + \langle\mathbf{t},\mathbf{z}-\mathbf{y}\rangle$, the convex set $\Sigma := \{\mathbf{z}:\ \widetilde{\psi}(\mathbf{z}) = l(\mathbf{z})\}$ is not a singleton. By subtracting an affine function, we can assume $\widetilde{\psi}(\mathbf{y}) = 0$ and $\widetilde{\psi}(\mathbf{z}) \geq 0$ for all $\mathbf{z}$; then, $\Sigma = \{\mathbf{z}:\ \widetilde{\psi}(\mathbf{z}) = 0\} = \{\mathbf{z}:\ \widetilde{\psi}(\mathbf{z})\leq 0\}$, which is closed since $\widetilde{\psi}$ is lower semicontinuous. Also, by adding the convex function $w(\mathbf{z}) := \frac{1}{2}(|\mathbf{z}|-1)_+^2$ (note that $\widetilde{\psi} = \widetilde{\psi}+w$ on $\bar{\mathbb{B}}_d$), we can assume that $\Sigma\subset\bar{\mathbb{B}}_d$. Being compact and convex, Σ coincides with the closed convex hull of its extreme points; as a consequence, it must have at least two exposed points (otherwise it would be empty or a singleton). Let $\bar{\mathbf{y}}\in\bar{\mathbb{B}}_d\setminus\{\mathbf{0}\}$ be one of them. If $\bar{\mathbf{y}}\in\mathbb{B}_d\setminus\{\mathbf{0}\}$, we consider a small ball $C_{\bar{\mathbf{y}}}$, say, around $\bar{\mathbf{y}}$, such that $\bar{C}_{\bar{\mathbf{y}}}\subset\mathbb{B}_d\setminus\{\mathbf{0}\}$. Then $\partial\widetilde{\psi}(\bar{C}_{\bar{\mathbf{y}}})$ is a compact set, and hence $\partial\widetilde{\psi}(\bar{C}_{\bar{\mathbf{y}}})\subset R\,\mathbb{B}_d$ for some $R>0$. By Proposition

6.2.1 (ii), we have constants $0 < \lambda_{C_{\bar{\mathbf{y}}}} \leq \Lambda_{C_{\bar{\mathbf{y}}}}$ such that the Monge-Ampère measure $\mu_{\widetilde{\psi}}$ satisfies $\lambda_{C_{\bar{\mathbf{y}}}} d\mathbf{x} \leq \mu_{\widetilde{\psi}} \leq \Lambda_{C_{\bar{\mathbf{y}}}} d\mathbf{x}$ in $C_{\bar{\mathbf{y}}}$. But the set Σ has an exposed point in $C_{\bar{\mathbf{y}}}$ and this contradicts Theorem 4.10 in Figalli (2017). Consequently, we must assume that $\bar{\mathbf{y}} \in \partial\mathbb{B}_d$. Observe that $\widetilde{\psi}(\bar{\mathbf{y}}) = 0$, hence $\bar{\mathbf{y}} \in \mathrm{dom}(\widetilde{\psi})$. First consider the case where $\bar{\mathbf{y}} \notin \partial\big(\mathrm{dom}(\widetilde{\psi})\big)$. Let $\mathbb{B}_r(\mathbf{x}) := \mathbf{x} + r\mathbb{B}_d$ and $\bar{\mathbb{B}}_r(\mathbf{x})$ denote, respectively the open and the closed ball of radius r centered at $\mathbf{x}$. Then, for $\eta > 0$ small enough, $\bar{\mathbb{B}}_\eta(\mathbf{y}) \subset \mathrm{dom}(\widetilde{\psi})$; consequently, there exists some R_0 such that $\partial\widetilde{\psi}\big(\bar{\mathbb{B}}_\eta(\bar{\mathbf{y}})\big) \subset R_0\,\mathbb{B}_d$. For η small enough, we further can ensure that $\bar{\mathbb{B}}_\eta(\bar{\mathbf{y}}) \subset 2\mathbb{B}_d$.

Without any loss of generality, let us assume that $\bar{\mathbf{y}} = \mathbf{e}_1$ where $\mathbf{e}_1$ stands for the first vector in the canonical basis of $\mathbb{R}^d$ (we can use a rotation otherwise). Then,

$$\Sigma \subset \big\{\mathbf{z} = (z_1, \ldots, z_d)' \in \mathbb{R}^d : \ z_1 \leq 1\big\} \quad \text{and} \quad \Sigma \cap \big\{\mathbf{z} = (z_1, \ldots, z_d)' \in \mathbb{R}^d : \ z_1 = 1\big\} = \{\mathbf{e}_1\}.$$

For $\sigma \in (0,1)$ small enough, we have

$$\Sigma \cap \big\{\mathbf{z} \in \mathbb{R}^d : \ z_1 \geq 1 - \sigma\big\} \subset \mathbb{B}_d \cap \big\{\mathbf{z} \in \mathbb{R}^d : \ z_1 \geq 1 - \sigma\big\} \subset \mathbb{B}_\eta(\mathbf{e}_1).$$

For such σ, defining

$$\psi_\epsilon(\mathbf{z}) := \widetilde{\psi}(\mathbf{z}) - \epsilon(z_1 - 1 + \sigma) \quad \text{and} \quad S_\epsilon := \{\mathbf{z} : \psi_\epsilon(\mathbf{z}) < 0\}, \tag{6.18}$$

observe that

$$S_\epsilon \longrightarrow \Sigma \cap \big\{\mathbf{z} \in \mathbb{R}^d : \ z_1 \geq 1 - \sigma\big\} \tag{6.19}$$

in the Hausdorff distance d_H as $\epsilon \to 0$, where we recall that, for $A, B \subseteq \mathbb{R}^d$,

$$\mathrm{d}_H(A, B) := \max\big\{\sup_{\mathbf{a}\in A}\inf_{\mathbf{b}\in B} |\mathbf{a} - \mathbf{b}|, \sup_{\mathbf{b}\in B}\inf_{\mathbf{a}\in A} |\mathbf{a} - \mathbf{b}|\big\}.$$

Hence, for $\epsilon > 0$ small enough, the sets S_ϵ are bounded open convex subsets of the ball $\mathbb{B}_\eta(\mathbf{e}_1)$. By Lemma 6.2.3, there exists some $M > 0$ such that

$$\mu_{\psi_\epsilon}(A) = \mu_{\widetilde{\psi}}(A) \leq M(\ell_d(A))^{1/d}$$

for every $A \subset S_\epsilon$ and ϵ small enough.

Next, fix $\mathbf{z}_0 \in \mathbb{B}_d \cap \Sigma$ and $\delta > 0$ such that $\bar{\mathbb{B}}_\delta(\mathbf{z}_0) \subset \mathbb{B}_d \cap \mathbb{B}_\eta(\mathbf{e}_1)$ and consider the normalizing map L_ϵ—namely, the affine transformation L_ϵ that normalizes S_ϵ. Denote by v_ϵ the normalized solution in $S^L_\epsilon := L_\epsilon(S_\epsilon)$ of $\mu_{v_\epsilon} = f \circ L_\epsilon^{-1}$ with the boundary condition $v_\epsilon = 0$ on ∂S^L_ϵ (v_ϵ is the convex map that has Monge-Ampère measure

$$d\mu_{v_\epsilon}(\mathbf{x}) = f \circ L_\epsilon^{-1}(\mathbf{x}) d\mathbf{x}$$

in S^L_ϵ and vanishes at the boundary of S^L_ϵ; its existence and uniqueness is guaranteed, for instance, by Proposition 4.2 in Figalli (2017)). Since $\mathbb{B}_d \subset S^L_\epsilon$, we have that $L_\epsilon^{-1}(\mathbb{B}_d) \subset 2\mathbb{B}_d$ and, therefore, the map L satisfies

$$|L_\epsilon(\mathbf{x}) - L_\epsilon(\mathbf{z})| \geq \frac{1}{2}|\mathbf{x} - \mathbf{z}| \quad \text{for all} \quad \mathbf{x}, \mathbf{z} \in \mathbb{R}^d.$$

It follows that

$$L_\epsilon(\mathbb{B}_d) \supset L_\epsilon\left(B_\delta\left(\mathbf{z}_0\right)\right) \supset B_{\delta/2}\left(L_\epsilon(\mathbf{z}_0)\right). \tag{6.20}$$

We consider the sets $S^L_{\epsilon,\delta} := \left\{\mathbf{z} \in S^L_\epsilon : \mathrm{d}_H\left(\mathbf{z}, \partial S^L_\epsilon\right) \geq \delta/4\right\}$. Now $L_\epsilon(\mathbf{z}_0) \in S^L_\epsilon$, a normalized set (it contains the unit ball and is contained in the ball of radius d, the dimension of the Euclidean space). Hence, there exists a constant $k_d > 0$, depending only on d such that (see Theorem 4.23 in Figalli (2017) or Lemma 3 in Caffarelli (1992))

$$\ell_d\big(S^L_{\epsilon,\delta} \cap \mathbb{B}_{\delta/2}(L_\epsilon(\mathbf{z}_0))\big) \geq k_d \delta^d.$$

In view of Lemma 6.2.3, the subsequent remark, and the fact that $\bar{\mathbb{B}}_\delta(\mathbf{z}_0) \subset \mathbb{B}_\eta(\mathbf{e}_1)$, we have that μ_{ψ_ϵ} is lower bounded over $\bar{\mathbb{B}}_\delta(\mathbf{z}_0)$, that is, there exists $\lambda > 0$ such that $\mu_{\psi_\epsilon}(A) \geq \lambda \ell_d(A)$ for every $A \subseteq \bar{\mathbb{B}}_\delta(\mathbf{z}_0)$. This and (6.20) thus imply that μ_{v_ϵ} is bounded from below on $\mathbb{B}_{\delta/2}(L_\epsilon(\mathbf{z}_0))$. It follows that, for some $\lambda > 0$,

$$\mu_{v_\epsilon}\left(S^L_{\epsilon,\delta}\right) \geq \lambda \ell_d\Big(S^L_{\epsilon,\delta} \cap \mathbb{B}_{\delta/2}\left(L_\epsilon(\mathbf{z}_0)\right)\Big) \geq C\delta^d.$$

This implies that, for c' small enough, no ball of radius $c'\delta/2$ can contain $\partial v_\epsilon\big(S^L_{\epsilon,\delta}\big)$. As a consequence, there exists a constant $c > 0$ such that $\sup_{\mathbf{p} \in \partial v_\epsilon(S^L_{\epsilon,\delta})} |\mathbf{p}| \geq c\delta$. Using Corollary A.23 in Figalli (2017), we conclude that

$$\big|\min_{S^L_\epsilon} v_\epsilon\big| \geq c''\left(\delta/2\right)^2$$

for some $c'' > 0$. On the other hand, using Lemma 6.14 again to upper bound μ_{ψ_ϵ}, we obtain

$$\mu_{v_\epsilon}\left(S^L_\epsilon\right) = \mu_{\psi_\epsilon}\left(S_\epsilon\right) \leq M(\ell_d(2\mathbb{B}_d))^{1/d}$$

which, by the Alexandrov maximum principle (e.g., Theorem 2.8. in Figalli (2017)), entails

$$|v_\epsilon\left(L_\epsilon \mathbf{e}_1\right)| \leq C\left(\mathrm{d}_H(L_\epsilon \mathbf{e}_1, \partial S^L_\epsilon)\right)^{1/d}.$$

This means that the same arguments as in the proof of Theorem 4.10 in Figalli (2017) yield

$$\lim_{\epsilon \to 0} \mathrm{d}_H\left(L_\epsilon \mathbf{e}_1, \partial(S^L_\epsilon)\right) = 0 \quad \text{and} \quad \lim_{\epsilon \to 0} \frac{v_\epsilon\left(L_\epsilon \mathbf{e}_1\right)}{\min_{S^L_\epsilon} v_\epsilon} = 1,$$

which is a contradiction.

Finally, consider the case where the exposed point of Σ belongs to $\partial\big(\mathrm{dom}(\widetilde{\psi})\big)$; here again, it can be assumed that $\Sigma = \{\mathbf{z} : \widetilde{\psi}(\mathbf{z}) = 0\}$ and that the exposed point of Σ is $\mathbf{e}_1$. We also can assume, without loss of generality, that $\mathrm{dom}(\widetilde{\psi}) \subset \left\{\mathbf{z} \in \mathbb{R}^d : \ z_1 \leq 1\right\}$. Hence, $\{c\,\mathbf{e}_1 : c \geq 0\} \subset \partial\widetilde{\psi}(\mathbf{e}_1)$. For small $\theta > 0$ we consider the sets

$$A_\theta := \mathbb{B}_d \cap \{\mathbf{z} = (z_1, \mathbf{z}') \in \mathbb{R} \times \mathbb{R}^{d-1} : \ -\theta|\mathbf{z}'| \leq z_1 - 1 \leq 0\}$$

and

$$C_\theta := \mathcal{X} \cap \{\mathbf{x} = (x_1, \mathbf{x}') \in \mathbb{R} \times \mathbb{R}^{d-1} : x_1 > 0, \, |\mathbf{x}'| \le \theta x_1\}.$$

Let $\mathbf{x} \in C_\theta$ and $\mathbf{z} \in \partial\widetilde{\psi}^*(\mathbf{x})$. Then $\mathbf{x} \in \partial\widetilde{\psi}(\mathbf{z})$ and, thanks to the monotonicity of $\partial\widetilde{\psi}$, we have that $\langle \mathbf{x} - t\,\mathbf{e}_1, \mathbf{z} - \mathbf{e}_1 \rangle \ge 0$ for every $t \ge 0$, which entails $\langle \mathbf{x}, \mathbf{z} - \mathbf{e}_1 \rangle \ge 0$ (take $t = 0$) and $\langle \mathbf{e}_1, \mathbf{z} - \mathbf{e}_1 \rangle \ge 0$ (take $t \to \infty$). This means that $z_1 \le 1$ and $x_1(z_1 - 1) + \langle \mathbf{x}', \mathbf{z}' \rangle \ge 0$, from which we deduce that $z_1 - 1 \ge \theta|\mathbf{z}'|$. Since, by Lemma 6.2.2, we have $\mathbf{z} \in \mathbb{B}_d$, it follows that $\partial\widetilde{\psi}(A_\theta) \supset C_\theta$. Also, since both $\mathbf{0}$ and $\mathbf{e}_1$ belong to $\partial\widetilde{\psi}(\mathbb{R}^d) \subset \bar{\mathcal{X}}$, which is a convex set with nonempty interior, we can argue as in pp. 8-9 of Cordero-Erausquin and Figalli (2019) to conclude that $\ell_d(C_\theta \cap 2\mathbb{B}_d) \gtrsim \theta^{d-1}$ for $\theta > 0$ small enough. From the transport equation, we have that

$$\mathrm{U}_d(A_\theta) = \int_{\partial\widetilde{\psi}(A_\theta)} p(\mathbf{x})d\mathbf{x} \ge \int_{C_\theta} p(\mathbf{x})d\mathbf{x} \gtrsim \ell_d(C_\theta \cap 2\mathbb{B}_d) \gtrsim \theta^{d-1}, \tag{6.21}$$

where we have used that p is lower bounded on bounded subsets of $\mathcal{X}$. However, for small θ, A_θ is well separated from $\mathbf{0}$ and, consequently, u_d is upper bounded on A_θ. This means that

$$\mathrm{U}_d(A_\theta) \lesssim \ell_d(A_\theta) \lesssim \theta^{d+1},$$

which contradicts (6.21). This completes the proof of the claim that $\widetilde{\psi}$ (equivalently, ψ) is strictly convex in $\mathbb{B}_d$. □

A careful look at the proof of Lemma 6.2.4 shows that a similar conclusion could be obtained if the domain of $\widetilde{\psi}$ were a compact, strictly convex set different from $\mathbb{B}_d$. However, the argument breaks down without the strict convexity assumption. In practical terms this means that strict convexity of $\widetilde{\psi}$ (which is the key to the main result in this paper) may fail if, instead of U_d, we choose, for instance, the uniform distribution on the unit hypercube as the reference measure.

We can now state and, based on Lemma 6.2.4, prove our main result, which extends Theorem 1.1 in Figalli Figalli (2018) to the case of convexly supported (bounded or unbounded support) distributions. Recall that $\mathcal{C}^{k,\alpha}_{\mathrm{loc}}(\mathcal{X})$ denotes the set of real functions on $\mathcal{X}$ which are k times differentiable and have derivatives of order k which are locally Hölder-continuous of order $\alpha \in (0, 1]$.

Theorem 6.2.5. *Let* P *be a probability measure with density* p *supported on the open convex set* $\mathcal{X} \subseteq \mathbb{R}^d$.

(i) *If* p *satisfies* (6.11)*, then* $K := \partial\psi(\mathbf{0})$ *is a compact, convex set with Lebesgue measure* 0 *such that the center-outward quantile function* $\mathbf{Q}_\pm := \nabla\psi$ *and the center-outward distribution function* $\mathbf{F}_\pm := \nabla\psi^*$ *are homeomorphisms between* $\mathbb{B}_d \setminus \{\mathbf{0}\}$ *and* $\mathcal{X} \setminus K$*, inverses of each other.*

(ii) *If, moreover,* $p \in \mathcal{C}^{k,\alpha}_{\mathrm{loc}}(\mathcal{X})$ *for some* $k \in \mathbb{N}$ *and* $\alpha \in (0, 1)$*, then* $\mathbf{Q}_\pm$ *and* $\mathbf{F}_\pm$ *are diffeomorphisms of class* $\mathcal{C}^{k+1,\alpha}_{\mathrm{loc}}$ *between* $\mathbb{B}_d \setminus \{\mathbf{0}\}$ *and* $\mathcal{X} \setminus K$.

Proof. Since $\mathbf{0}$ belongs to the interior of the domain of ψ, we have that K is compact and convex (see, e.g., Lemma A.22 in Figalli (2017)). Moreover, Proposition 6.2.1, part (ii), implies that K has Lebesgue measure zero. Now, Assumption A implies, for any closed ball $B \subseteq \mathbb{B}_d \setminus \{\mathbf{0}\}$, the existence of constants $0 < \lambda_B \leq \Lambda_B$ such that $\lambda_B d\mathbf{x} \leq \mu_{\widetilde{\psi}} \leq \Lambda_B d\mathbf{x}$. It follows from Caffarelli's regularity theory (see Corollary 4.21 in Figalli (2017)) that ψ is locally of class $C^{1,\alpha}$. Since the constant $\alpha > 0$ depends on λ and Λ we cannot conclude that, for some $\alpha > 0$, $\psi \in C^{1,\alpha}_{\text{loc}}(\mathbb{B}_d \setminus \{\mathbf{0}\})$. However, ψ is continuously differentiable on $\mathbb{B}_d \setminus \{\mathbf{0}\}$ and, therefore, $\mathbf{Q}_\pm = \nabla\psi$ is a (single-valued) continuous function on $\mathbb{B}_d \setminus \{\mathbf{0}\}$. Furthermore, the strict convexity of ψ (Lemma 6.2.4) implies that $\mathbf{Q}_\pm$ is injective. By Brouwer's theorem on invariance of domain (see, e.g., Theorem 2B.3, p. 172 in Hatcher (2002)), $\mathbf{Q}_\pm(\mathbb{B}_d \setminus \{\mathbf{0}\})$ is open and $\mathbf{Q}_\pm$ is a homeomorphism between $\mathbb{B}_d \setminus \{\mathbf{0}\}$ and $\mathbf{Q}_\pm(\mathbb{B}_d \setminus \{\mathbf{0}\})$. But then, necessarily, $\mathbf{Q}_\pm(\mathbb{B}_d \setminus \{\mathbf{0}\}) = \mathcal{X} \setminus K$, which completes the proof of part (i) of the theorem.

Turning to part (ii), assume that $p \in \mathcal{C}^{k,\alpha}_{\text{loc}}(X)$, fix $\mathbf{x} \in \mathbb{B}_d$, and consider a neighborhood V of $\mathbf{x}$ such that its closure $\bar{V}$ is contained in $\mathbb{B}_d \setminus \{\mathbf{0}\}$. Now, ψ is strictly convex over V and there exist constants $0 < \lambda_V < \Lambda_V$ such that $\lambda_V d\mathbf{x} \leq \mu_{\widetilde{\psi}} \leq \Lambda_V d\mathbf{x}$ on V. Note that $u_d \in \mathcal{C}^{k,\alpha}_{\text{loc}}(V)$ for every k and α. Hence, we can apply Remark 4.44 in Figalli (2017) to conclude that $\nabla\psi \in \mathcal{C}^{k+1,\alpha}_{\text{loc}}(V)$. This completes the proof. □

We conclude this section with several remarks. First, we observe that the center-outward quantile function of a probability measure P satisfying the assumptions of Theorem 6.2.5 may fail to be continuous (singled-valued) at the origin. However, the center-outward distribution function is single-valued (and consequently continuous) at every point of the support of P, since the points in the set K which had to be removed to guarantee that $\mathbf{Q}_\pm$ is a homeomorphism are all mapped by $\mathbf{F}_\pm$ to the origin.

Our second remark concerns the convexity assumption on the support of P. While this may look unnatural, extensions to nonconvex domains enter into rather delicate technical issues. The bottom line is the problem of characterizing the class of distributions that are the push-forward, by some continuous and single-valued (on $\mathbb{B}_d \setminus \{\mathbf{0}\}$) gradient of convex quantile map, of U_d. Continuity of the quantile map requires, at least, that P has a connected support. However, strictly speaking, this is not a sufficient condition: Figalli (2017) (pp. 97-98) provides an example of a discontinuous optimal transport map from a constant density on a ball to a constant density on a connected set; the same map would yield an example of a discontinuous quantile map for a probability with a positive density on a connected support. The problem, thus, is connected to the shape of the support of P rather than to the singularity at the origin of the reference measure U_d: the example in Figalli (2017) is given for a reference measure with a constant, hence bounded, density. We must add, though, that the nature of the discontinuity may be regarded as mostly formal. For the two-dimensional case, Figalli (2010) shows that for probabilities P and Q with smooth densities supported on bounded open sets Ω and Λ, the optimal map from P to Q is a homeomorphism from Ω' to Λ', where Ω' and Λ' are open sets such that $\mathrm{P}(\Omega') = \mathrm{Q}(\Lambda') = 1$. Furthermore, a rather precise description of the discontinuity set is given. Roughly speaking, for smooth connected Λ (not necessarily convex), the discontinuity set is a smooth manifold of dimension ≤ 1 that touches the boundary of Ω. Extensions of these results to higher dimension and more general probabilities would be of great interest and would help to extend the theory presented in this work.

The median set, namely $K = \partial\psi(\mathbf{0})$ (equivalently, the set of points $\mathbf{x}$ such that $\mathbf{F}_\pm(\mathbf{x}) = \mathbf{0}$), in principle, is not necessarily a singleton. If P is the uniform distribution (constant density) on the annulus $\{\mathbf{x} \in \mathbb{R}^d : 0 < a \leq |\mathbf{x}| \leq b\}$, then K is the disk $\{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x}| \leq a\}$ (this example was kindly suggested by a referee). However, looking for an example of a probability P with convex support $\mathcal{X}$ for which this set is not a singleton seems to be a hard task. If $d = 2$ and $\mathcal{X} = \mathbb{R}^2$ then Figalli (2018) shows that K is indeed a singleton but conjectures that the result fails in higher dimension. We believe that this issue deserves further analysis, but we refrain from pursuing it further here.

Our final comment in this section deals with the smoothness of center-outward quantile contours. Part *(ii)* of Theorem 6.2.5 guarantees that $\mathbf{Q}_\pm(r\mathbb{S}_{d-1})$ is a hypersurface of class $\mathcal{C}^{k+1,\alpha}$ if P has a smooth density of class $\mathcal{C}^{k,\alpha}$. This is in sharp contrast to other notions of quantile or depth contours. As an example, we could mention the bivariate distribution with independent Cauchy marginals. In this case Tukey's halfspace depth contours are squares, hence they have points of nondiffererentiability (see Proposition 13 in Rousseeuw and Ruts (1999)), while the center-outward quantile contours are $\mathcal{C}^\infty$ curves.

6.3 Some further properties of center-outward distribution and quantile functions.

To conclude this note, we present three results that more or less directly follow as consequences of Theorem 6.2.5. The first one is a kind of “extreme" invariance of center-outward distribution functions: the limit of center-outward distribution functions along rays in direction $\mathbf{u} \in \mathbf{S}_{d-1}$ is $\mathbf{u}$. The second result deals with the ability of center-outward quantile functions to capture the shape of a convex supporting set. The third one is about the shape of quantile contours, which turn out to satisfy a kind of relaxed version of convexity, connected to the so-called “lighthouse convexity” property (see, e.g., pp. 263-264 in Cholaquidis et al. (2017)).

To motivate the announced invariance result, we recall that a classical univariate distribution function F trivially satisfies

$$\lim_{x\to-\infty} F(x) = 0, \quad \lim_{x\to\infty} F(x) = 1,$$

hence, in terms of the univariate center-outward distribution function $\mathbf{F}_\pm := 2F - 1$,

$$\lim_{t\to\infty} \mathbf{F}_\pm(tu) = u \qquad \text{for all } u \text{ such that } |u| = 1.$$

The following proposition shows that this carries over to $\mathbf{F}_\pm$ in general dimension. For probability measures supported on all of $\mathbb{R}^d$, this implies a similar result for extreme quantiles, a characteristic shared with other notions of multivariate quantiles. Details about this connection are given after Corollary 6.3.2 below.

Proposition 6.3.1. *Let the probability measure* P *have a density on* $\mathbb{R}^d$*. For any* $\mathbf{u}$ *on the unit sphere* $\mathbb{S}_{d-1}$*, any sequence* $(t_n)_{n\in\mathbb{N}}$ *of real numbers such that* $t_n \to \infty$*, and any* $\mathbf{y}_n \in \partial\varphi(t_n\mathbf{u})$*,*

$$\lim_{n\to\infty} \mathbf{y}_n = \mathbf{u}.$$

Proof. It follows from (6.6) that $\mathbf{y}_n \in \bar{\mathbb{B}}_d$. Hence, by compactness, there exists a subsequence along which $\mathbf{y}_n \to \mathbf{y}_\infty \in \bar{\mathbb{B}}_d$. On the other hand, monotonicity of the subdifferential implies that, for all $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{y} \in \partial\varphi(\mathbf{x})$,

$$\langle \mathbf{y} - \mathbf{y}_n, \mathbf{x} - t_n\mathbf{u} \rangle \geq 0$$

or, equivalently, for all $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{y} \in \partial\psi(\mathbf{x})$,

$$\langle \mathbf{y} - \mathbf{y}_\infty, \mathbf{x} \rangle + \langle \mathbf{y}_\infty - \mathbf{y}_n, \mathbf{x} \rangle \geq t_n \left(\langle \mathbf{y} - \mathbf{y}_\infty, \mathbf{u} \rangle + \langle \mathbf{y}_\infty - \mathbf{y}_n, \mathbf{u} \rangle \right).$$

Fixing $\epsilon > 0$ and $N = N(\epsilon)$ such that $|\mathbf{y}_n - \mathbf{y}_\infty| < \epsilon$ for all $n \geq N$, we obtain

$$\langle \mathbf{y} - \mathbf{y}_\infty, \mathbf{x} \rangle + \epsilon|\mathbf{x}| \geq t_n \left(\langle \mathbf{y} - \mathbf{y}_\infty, \mathbf{u} \rangle - \epsilon \right).$$

Hence, for n large enough, $\langle \mathbf{y} - \mathbf{y}_\infty, \mathbf{u} \rangle - \epsilon \leq 0$ for all $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{y} \in \partial\varphi(\mathbf{x})$. Since $\epsilon > 0$ is arbitrary, we conclude that

$$\partial\varphi(\mathbb{R}^d) \subset S := \{\mathbf{y} : \ \langle \mathbf{y} - \mathbf{y}_\infty, \mathbf{u} \rangle \leq 0\},$$

which is a halfspace. Now, the fact that $\nabla\varphi$ pushes P forward to U_d implies that $\partial\varphi(\mathbb{R}^d)$ contains almost every $\mathbf{x} \in \mathbb{B}_d$. Hence, $\mathbb{B}_d \subset S$, which only can happen if $\mathbf{y}_\infty = \mathbf{u}$. □

Under additional smoothness assumptions on P, the announced result for $\mathbf{F}_\pm$ follows as a corollary.

Corollary 6.3.2. *Let* P *satisfy the assumptions in Proposition 6.2.1* (ii)*. Then, for any* $\mathbf{u}$ *on the unit sphere* $\mathbb{S}_{d-1}$ *and any sequence* $(t_n)_{n\in\mathbb{N}}$ *of real numbers such that* $t_n \to \infty$*, we have*

$$\lim_{n\to\infty} \mathbf{F}_\pm(t_n\mathbf{u}) = \mathbf{u}.$$

Note that the proof Proposition 6.3.1 actually entails, with minimal and obvious changes, $\lim_{n\to\infty} \mathbf{F}_\pm(t_n\mathbf{u}_n) = \mathbf{u}$ for any sequence $t_n \to \infty$ and any $\mathbf{u}_n \in \mathbb{S}_{d-1}$ such that $\mathbf{u}_n \to \mathbf{u}$ (we assume for simplicity that P satisfies the assumptions in Proposition 6.2.1 (ii)). Also, the same proof shows that $\lim_{n\to\infty} \mathbf{F}_\pm(\theta + t_n\mathbf{u}) = \mathbf{u}$ for any $\theta \in \mathbb{R}^d$ (so that we could take, for instance, a center-outward median, that is, let $\theta = \theta_\mathrm{P}$ such that $\mathbf{F}_\pm(\theta_\mathrm{P}) = \mathbf{0}$, and get a more natural formulation).

We also would like to observe that Corollary 6.3.2 can be rewritten in terms of extreme quantiles. Assume for simplicity that P satisfies the assumptions in Proposition 6.2.1 (ii) with $\mathcal{X} = \mathbb{R}^d$. Take $\mathbf{v}_n \in \mathbb{B}_d$ with $\mathbf{v}_n \to \mathbf{v} \in \mathbb{S}_{d-1}$. Necessarily, $t_n := \|\mathbf{Q}_\pm(\mathbf{v}_n)\| \to \infty$. If we set $\mathbf{u}_n := \frac{\mathbf{Q}_\pm(\mathbf{v}_n)}{\|\mathbf{Q}_\pm(\mathbf{v}_n)\|}$, then, by compactness, we can assume that, along subsequences, $\mathbf{u}_n \to$

$\mathbf{u} \in \mathbb{S}_{d-1}$. Then, by Corollary 6.3.2 (and subsequent comments) we have $\mathbf{F}_\pm(t_n\mathbf{u}_n) \to \mathbf{u}$. Since, on the other hand,

$$\mathbf{F}_\pm(t_n\mathbf{u}_n) = \mathbf{F}_\pm(\mathbf{Q}_\pm(\mathbf{v}_n)) = \mathbf{v}_n \to \mathbf{v},$$

we conclude that

$$\lim_{n\to\infty} \frac{\mathbf{Q}_\pm(\mathbf{v}_n)}{\|\mathbf{Q}_\pm(\mathbf{v}_n)\|} = \mathbf{v} \tag{6.22}$$

whenever $\mathbf{v}_n \in \mathbb{B}_d$ is such that $\mathbf{v}_n \to \mathbf{v} \in \mathbb{S}_{d-1}$. This property of center-outward quantiles, hence, is shared with geometric or spatial quantiles (see Theorem 2.1 in Girard and Stupfler (2017) and Theorem 3 in Paindaveine and Virta (2020)). A major difference, though, is that for geometric quantiles $\|\mathbf{Q}(\mathbf{v}_n)\| \to \infty$ if $\mathbf{v}_n \to \mathbf{v} \in \mathbb{S}_{d-1}$, even when P has bounded support while, by construction, $\mathbf{Q}(\mathbf{v}_n) \in \mathcal{X}$ if $\mathbf{v}_n \in \mathbb{B}_d$ and, as a consequence, $\|\mathbf{Q}(\mathbf{v}_n)\|$ is necessarily bounded if $\mathcal{X}$ is bounded. We remark that Corollary 6.3.2 is valid for boundedly supported probabilities (as, for instance, in the example at the end of subsection 2.1, in which $\mathbf{F}_\pm(t_n\mathbf{u}) = \mathbf{u}$ for $t_n \geq 1$).

Next, we include the announced simple result showing that the outer quantile contours of a convexly supported P approach (in Hausdorff distance) the boundary of its support.

Lemma 6.3.3. *Let* P *be a probability measure on* $\mathbb{R}^d$ *with compact convex support* $\mathcal{X}$ *and a density* p *such that* $\lambda \leq p \leq \Lambda$ *for some* $0 < \lambda \leq \Lambda$. *Then, as* $r \to 1$, $\mathbf{Q}_\pm(r\,\mathbb{B}_d)$ *tends to* $\mathcal{X}$ *in Hausdorff distance:*

$$\lim_{r\to 1} \mathrm{d}_H\big(\mathbf{Q}_\pm(r\,\mathbb{B}_d), \mathcal{X}\big) = 0.$$

Proof. Since $\mathbf{Q}_\pm(r\,\mathbb{B}_d) = \nabla\psi(R\,\mathbb{B}_d)$ is contained in $\mathcal{X}$, we only need to analyse one of the two members of the maximum defining the Hausdorff distance: indeed,

$$\mathrm{d}_H(\nabla\psi(r\,\mathbb{B}_d), \mathcal{X}) = \max\{\sup_{\mathbf{a}\in\nabla\psi(r\,\mathbb{B}_d)} \inf_{\mathbf{x}\in\mathcal{X}} |\mathbf{a}-\mathbf{x}|, \sup_{\mathbf{x}\in\mathcal{X}} \inf_{\mathbf{a}\in\nabla\psi(r\,\mathbb{B}_d)} |\mathbf{a}-\mathbf{x}|\} = \sup_{\mathbf{y}\in\mathbb{B}_d} \inf_{\mathbf{b}\in r\,\mathbb{B}_d} |\nabla\psi(\mathbf{b}) - \nabla\psi(\mathbf{y})|.$$

On the other hand, since $r'\,\mathbb{B}_d \subset r\,\mathbb{B}_d \subset \mathcal{X}$, $\nabla\psi(r'\,\mathbb{B}_d) \subset \nabla\psi(r\,\mathbb{B}_d) \subset \mathcal{X}$ for $r' \leq r$, so that the map $r \mapsto \mathrm{d}_H(\nabla\psi(r\,\mathbb{B}_d), \mathcal{X})$ is a decreasing function. Suppose that $\mathrm{d}_H(\nabla\psi(R\,\mathbb{B}_d), \mathcal{X})$ does not tend to 0 when R tends to 1. Then, there exists $\epsilon > 0$ such that, for every r, $\mathrm{d}_H(\nabla\psi(r\,\mathbb{B}_d), \mathcal{X}) > \epsilon$; in particular, there exists $\mathbf{x}_r \in \mathcal{X}$ such that $|\mathbf{a}_r - \mathbf{x}_r| > \epsilon$ for all $\mathbf{a}_r \in \nabla\psi(r\,\mathbb{B}_d)$.

Now, for each $n \in \mathbb{N}$, consider the sequences $A_n := \nabla\psi\big((1-1/n)\mathbb{B}_d\big)$ and $\mathbf{y}_n := \mathbf{x}_{1-1/n} \in \mathcal{X}$. These sequences are such that

$$\inf_{\mathbf{a}\in A_m} \big|\mathbf{a}-\mathbf{y}_n\big| \geq \inf_{\mathbf{a}\in A_n} \big|\mathbf{a}-\mathbf{y}_n\big| > \epsilon \quad \text{for all } m \leq n.$$

By compactness, the sequence $\mathbf{y}_n$ admits a convergent subsequence, with limit $\mathbf{y}_\infty$, say, where $\mathbf{y}_\infty \in \mathcal{X}$. This limit satisfies $\inf_{\mathbf{a}\in A_n} \big|\mathbf{a}-\mathbf{y}_\infty\big| > \epsilon$ for all $n \in \mathbb{N}$, which is not possible since $\mathcal{X} = \bigcup_{n\in\mathbb{N}} A_n$. □

Our final result concerns the shape of the quantile contours of smooth probability measures (those satisfying the assumptions of Theorem 6.2.5). As a consequence of Theorem 6.2.5, the sets $\mathbf{Q}_{\pm}(r\,\mathbb{B}_d)$ are bounded, with connected boundary. Beyond this type of topological properties, results on the geometry of the quantile regions are not available. Here we prove that they satisfy a weak form of convexity. Recall from Cholaquidis et al. (2017) that a set $B \subseteq \mathbb{R}^d$ is ρ-lighthouse convex if, from every point $\mathbf{x}$ in the boundary of B, there exists an open cone with vertex $\mathbf{x}$ and opening angle $\rho > 0$ which is contained in $\mathbb{R}^d \setminus B$. The limiting version of this concept (obtained as $\rho \to 0$) is that for every point $\mathbf{x}$ in the boundary of B there exists a ray emanating from $\mathbf{x}$ that does not intersect B at any other point. This is precisely what can be proved for quantile sets.

Lemma 6.3.4. *Let* P *be a probability measure on* $\mathbb{R}^d$ *satisfying the assumptions of Theorem 6.2.5. Then, for all* $r \in (0, 1)$ *and all* $\mathbf{y}$ *belonging to the boundary of* $\mathbf{Q}_{\pm}(r\,\mathbb{B}_d)$*, there exists a ray* T *emanating from* $\mathbf{y}$ *for which* $\overline{\mathbf{Q}_{\pm}(r\,\mathbb{B}_d)} \cap T = \{\mathbf{y}\}$*.*

Proof. Assume, on the contrary, that there exists $\mathbf{y}$ in the boundary of $\mathbf{Q}_{\pm}(r\,\mathbb{B}_d)$ such that for every ray

$$T = \{\mathbf{z} \in \mathbb{R}^d : \ \mathbf{z} = \mathbf{y} + t\mathbf{s}, \ t \geq 0\},$$

there exists in $\overline{\mathbf{Q}_{\pm}(r\,\mathbb{B}_d)} \cap T$ at least one point $\mathbf{z}$ distinct from $\mathbf{y}$. Note that, necessarily, that point can be chosen in the boundary of $\mathbf{Q}_{\pm}(r\,\mathbb{B}_d)$. Now, since $\mathbf{Q}_{\pm}$ is a homeomorphism, it maps boundaries into boundaries. Therefore, we can assume, up to a rotation, that $\mathbf{y} = \mathbf{Q}_{\pm}(r\mathbf{e}_1)$. Monotonicity of $\mathbf{Q}_{\pm}$ implies that

$$\langle \mathbf{u} - r\mathbf{e}_1, \mathbf{Q}_{\pm}(\mathbf{u}) - \mathbf{Q}_{\pm}(r\mathbf{e}_1)\rangle \geq 0 \qquad \text{for all } \mathbf{u} \in \mathbb{B}_d. \tag{6.23}$$

However, if $\mathbf{Q}_{\pm}(\mathbf{u}) \in T = \{\mathbf{z} \in \mathbb{R}^d : \ \mathbf{z} = \mathbf{y} + t\mathbf{e}_1, \ t \geq 0\}$, then $\mathbf{Q}_{\pm}(\mathbf{u}) = \mathbf{Q}_{\pm}(r\mathbf{e}_1) + t\mathbf{e}_1$ for some $t > 0$. Hence, by (6.23) $\langle \mathbf{u} - r\mathbf{e}_1, r\mathbf{e}_1\rangle \geq 0$. This implies that $\mathbf{u} \notin r\,\mathbb{B}_d$, thus contradicting the assumption that T has a common point with the boundary of $\mathbf{Q}_{\pm}(r\,\mathbb{B}_d)$ other than $\mathbf{y}$. □

Acknowledgments

Eustasio del Barrio's research was partially supported by FEDER, Spanish Ministerio de Economía y Competitividad, Grant MTM2017-86061-C2-1-P and Junta de Castilla y León, Grants VA005P17 and VA002G18.

The authors thank two referees for their very careful and insightful reports.

Bibliography

L A. Caffarelli. A localization property of viscosity solutions to the Monge-Ampère equation and their strict convexity. *Ann. Math.*, 131:129–134, 1990.

L A. Caffarelli. Some regularity properties of solutions of Monge-Ampère equation. *Comm. Pure Appl. Math.*, 44:965–969, 1991.

Luis A. Caffarelli. The regularity of mappings with a convex potential. *J. Amer. Math. Soc.*, 5:99–104, 1992.

V. Chernozhukov, A. Galichon, M. Hallin, and M. Henry. Monge-kantorovich depth, quantiles, ranks and signs. *Ann. Statist.*, 45:223–256, 2017.

A. Cholaquidis, A. Cuevas, and R. Fraiman. On Poincaré cone property. *Ann. Statist.*, 42: 255–284, 2017.

D. Cordero-Erausquin and A. Figalli. Regularity of monotone transport maps between unbounded domains. *Discrete Cont. Dyn.-A*, 39:7101–7112, 2019.

C. de Valk and J. Segers. Tails of optimal transport plans for regularly varying probability measures. *https://arxiv.org/abs/1811.12061*, 2018.

N. Deb and B. Sen. Multivariate rank-based distribution-free nonparametric testing using measure transportation. *https://arxiv.org/abs/1909.08733*, 2019.

E. del Barrio and Jean M. Loubes. Central limit theorems for empirical transportation cost in general dimension. *Ann. Probab.*, 47:926–951, 2019.

E. del Barrio, J. Cuesta Albertos, M. Hallin, and C. Matrán. Smooth cyclically monotone interpolation and empirical center-outward distribution functions. *https://ideas.repec.org/p/eca/wpaper/2013-271399.html*, 2018.

Eustasio del Barrio, Jan Beirlant, Sven Buitendag, and Marc Hallin. Center-outward quantiles and the measurement of multivariate risk. *https://arxiv.org/abs/1912.04924*, 2019.

Eustasio del Barrio, Alberto González-Sanz, and Marc Hallin. A note on the regularity of optimal-transport-based center-outward distribution and quantile functions. *Journal of Multivariate Analysis*, 180:104671, 2020. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2020.104671. URL https://www.sciencedirect.com/science/article/pii/S0047259X20302529.

Ivar Ekeland, Alfred Galichon, and Marc Henry. Comonotonic measures of multivariate risks. *Math. Financ.*, 22:109–132, 2012.

L.C. Evans and R.F. Gariepy. *Measure theory and fine properties of functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

O.P. Faugeras and L. Rüschendorf. Markov morphisms: a combined copula and mass transportation approach to multivariate quantiles. *Math. Appl. (Warsaw)*, 45:21–63, 2017.

O.P. Faugeras and L. Rüschendorf. Functional, randomized and smoothed multivariate quantile regions. Tse working papers, Toulouse School of Economics (TSE), 2019. URL https://ideas.repec.org/p/tse/wpaper/123569.html.

A. Figalli. Regularity properties of optimal maps between nonconvex domains in the plane. *Commun. Part. Diff. Eq.*, 35:465–479, 2010.

A. Figalli. *The Monge-Ampère equation and its applications*. Zurich Lectures in Advanced Mathematics. European Mathematical Society (EMS), Zürich, 2017.

Alessio Figalli. On the continuity of center-outward distribution and quantile functions. *Nonlinear Anal.*, 177:413–421, 2018.

Christian Genest and Louis-Paul Rivest. On the multivariate probability integral transformation. *Statist. Probab. Lett.*, 53:391–399, 2001.

P. Ghosal and B. Sen. Multivariate ranks and quantiles using optimal transportation and applications to goodness-of-fit testing. *https://arxiv.org/abs/1905.05340*, 2019.

Stéphane Girard and Gilles Stupfler. Intriguing properties of extreme geometric quantiles. *REVSTAT*, 15:107–139, 2017.

Jaroslav Hájek, Zbyněk Šidák, and Pranab K. Sen. *Theory of Rank Tests, 2nd Ed.* Probability and Mathematical Statistics. Academic Press, Inc., San Diego, CA, 1999.

M. Hallin. On the Pitman nonadmissibility of correlogram-based time series methods. *J. Time Ser. Anal.*, 15:607–611, 1994.

M. Hallin. On distribution and quantile functions, ranks and signs in $\mathbb{R}^d$. Working papers ecares, ULB – Université libre de Bruxelles, 2017. URL https://ideas.repec.org/p/eca/wpaper/2013-258262.html.

M. Hallin and D. Paindaveine. Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.*, 30:1103–1133, 2002a.

M. Hallin and D. Paindaveine. Optimal procedures based on interdirections and pseudo-Mahalanobis ranks for testing multivariate elliptic white noise against ARMA dependence. *Bernoulli*, 8:787–815, 2002b.

M. Hallin and D. Paindaveine. Rank-based optimal tests of the adequacy of an elliptic VARMA model. *Ann. Statist.*, 32:2642–2678, 2004.

M. Hallin and O. Tribel. The efficiency of some nonparametric rank-based competitors to correlogram methods. In *Game Theory, Optimal Stopping, Probability and Statistics*, volume 35 of IMS Lecture Notes Monogr. Ser., pages 249–262. Inst. Math. Statist., Beachwood, OH, 2000.

M. Hallin and B.J.M. Werker. Semi-parametric efficiency, distribution-freeness and invariance. *Bernoulli*, 9:137–165, 2003.

M. Hallin, D. Paindaveine, and T. Verdebout. Optimal rank-based testing for principal components. *Ann. Statist.*, 38:3245–3299, 2010.

M. Hallin, D. La Vecchia, and H. Liu. Center-outward R-estimation for semiparametric VARMA models. *https://arxiv.org/abs/1910.08442*, 2019.

M. Hallin, E. del Barrio, J. Cuesta Albertos, and C. Matrán. Center-outward distribution and quantile functions, ranks, and signs in dimension d: a measure transportation approach. *Ann. Statist., to appear*, 2020a.

M. Hallin, D. Hlubinka, and Š Hudecová. Fully distribution-free center-outward rank tests for multiple-output regression and MANOVA. *http://arxiv.org/abs/2007.15496*, 2020b.

A. Hatcher. *Algebraic Topology*. Cambridge University Press, Cambridge, 2002.

X. He and G. Wang. Convergence of depth contours for multivariate datasets. *Ann. Statist.*, 25:495–504, 1997.

R.J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80:309–323, 1995.

R.J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128:153–179, 1997.

H. Oja. Affine-invariant multivariate sign and rank tests and corresponding estimates: a review. *Scand. J. Statist.*, 26:319–343, 1999.

H. Oja. *Multivariate Nonparametric Methods with R. An Approach Based on Spatial Signs and Ranks*, volume 199 of Lecture Notes in Statistics. Springer, New York, 2010.

D. Paindaveine and J. Virta. On the behavior of extreme d-dimensional spatial quantiles under minimal assumptions. *arXiv:2001.10877*, 2020.

M.L. Puri and P.K. Sen. *Nonparametric Methods in Multivariate Analysis*. John Wiley & Sons, Inc., New York-London-Sydney, 1971.

R.T. Rockafellar. *Convex Analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

P.J. Rousseeuw and I. Ruts. The depth function of a population distribution. *Metrika*, 49: 213–244, 1999.

R. Serfling. Quantile functions for multivariate analysis: approaches and applications. *Stat. Neerl.*, 56:214–232, 2002.

Robert Serfling. Perspectives on depth functions on general data spaces, with consideration of the Tukey, projection, spatial, “density”, “local”, and “contour” depths. *http://www.utdallas.edu/ serfling/papers/I.Perspectives.pdf*, 2019.

H. Shi, M. Drton, and F. Han. Distribution-free consistent independence tests via Hallin’s multivariate rank. *https://arxiv.org/abs/1909.10024*, 2019.

H. Shi, M. Hallin, M. Drton, and F. Han. Rate-optimality of consistent distribution-free tests of independence based on center-outward ranks and signs. *arXiv:2007.02186*, 2020.

Cédric Villani. *Topics in optimal transportation*, volume 58 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2003.

Yijun Zuo and Xuming He. On the limiting distributions of multivariate depth-based rank sum statistics and related tests. *Ann. Statist.*, 34:2879–2896, 2006.

CHAPTER 7

Nonparametric Multiple-Output Center-Outward Quantile Regression

The content of this chapter is fully available online as del Barrio et al. (2022).

Contents

Based on the novel concept of multivariate center-outward quantiles introduced recently in Chernozhukov et al. (2017) and Hallin et al. (2021), we are considering the problem of nonparametric multiple-output quantile regression. Our approach defines nested *conditional center-outward quantile regression contours* and *regions* with given conditional probability content irrespective of the underlying distribution; their graphs constitute nested *center-outward quantile regression tubes*. Empirical counterparts of these concepts are constructed, yielding interpretable empirical regions and contours which are shown to consistently reconstruct their population versions in the Pompeiu-Hausdorff topology. Our method is entirely non-parametric and performs well in simulations including heteroskedasticity and nonlinear trends; its power as a data-analytic tool is illustrated on some real datasets.

7.1 Introduction

7.1.1 Quantile regression, single- and multiple-output

Forty-five years after its introduction by Koenker and Bassett (1978), quantile regression—arguably the most powerful tool in the statistical study of the dependence of a variable of interest Y on covariates $\mathbf{X} = (X_1, \ldots, X_m)$—has become part of statistical daily practice, with countless applications in all domains of scientific research, from economics and social sciences to astronomy, biostatistics, and medicine. Unlike classical regression, which, somewhat narrowly, is focused on conditional means $\mathrm{E}[Y|\mathbf{X}]$, quantile regression indeed is dealing with the complete conditional distributions $\mathrm{P}_{\mathrm{Y}|\mathbf{X}=\mathbf{x}}$ of Y conditional on $\mathbf{X} = \mathbf{x}$. Building on that pioneering contribution, a number of quantile regression methods, parametric, semiparametric, and nonparametric, have been developed for an extremely broad range of statistical topics, including time series, survival analysis, instrumental variables, measurement errors, and functional data—to quote only a few. Sometimes, a simple parametrized regression model allows for a parametric approach, yielding, for instance, linear quantile regression. In most situations, however, parametric models are too rigid and a more agnostic nonparametric approach is in order. We refer to Koenker (2005) for an introductory text and to Koenker et al. (2018) for a comprehensive survey.

In single-output models (univariate variable of interest Y), this nonparametric approach is well understood and well studied, and the history of non-parametric estimation of conditional quantile functions goes back, at least, to the seminal paper by Stone (1977). The results are much scarcer, however, in the ubiquitous multiple-output case (d-dimensional variable of interest $\mathbf{Y}$, with $d > 1$), and the few existing ones are less satisfactory—the simple reason for this being the absence of a fully satisfactory concept of multivariate quantiles.

A major difficulty with quantiles in dimension $d > 1$, indeed, is the fact that $\mathbb{R}^d$, contrary to $\mathbb{R}$, is not canonically ordered. A number of attempts have been made to overcome that issue, the most remarkable of which is the theory of statistical depth. That theory has generated an abundant literature which we cannot summarize here—we refer to (Serfling and Zuo, 2000) or (Serfling, 2002, 2019a,b) for general expositions and authoritative surveys.

Several depth concepts coexist. The most popular of them is Tukey's halfspace depth (Tukey, 1975), but all depth concepts (including the most recent ones: see, e.g., Konen and Paindaveine (2022)) are sharing the same basic properties. Tukey's halfspace depth characterizes, for each distribution P over $\mathbb{R}^d$ (for simplicity, assume P to be Lebesgue-absolutely continuous) *depth regions* $\mathbb{D}_{\mathrm{P}}(\delta)$ (resp., *depth contours* $\mathcal{D}_{\mathrm{P}}(\delta)$) as collections of points with depth (relative to P) larger than or equal to δ (resp., equal to δ), $\delta \in (0, 1/2]$. Depth regions are convex, closed, and nested as δ increases, and have been proposed as notions of quantile regions and contours—an interpretation that is supported by the L_1 nature of Tukey depth (Hallin et al., 2010).

Among the merits of this interpretation is that it has imposed the idea that quantiles, in dimension $d \geq 2$, should rely on some center-outward ordering with central region of depth $\delta = 1/2$ rather than a southwest-northeast extension of the classical univariate "left-to-right" linear ordering of the real line. Unfortunately, depth regions fail to satisfy

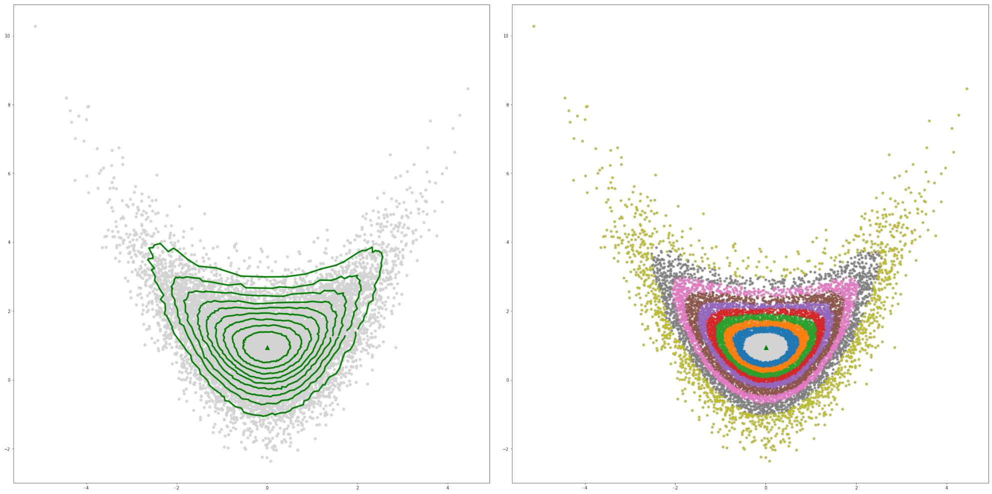

Figure 7.1.1: Quantile contours $\mathcal{C}_{\pm}(\tau)$ (left panel) and regions $\mathbb{C}_{\pm}(\tau)$ (right panel) for the banana-shaped Gaussian mixture of Section 7.4.1 and the quantile orders $\tau = 0.1,\ 0.2, \ldots, 0.9$.

the quintessential property of quantile regions: the P-probability content $\mathrm{P}\big[\mathbb{D}_{\mathrm{P}}(\delta)\big]$ of the quantile region $\mathbb{D}_{\mathrm{P}}(\delta)$ indeed very much depends on P. This is not a minor weakness: the univariate median $Y^{\mathrm{P}}_{1/2}$ of an absolutely continuous distribution P, for instance, is characterized by the fact that $\mathrm{P}\big[(-\infty, \mathrm{Y}^{\mathrm{P}}_{1/2}]\big] = 1/2$ irrespective of P—who would call *median* a quantity $Y^{\mathrm{P}}_{\mathrm{med}}$ such that $\mathrm{P}_1\big[(-\infty, Y^{\mathrm{P}_1}_{\mathrm{med}}]\big] = 0.4$ while $\mathrm{P}_2\big[(-\infty, Y^{\mathrm{P}_2}_{\mathrm{med}}]\big] = 0.6$? None of the depth concepts in the literature is satisfying that essential property of quantiles, though, and depth regions, therefore, cannot be considered as *bona fide* quantile regions.

A review of the various existing multiple-output quantile regression models (linear and nonparametric, depth-based and others) can be found in Hallin and Šiman (2018). A nonparametric quantile regression model based on a directional form of Tukey depth is developed in Hallin et al. (2015) but suffers the same lack of control over the probability contents of the quantile regions involved as the depth-based quantile concept itself. So does also the directional concept of M-quantiles proposed by Merlo et al. (2022).

Recently, based on measure transportation ideas, new concepts of quantiles in dimension $d > 1$ have been introduced in (Chernozhukov et al., 2017; Hallin, 2017; Hallin et al., 2021) under the names of *Monge-Kantorovich depth* and *center-outward quantile function*. Center-outward quantile functions define nested closed regions $\mathbb{C}_{\mathrm{P}}(\tau)$ and continuous contours $\mathcal{C}_{\mathrm{P}}(\tau)$ indexed by $\tau \in ([0, 1)$ such that, for any absolutely continuous P, $\mathrm{P}\big[\mathbb{C}_{\mathrm{P}}(\tau)\big] = \tau$ irrespective of P. Unlike the previous concepts, thus, and unlike the depth-based ones (as proposed in Hallin et al. (2010) or Kong and Mizera (2012)), these measure-transportation-based quantiles do satisfy the essential property that the P-probability contents of the resulting quantile regions do not depend on P. Moreover, the corresponding quantile regions are not necessarily convex and, as shown in Figure 7.1.1, they are able to capture the "shape" of the underlying distribution. We refer to Hallin (2022) for a survey of measure-transportation-based center-outward quantiles, the dual concepts of multivariate ranks and signs, and their many applications in inference problems (Ghosal and Sen (2019); Hallin et al. (2020); Shi et al. (2021a); Deb and Sen (2021); Shi et al. (2021b); Hallin et al. (2022b,a), among others).

Motivated by this long list of successful applications, we are proposing in this paper a novel and meaningful solution, based on the concept of center-outward quantiles, to the problem of nonparametric multiple-output quantile regression. Namely, for a pair of multidimensional random variables $(\mathbf{X}, \mathbf{Y})$ with values in $\mathbb{R}^m \times \mathbb{R}^d$ ($\mathbf{Y}$ the variable of interest, $\mathbf{X}$ the vector of covariates) and joint distribution[1] $\mathbb{P}$, we define (Section 7.2.2) the *center-outward quantile map* $\mathbf{Q}_\pm$ *of* $\mathbf{Y}$ *conditional on* $\mathbf{X} = \mathbf{x}$ as

$$\mathbf{u} \in \mathbb{S}_d \mapsto \mathbf{Q}_\pm(\mathbf{u}|\,\mathbf{x}) \in \mathbb{R}^d \tag{7.1}$$

($\mathbb{S}_d$ the open unit ball in $\mathbb{R}^d$), with the essential property that, letting

$$\mathbb{C}_\pm(\tau|\mathbf{x}) := \mathbf{Q}_\pm(\tau\,\overline{\mathbb{S}}_d|\mathbf{x}) \quad \tau \in (0,1), \;\; \mathbf{x} \in \mathbb{R}^m, \tag{7.2}$$

we have

$$\mathbb{P}\left[\mathbf{Y} \in \mathbb{C}_\pm(\tau|\,\mathbf{x})\big|\,\mathbf{X} = \mathbf{x}\right] = \tau \quad \text{for all } \mathbf{x} \in \mathbb{R}^m,\, \tau \in (0,1), \text{ and } \mathbb{P}, \tag{7.3}$$

justifying the interpretation of $\mathbf{x} \mapsto \mathbb{C}_\pm(\tau|\,\mathbf{x})$ as the value at $\mathbf{x}$ of a *regression quantile region of order* τ of $\mathbf{Y}$ with respect to $\mathbf{X}$. For $\tau = 0$,

$$\mathbb{C}_\pm(0\,|\,\mathbf{x}) := \bigcap_{\tau\in(0,1)} \mathbb{C}_\pm(\tau|\,\mathbf{x}) \tag{7.4}$$

yields the value at $\mathbf{X} = \mathbf{x}$ of the *regression median* $\mathbf{x} \mapsto \mathbb{C}_\pm(0\,|\,\mathbf{x})$ of $\mathbf{Y}$ with respect to $\mathbf{X}$. The same conditional quantile map characterizes nested (no "quantile crossing" phenomenon) *"regression quantile tubes of order* τ*"* (in $\mathbb{R}^{m+d}$)

$$\mathbb{T}_\pm(\tau) := \left\{\left(\mathbf{x}, \mathbf{Q}_\pm(\tau\,\overline{\mathbb{S}}_d\big|\,\mathbf{x})\right)\big|\mathbf{x} \in \mathbb{R}^m\right\}, \quad \tau \in (0,1) \tag{7.5}$$

which are such that

$$\mathbb{P}\left[(\mathbf{X}, \mathbf{Y}) \in \mathbb{T}_\pm(\tau)\right] = \tau \quad \text{irrespective of } \mathbb{P},\, \tau \in (0,1). \tag{7.6}$$

For $\tau = 0$, define

$$\mathbb{T}_\pm(0) := \left\{(\mathbf{x}, \mathbf{y}))\,\big|\mathbf{x} \in \mathbb{R}^m,\, \mathbf{y} \in \mathbb{C}_\pm(0\,|\,\mathbf{x})\right\} = \bigcap_{\tau\in(0,1)} \mathbb{T}_\pm(\tau)$$

(the *graph* of $\mathbf{x} \mapsto \mathbb{C}_\pm(\tau|\,\mathbf{x})$); with a slight abuse of language, also call $\mathbb{T}_\pm(0)$ the *regression median* of $\mathbf{Y}$ with respect to $\mathbf{X}$.

[1]For simplicity, in this introduction, we tacitly assume all distributions to be Lebesgue-absolutely continous.

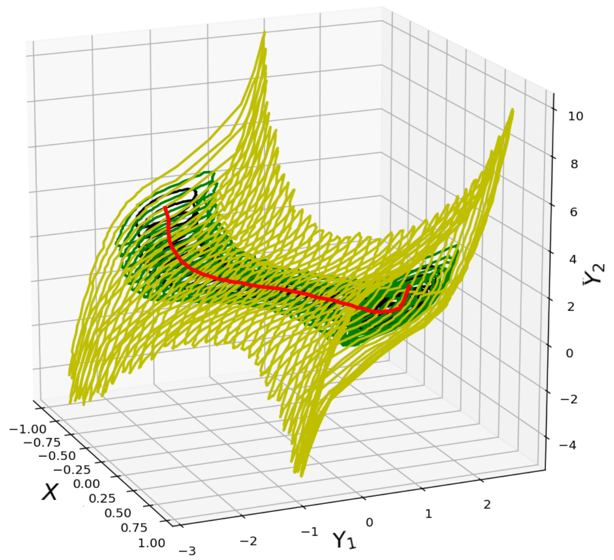

Figure 7.1.2: Quantile regression (two-dimensional variable of interest $\mathbf{Y}$; univariate regressor X), showing the conditional center-outward medians (red) and the conditional quantile contours $(\mathbf{x}, \mathbb{C}_{\pm}(\tau|\,\mathbf{x}))$of order $\tau = 0.2$ (black), $\tau = 0.4$ (green), and $\tau = 0.8$ (yellow).

None of the earlier attempts to define multiple-output regression quantiles—neither the depth-based definitions in Hallin et al. (2015), the directional concepts of marginal M-quantiles (Breckling and Chambers, 1988) considered by Merlo et al. (2022), nor the measure-transportation-based approach proposed by Carlier et al. (2016)—is characterizing quantile regions that satisfy requirements (7.3) and (7.6).

Carlier et al. (2016) deserves special attention, though, as the first attempt to break with directional and depth-based approaches to multiple-output quantile regression by means of innovative measure transportation ideas. Focusing on linear quantile regression, their method is based on a concept of multivariate quantile functions defined over the unit cube $[0, 1]^d$ rather than the unit ball $\mathbb{S}_d$. While yielding (under the assumption of a linear regression) an asymptotic reconstruction of the distributions of $\mathbf{Y}$ conditional on $\mathbf{X} = \mathbf{x}$, however, their choice of the unit cube does not directly allow for the definition of quantile regions of given order similar to the center-outward regression quantile regions $\mathbb{C}_{\pm}(\tau|\,\mathbf{x}) \subset \mathbb{R}^d$ or the quantile regression tubes $\mathbb{T}_{\pm}(\tau) \subset \mathbb{R}^{m+d}$.

As for the depth-based quantile regions, moreover, they are necessarily convex, even for distributions with highly non-convex shapes as in Figure 7.1.1. To circumvent this convexity problem, Feldman et al. (2021) propose a clever transformation of the data turning its distribution into a latent one with convex level sets by fitting a conditional variational auto-encoder (Sohn et al. (2015)). The probability contents of the quantile regions resulting from this machine-learning type of *"convexity reparation,"* however, remain out of control (they still depend on $\mathbb{P}$).

Figure 7.1.2 provides, for $m = 1$ and $d = 2$, a visualization of the regression median and quantile tubes of orders $\tau = 0.2$, 0.4, and 0.8, along with some of the corresponding *conditional regression quantile contours* $\mathcal{C}_{\pm}(\tau|\mathbf{x}) := \big\{\big(\mathbf{x}, \mathbf{Q}_{\pm}(\tau\, \mathcal{S}_{d-1}\big|\, \mathbf{x})\big)\big\}$ ($\mathcal{S}_{d-1}$ the unit sphere in $\mathbb{R}^d$), for the model

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \sin(\frac{2\pi}{3}X) + 0.575\, e_1 \\ \cos(\frac{2\pi}{3}X) + X^2 + \frac{e_2^3}{2.3} + \frac{1}{4}e_1 + 2.65\, X^4 \end{pmatrix} \tag{7.7}$$

where $X \sim \mathrm{U}_{[-1,1]}$, $\mathbf{e} = \sqrt{1 + \frac{3}{2}\sin(\pi X/2)^2}\mathbf{v}$, $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{Id})$, and X and $\mathbf{v}$ are mutually independent. Actually, since exact values cannot be obtained for these population concepts, a very large sample of $n = 400,040$ observations was generated from (7.7), and the consistent estimation procedure described in Section 7.3 was performed to obtain the picture. Note the non-convexity of the conditional contours for $\tau = 0.8$, the non-linearity of the regression median, and the marked heteroskedasticity of the regression.

7.1.2 Outline of the paper

The paper is organized as follows. Section 7.2 is dealing with the population concept of *conditional center-outward quantile map* and the resulting *center-outward regression quantile contours* and *regions*, Section 7.3 with their estimation. In Section 7.3.1, we show how to construct empirical quantile contours and regions, the consistency of which is established in Section 7.3. Numerical results are provided in Section 7.4. Monte Carlo experiments in Section 7.4.1 show the ability of our method to pick heteroskedasticity, nonlinear trends, and the shape of conditional distributions; comparisons also are made with the results of the depth-based method of Hallin et al. (2015). Real datasets are analyzed in Section 7.4.2, illustrating the power of our method as a data-analytic tool. Section 7.5 concludes with references to the recent literature on the numerical aspects of optimal transportation and perspectives for future developments.

7.2 Nonparametric center-outward quantile regression

7.2.1 Notation

For convenience, we are listing here the main notation to be used throughout the paper. Unless otherwise stated, we denote by $(\Omega, \mathcal{A}, \mathbb{P})$ the triple defining the underlying probability space. Let ℓ_d be the d-dimensional Lebesgue measure, $\mathcal{B}_d$ the Borel σ-field, and $\mathcal{P}(\mathbb{R}^d)$ the space of Borel probability measures on $\mathbb{R}^d$. The support of a probability $\mathrm{P} \in \mathcal{P}(\mathbb{R}^\mathrm{d})$ is denoted as $\mathrm{supp}(\mathrm{P})$; its closure as $\overline{\mathrm{supp}}(\mathrm{P})$. Throughout, $(\mathbf{X}, \mathbf{Y})$ denotes an $\mathbb{R}^{m+d}$-valued random vector with probability distribution $\mathbb{P} = \mathrm{P}_{\mathbf{X},\mathbf{Y}} \in \mathcal{P}(\mathbb{R}^{m+d})$, m-dimensional $\mathbf{X}$-marginal $\mathrm{P}_{\mathbf{X}}$ and d-dimensional $\mathbf{Y}$-marginal $\mathrm{P}_{\mathbf{Y}}$. The distribution of $\mathbf{Y}$ conditional on $\mathbf{X} = \mathbf{x}$ is denoted as $\mathrm{P}_{\mathbf{Y}\,|\mathbf{X}=\mathbf{x}}$[2]. The open unit ball, the closed unit ball, and the unit hypersphere in $\mathbb{R}^d$ are denoted by $\mathbb{S}_d$, $\overline{\mathbb{S}}_d$, and $\mathcal{S}_{d-1}$, respectively. We denote by U_d the spherical uniform over $\mathbb{S}_d$—that is, the product of a uniform distribution over the unit hypersphere $\mathcal{S}_{d-1}$ (for the directions) and a uniform distribution over $[0,1]$ (for the distance to the center).

7.2.2 Conditional center-outward quantiles, regions, and contours.

Let us provide precise definitions for the concepts we briefly presented in the Introduction and properly introduce center-outward quantiles, regions, and contours.

For any $\mathrm{P} \in \mathcal{P}(\mathbb{R}^{\mathrm{d}})$, denote by $\mathbf{Q}_\pm = \nabla\varphi$ and call *center-outward quantile map* the (Lebesgue-a.e.) unique gradient of a convex function $\varphi : \mathbb{S}_d \to \mathbb{R}$ such that $\mathbf{Q}_\pm(\mathbf{U}) \sim \mathrm{P}$, for any $\mathbf{U} \sim \mathrm{U}_d$—in the measure transportation convenient terminology, $\mathbf{Q}_\pm$ is *pushing* P *forward to* U_d, which we denote as $\mathbf{Q}_\pm \# \mathrm{P} = \mathrm{U}_\mathrm{d}$. This only defines $\mathbf{Q}_\pm$ at φ's points of differentiability (recall that convex functions are differentiable at almost every point in the interior of their domain: see Theorems 26.1 and 25.5 in Rockafellar (1970)). At φ's points of non-differentiability $\mathbf{u}$, let us define $\mathbf{Q}_\pm(\mathbf{u})$ as the subdifferential $\partial\varphi(\mathbf{u})$ of φ, namely,

$$\mathbf{Q}_\pm(\mathbf{u}) = \partial\varphi(\mathbf{u}) := \Big\{\mathbf{y} \in \mathbb{R}^d |\ \text{for all } \mathbf{z} \in \mathbb{R}^d,\ \varphi(\mathbf{z}) - \varphi(\mathbf{u}) \geq \langle \mathbf{y}, \mathbf{z} - \mathbf{u} \rangle \Big\}, \quad \mathbf{u} \in \mathbb{S}_d;$$

then, $\mathbf{Q}_\pm$ is an everywhere-defined set-valued function. Slightly abusing the notation, we also write $\partial\varphi$ for the set of all points $(\mathbf{u}, \mathbf{y}) \in \mathbb{R}^{m+d}$ such that $\mathbf{y} \in \partial\varphi(\mathbf{u})$. With this notation, we can introduce the concepts of conditional center-outward quantiles, contours, and regions.

Definition 7.2.1. *Call* conditional center-outward quantile map of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$ *the center-outward quantile map* $\mathbf{u} \mapsto \mathbf{Q}_\pm(\mathbf{u}|\mathbf{X} = \mathbf{x})$, $\mathbf{u} \in \mathbb{S}_d$ *of* $\mathrm{P}_{\mathbf{Y}\,|\mathbf{X}=\mathbf{x}}$, $\mathbf{x} \in \mathbb{R}^m$. *The corresponding* conditional center-outward quantile regions and contours of order $\tau \in (0,1)$ *are the sets*

$$\mathbb{C}_\pm(\tau\,\big|\mathbf{x}) := \mathbf{Q}_\pm\left(\tau\,\overline{\mathbb{S}}_d|\mathbf{X} = \mathbf{x}\right) \ \textit{and}\ \mathcal{C}_\pm(\tau\,\big|\mathbf{x}) := \mathbf{Q}_\pm\left(\tau\mathcal{S}_{d-1}\,|\mathbf{X} = \mathbf{x}\right),$$

respectively. The conditional center-outward quantile maps also characterize (see Definitions (7.4), (7.5), and (7.6)) conditional medians $\mathbb{C}_\pm(0\,|\,\mathbf{x})$ *and regression quantile tubes* $\mathbb{T}_\pm(\tau)$.

When no confusion is possible, we also write $\mathbf{Q}_\pm(\mathbf{u}|\mathbf{x})$ for $\mathbf{Q}_\pm(\mathbf{u}|\mathbf{X} = \mathbf{x})$. The terminology center-outward *regression* quantile region, contour, and median is used for the mappings $\mathbf{x} \mapsto \mathbb{C}_\pm(\tau\,\big|\mathbf{x})$, $\mathbf{x} \mapsto \mathcal{C}_\pm(\tau\,\big|\mathbf{x})$, and $\mathbf{x} \mapsto \mathbb{C}_\pm(0\,|\,\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^m$.

Recall, however, that, in the absence of any assumptions on the conditional probabilities $\mathrm{P}_{\mathbf{Y}\,|\mathbf{X}=\mathbf{x}}$, the mappings $\mathbf{u} \mapsto \mathbf{Q}_\pm(\mathbf{u}\,|\mathbf{X} = \mathbf{x})$ typically are set-valued, see Rockafellar and Wets (1998). Whenever continuous, single-valued functions (typically, on the punctured unit ball $\mathbb{S}_d \setminus \{\mathbf{0}\}$) are needed, we will make the following assumption.

[2]The existence of the regular conditional probability is a direct consequence of the disintegration theorem (see, e.g., Theorem 2.5.1 in Lehmann and Romano (2005)).

Assumption (R) *For* $\mathrm{P}_{\mathbf{X}}$*-a.e.* $\mathbf{x} \in \mathbb{R}^m$*, the conditional distribution* $\mathrm{P}_{\mathbf{Y}\,|\mathbf{X}=\mathbf{x}}$ *admits, with respect to the Lebesgue measure, a density* $\mathrm{p}_{\mathbf{Y}\,|\mathbf{X}=\mathbf{x}}$ *with convex support* $\mathrm{supp}(\mathrm{P}_{\mathbf{Y}\,|\mathbf{X}=\mathbf{x}})$*; moreover, for every* $R > 0$*, there exist constants* $0 < \lambda_R^{\mathbf{x}} \leq \Lambda_R^{\mathbf{x}} < \infty$ *such that*

$$\lambda_R^{\mathbf{x}} \leq \mathrm{p}_{\mathbf{Y}\,|\mathbf{X}=\mathbf{x}}(\mathbf{y}) \leq \Lambda_R^{\mathbf{x}} \quad \textit{for all } \mathbf{y} \in \mathrm{supp}(\mathrm{P}_{\mathbf{Y}\,|\mathbf{X}=\mathbf{x}}) \cap \mathrm{R}\,\mathbb{S}_{\mathrm{d}}.$$

In the classical single-output case ($d = 1$), consistent estimation of conditional quantiles similarly requires the continuity of the conditional quantile maps (see Stone (1977)). In dimension $d > 1$, the continuity of center-outward quantile maps follows from assumptions similar to Assumption (R)—see Figalli (2018) and del Barrio et al. (2020).

7.3 Empirical center-outward quantile regression

We now proceed with the construction of empirical versions of the conditional center-outward quantile concepts defined in Section 7.2.2 and their consistency properties.

7.3.1 Empirical conditional center-outward quantiles

Let $(\mathbf{X}, \mathbf{Y})^{(n)} := \big((\mathbf{X}_1, \mathbf{Y}_1), \ldots, (\mathbf{X}_n, \mathbf{Y}_n)\big)$ be a sample of n i.i.d. copies of $(\mathbf{X}, \mathbf{Y}) \sim= \mathrm{P}_{\mathbf{XY}}$. In this section, we develop an estimator of the conditional center-outward quantile maps $\mathbf{u} \mapsto \mathbf{Q}_{\pm}(\mathbf{u}|\mathbf{X} = \mathbf{x})$, $\mathbf{x} \in \mathbb{R}^m$. Our estimator is obtained in two steps: in Step 1, we construct an empirical distribution of $\mathbf{Y}$ conditional on $\mathbf{X} = \mathbf{x}$ and, in Step 2, we compute the corresponding empirical center-outward quantile map.

Step 1. For each value of $\mathbf{x} \in \mathbb{R}^m$, our estimation of the conditional distribution of $\mathbf{Y}$ conditional on $\mathbf{X} = \mathbf{x}$ involves a sequence of *weight functions* $w^{(n)} : \mathbb{R}^{m(n+1)} \to \mathbb{R}^n$ measurable with respect to $\mathbf{x}$ and the sample $\mathbf{X}^{(n)} := (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ of $\mathbf{X}$ observations, of the form

$$(\mathbf{x},\, \mathbf{X}^{(n)}) \mapsto w^{(n)}\left(\mathbf{x},\, \mathbf{X}^{(n)}\right) := \left(w_1(\mathbf{x}; \mathbf{X}^{(n)}), \ldots, w_n(\mathbf{x}; \mathbf{X}^{(n)})\right) \tag{7.8}$$

where $w_j^{(n)} : \mathbb{R}^{m(n+1)} \to \mathbb{R}$, $j = 1, \ldots, n$ are such that

$$w_j^{(n)}(\mathbf{x}; \mathbf{X}^{(n)}) \geq 0 \quad \text{and} \quad \sum_{j=1}^{n} w_j^{(n)}(\mathbf{x}; \mathbf{X}^{(n)}) = 1 \quad \text{a.s. for all } n. \tag{7.9}$$

We refer to a function $w^{(n)}$ of the form (7.8) satisfying (7.9) as a *probability weight function* and define the *empirical conditional distribution* of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$ as

$$\mathrm{P}^{(n)}_{w(\mathbf{x})} := \sum_{j=1}^{n} w_j^{(n)}(\mathbf{x}; \mathbf{X}^{(n)}) \delta_{\mathbf{Y}_j}, \tag{7.10}$$

where $\delta_{\mathbf{Y}_j}$ is the Dirac function computed at $\mathbf{Y}_j$. Following Stone (1977), we say that the sequence $w^{(n)}$ is a *consistent* weight function if, whenever $(\mathbf{X}, Y), (\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ are i.i.d., where Y is real-valued and such that $\mathrm{E}|Y|^r < \infty$ for $r > 1$,

$$\mathrm{E}\left|\sum_{j=1}^{n} w_j^{(n)}(\mathbf{X}; \mathbf{X}^{(n)}) Y_j - \mathrm{E}(Y|\mathbf{X})\right|^r \longrightarrow 0 \quad \text{as } n \to \infty. \tag{7.11}$$

Step 2. To estimate the conditional quantiles, consider a *regular grid* $\mathfrak{G}^{(N)}$ of $\mathbb{U}_d$ consisting of N gridpoints denoted as $\mathfrak{s}_1^{(N)}, \ldots, \mathfrak{s}_N^{(N)}$. The number N here is arbitrarily chosen as factorizing into a product of integers of the form $N = N_R N_S + N_0$ with $N_0 = 0$ or 1. That regular grid is created as the intersection between

- the rays generated by an N_S-tuple $\mathbf{u}_1, \ldots, \mathbf{u}_{N_S} \in \mathcal{S}_{d-1}$ of unit vectors such that ${N_S}^{-1} \sum_{j=1}^{N_S} \delta_{\mathbf{u}_j}$ converges weakly to the uniform over $\mathcal{S}_{d-1}$ as $N_S \to \infty$, and
- the N_R hyperspheres with center $\mathbf{0}$ and radii $j/(N_R + 1)$, $j = 1, \ldots, N_R$,

along with the origin if $N_0 = 1$. Based on this grid, we define the sequence of discrete uniform measures

$$\mathrm{U}_d^{(N)} := \frac{1}{N} \sum_{j=1}^{N} \delta_{\mathfrak{G}_j^{(N)}} \in \mathcal{P}(\mathbb{R}^d), \quad N \in \mathbb{N}$$

over $\mathfrak{G}^{(N)}$ and require that $N \to \infty$ with both $N_R \to \infty$ and $N_S \to +\infty$. By construction, $\mathrm{U}_d^{(N)}$ converges weakly to U_d as $N \to \infty$. Note that imposing $N_0 = 0$ or 1 is not a problem, since N, unlike n, is chosen by the practitioner. Having $N_0 = 0$ or 1 yields the fundamental advantage that all points of $\mathfrak{G}^{(N)}$ have multiplicity one and that Corollary 3.1 in Hallin et al. (2021), to be used below, applies.

Our estimation of the conditional center-outward quantile maps relies on the optimal transport pushing $\mathrm{U}_d^{(N)}$ forward to $\mathrm{P}_{w(\mathbf{x})}^{(n)}$—more precisely, adopting (since typically $N \neq n$) the Kantorovich formulation of the optimal transport problem, on the solution of the linear program (solvable using efficient numerical methods such as the auction or Hungarian algorithms—see PeyrÃ© and Cuturi (2019) and references therein)[3]

$$\begin{aligned}
&\min_{\pi := \{\pi_{i,j}\}} \sum_{i=1}^{N} \sum_{j=1}^{n} \frac{1}{2} \left|\mathbf{Y}_j - \mathfrak{s}_i\right|^2 \pi_{i,j}\ , \\
&\text{s.t.} \ \sum_{j=1}^{n} \pi_{i,j} = \frac{1}{N}, \quad i \in \{1, 2, \ldots, N\}, \\
&\sum_{i=1}^{N} \pi_{i,j} = w_j^{(n)}(\mathbf{x}; \mathbf{X}^{(n)}), \quad j \in \{1, 2, \ldots, n\}, \\
&\pi_{i,j} \geq 0, \quad i \in \{1, 2, \ldots, N\},\ j \in \{1, 2, \ldots, n\}.
\end{aligned} \tag{7.12}$$

[3] We are dropping the superscripts $^{(N)}$ and $^{(n)}$ when no confusion is possible.

Here, any Nn-tuple $\pi := \{\pi_{i,j} | i = 1, \dots, N \; j = 1, \dots, n\}$ satisfying the constraints in (7.12) represents a transport plan from $\mathrm{U}_d^{(N)}$ to $\mathrm{P}_{w(\mathbf{x})}^{(n)}$—that is, a discrete distribution over $\mathbb{R}^d \times \mathbb{R}^d$ with marginals $\mathrm{U}_d^{(N)}$ and $\mathrm{P}_{w(\mathbf{x})}^{(n)}$. Let $\pi^*(\mathbf{x}) = \{\pi^*_{i,j}(\mathbf{x}) | i = 1, \dots, N \; j = 1, \dots, n\}$ be a solution of (7.12) (an optimal *transport plan*). Theorem 2.12(i) in Villani (2003) implies that its support $\mathrm{supp}(\pi^*(\mathbf{x})) := \{(\mathfrak{G}_i, \mathbf{Y}_j) | \pi^*_{i,j}(\mathbf{x}) > 0\}$ is *cyclically monotone*,[4] hence is contained in the graph of the *subdifferential* of a convex function. Therefore, the idea is to construct a smooth interpolation of $\pi^*(\mathbf{x})$ that maintains this property.

Note that, for any gridpoint $\mathfrak{G}_i$, $i \in \{1, \dots, N\}$, the constraints in (7.12) imply that there exists at least one $j \in \{1, \dots, n\}$ such that $(\mathfrak{G}_i, \mathbf{Y}_j) \in \mathrm{supp}(\pi^*(\mathbf{x}))$. Since more than one such j may exist, we choose the one which *"gets the highest mass"* from $\mathfrak{G}_i$, and in case of ties we choose the *"smallest"* one: namely, let

$$\mathbf{T}^*(\mathfrak{G}_i | \, \mathbf{x}) := \arg\inf \left\{ \| \mathbf{y} \| : \; \mathbf{y} \in \mathrm{conv}\Big(\{ \mathbf{Y}_J : \; J \in \arg\max_j \pi^*_{i,j}(\mathbf{x}) \} \Big) \right\}, \tag{7.13}$$

where $\mathrm{conv}(A)$ denotes the convex hull of a set A. Since $\mathrm{conv}\Big(\{\mathbf{Y}_J : \; J \in \arg\max_j \pi^*_{i,j}(\mathbf{x})\}\Big)$ is closed and convex in $\mathbb{R}^d$, (7.13) defines a unique $\mathbf{T}^*(\mathfrak{G}_i | \, \mathbf{x})$. Due to the cyclical monotonicity of $\mathrm{supp}(\pi^*(\mathbf{x}))$, there exists a convex function $\varphi^*(\cdot \, | \mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$ with subdifferential $\partial \varphi^*(\cdot \, | \mathbf{x})$ such that, for all $1 \le i \le N$,

$$\emptyset \neq \{ \mathbf{Y}_j : \; (\mathfrak{G}_i, \mathbf{Y}_j) \in \mathrm{supp}(\pi^*(\mathbf{x})) \} \subset \partial \varphi^*(\mathfrak{G}_i \, | \mathbf{x}).$$

Since sub-differentials are convex sets, this entails

$$\mathrm{conv}\{ \mathbf{Y}_j : \; (\mathfrak{G}_i, \mathbf{Y}_j) \in \mathrm{supp}(\pi^*(\mathbf{x})) \} \subset \partial \varphi^*(\mathfrak{G}_i \, | \mathbf{x}).$$

Consequently, $\{(\mathfrak{G}_i, \mathbf{T}^*(\mathfrak{G}_i | \, \mathbf{x})) : \; i = 1, \dots, N\}$ is cyclically monotone and satisfies the assumptions of Corollary 3.1 in Hallin et al. (2021). This implies the existence, for all $\mathbf{x}$, of a continuous cyclically monotone map $\mathbf{u} \mapsto \mathbf{Q}^{(n)}_{w,\pm}(\mathbf{u} \, | \mathbf{x})$, say, interpolating the N-tuple $(\mathfrak{G}_1, \mathbf{T}^*(\mathfrak{G}_1 | \, \mathbf{x})), \dots, (\mathfrak{G}_N, \mathbf{T}^*(\mathfrak{G}_N | \, \mathbf{x}))$, i.e., such that $\mathbf{Q}^{(n)}_{w,\pm}(\mathfrak{G}_i | \, \mathbf{x}) = \mathbf{T}^*(\mathfrak{G}_i | \, \mathbf{x})$ for $i = 1, \dots, N$.

In particular, we proceed as in Hallin et al. (2021) by choosing the smooth cyclically monotone interpolation with largest Lipschitz constant. This continuous map $\mathbf{u} \mapsto \mathbf{Q}^{(n)}_{w,\pm}(\mathbf{u} | \mathbf{x})$ from $\mathbb{S}_d$ to $\mathbb{R}^d$ will be called the *empirical conditional center-outward quantile function* of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$. It defines the *empirical center-outward regression quantile regions* and *contours*

$$\mathbb{C}^{(n)}_{w,\pm}(\tau \, | \mathbf{x}) := \mathbf{Q}^{(n)}_{w,\pm}\left(\tau \, \overline{\mathbb{S}}_d \, | \mathbf{x}\right) \;\; \text{and} \;\; \mathcal{C}^{(n)}_{w,\pm}(\tau \, | \mathbf{x}) := \mathbf{Q}^{(n)}_{w,\pm}\left(\tau \mathcal{S}_{d-1} \, | \mathbf{x}\right), \quad \tau \in (0,1) \tag{7.14}$$

[4] Recall from Rockafellar (1970) that a set $S \subset \mathbb{R}^d \times \mathbb{R}^d$ is *cyclically monotone* if any finite subset $\{(x_{k_1}, y_{k_1}), \dots, (x_{k_\nu}, y_{k_\nu})\} \subset S$, $\nu \in \mathbb{N}$ satisfies $\sum_{\ell=1}^{\nu-1} \langle y_{k_\ell}, x_{k_{\ell+1}} - x_{k_\ell} \rangle + \langle y_{k_\nu}, x_{k_1} - x_{k_\nu} \rangle \le 0$, where $\langle \cdot, \cdot \rangle$ stands for the scalar product in $\mathbb{R}^d$.

which we are proposing as estimators of $\mathbb{C}_{\pm}(\tau\,|\mathbf{x})$ and $\mathcal{C}_{\pm}(\tau\,|\mathbf{x})$, respectively. The intersection $\bigcap_{\tau\in(0,1)}\mathbb{C}^{(n)}_{w,\pm}(\tau\,|\mathbf{x})$ yields the *empirical conditional center-outward regression median region*. The definition of *empirical regression quantile tubes*

$$\mathbb{T}^{(n)}_{w,\pm}(\tau) := \left\{\left(\mathbf{x}, \mathbf{Q}^{(n)}_{\pm}(\tau\,\overline{\mathbb{S}}_d\big|\,\mathbf{x})\right)\big|\mathbf{x}\in\mathbb{R}^m\right\}, \quad \tau\in(0,1)$$

naturally follows.

Remark 7.3.1. Note that the results of this section and the next one still hold for any continuous map with cyclically monotone graph satisfying

$$(\mathbf{u}_i, \mathbf{Q}^{(n)}_{w,\pm}(\mathbf{u}_i\,|\mathbf{x})) \in \operatorname{conv}\left(\{\mathbf{Y}_j : \ (\mathbf{u}_i, \mathbf{Y}_j) \in \operatorname{supp}(\pi^*(\mathbf{x}))\}\right) \quad \text{for all } i=1,\ldots,N.$$

The reason for choosing the "smallest" $\mathbf{y}$ in (7.13) is to have a "universal criterion."

7.3.2 Consistency of empirical conditional center-outward quantiles, regression quantile regions, and regression quantile contours

The objective of this section is to justify the definitions of Section 7.3.1 by showing the consistency of the empirical quantile regions and contours defined in (7.14). The asymptotic behavior of these regions and contours, quite naturally, depends on the regularity of the conditional distributions involved. In fact, as discussed before, when Assumption (R) does not hold, the population conditional quantile maps are not necessarily defined for every $\mathbf{u}\in\mathbb{S}_d$, but only for a set of U_d-probability one. Consistency results can be obtained despite this a.s. definition, provided that population quantile maps are extended into set-valued maps. The following theorem shows[5], under mild assumptions, that any possible limit of $\mathbf{Q}^{(n)}_{w,\pm}(\,\cdot\,|\mathbf{x})$ asymptotically belongs to the set $\mathbf{Q}_{\pm}(\,\cdot\,|\mathbf{X}=\mathbf{x})$.

Theorem 7.3.2. *Let* $(\mathbf{X},\mathbf{Y}),(\mathbf{X}_1,\mathbf{Y}_1),\ldots,(\mathbf{X}_n,\mathbf{Y}_n)$ *be pairs of i.i.d. random vectors with values in* $\mathbb{R}^m\times\mathbb{R}^d$ *and let* $w^{(n)}$ *be a consistent sequence of weight functions. Then, for every* $\mathbf{u}\in\mathbb{S}_d$ *and* $\epsilon>0$*,*

$$\mathbb{P}\left(\,\mathbf{Q}^{(n)}_{w,\pm}(\mathbf{u}\,|\mathbf{X}) \notin \mathbf{Q}_{\pm}(\mathbf{u}\,|\mathbf{X}) + \epsilon\mathbb{S}_d\,\right) \longrightarrow 0 \quad \text{as } n \text{ and } N\to\infty,$$

and, for every $\tau\in(0,1)$*,*

$$\mathbb{P}\left(\,\mathbb{C}^{(n)}_{\pm}(\tau\,|\mathbf{X}) \not\subset \mathbb{C}_{\pm}(\tau\,|\mathbf{X}) + \epsilon\mathbb{S}_d\,\right) \to 0 \text{ and } \mathbb{P}\left(\,\mathcal{C}^{(n)}_{\pm}(\tau\,|\mathbf{X}) \not\subset \mathcal{C}^{(n)}_{\pm}(\tau\,|\mathbf{X}) + \epsilon\mathbb{S}_d\,\right) \to 0$$

as n *and* $N\to\infty$*.*

Neater convergence results—avoiding the notion of set-valued maps— are obtained if it can be assumed that Assumption (R) holds, which implies that, for any $\mathbf{X}=\mathbf{x}$ and

[5] Note that, although N does not appear in the notation, $\mathbf{Q}^{(n)}_{w,\pm}$ depends on both N and n.

any $\mathbf{u} \in \mathbb{S}_d \setminus \{\mathbf{0}\}$, the set $\mathbf{Q}_\pm(\mathbf{u}\,|\mathbf{X} = \mathbf{x})$ is a singleton. Then, the map $\mathbf{u} \mapsto \mathbf{Q}_\pm(\mathbf{u}\,|\mathbf{X} = \mathbf{x})$ can be seen as continuous on $\mathbb{S}_d \setminus \{\mathbf{0}\}$, see Theorem 25.5 in Rockafellar (1970), hence single-valued on $\mathbb{S}_d \setminus \{\mathbf{0}\}$ since the gradient of a convex function is single-valued at a point if and only if it is continuous at this point. We then can state the following theorem (see Appendix A.1 for the proof), the second part of which describes the convergence of the contours in terms of the Pompeiu-Hausdorff distance d_∞. Recall that he Pompeiu-Hausdorff distance between two sets A and B in $\mathbb{R}^d$ is defined as

$$d_\infty(A, B) \coloneqq \inf\{\nu \geq 0 : \; A \subset B + \nu\mathbb{S}_d \text{ and } B \subset A + \nu\mathbb{S}_d\}$$

(see Rockafellar and Wets (1998)).

Theorem 7.3.3. *Let $(\mathbf{X}, \mathbf{Y}), (\mathbf{X}_1, \mathbf{Y}_1), \ldots, (\mathbf{X}_n, \mathbf{Y}_n)$ be pairs of i.i.d. random vectors with values in $\mathbb{R}^m \times \mathbb{R}^d$ and let $w^{(n)}$ be a consistent sequence of weight functions. Suppose moreover that Assumption (R) holds. Then, for every compact $K \subset \mathbb{S}_d \setminus \{\mathbf{0}\}$, as n and $N \to \infty$,*

$$\sup_{\mathbf{u} \in K} |\mathbf{Q}^{(n)}_{w,\pm}(\mathbf{u}\,|\mathbf{X}) - \mathbf{Q}_\pm(\mathbf{u}\,|\mathbf{X})| \xrightarrow{\mathbb{P}} 0$$

and, for every $\tau \in (0, 1)$ and $\epsilon > 0$,

$$\mathbb{P}\left(d_\infty\left(\mathcal{C}^{(n)}_\pm(\tau\,|\mathbf{X}), \mathcal{C}_\pm(\tau\,|\mathbf{X})\right) > \epsilon\right) \to 0.$$

Under the assumptions of Theorem 7.3.3, consistency in Pompeiu-Hausdorff distance of the quantile contours holds in case the median is a single point[6]—the continuity of quantile maps then extends to the whole open unit ball. This, however, is not necessarily the case for $d > 3$ (see Figalli (2018)), and Pompeiu-Hausdorff consistency may fail due to the fact that our empirical version is continuous over $\mathbb{S}_d$ while $\mathbf{Q}_\pm(\mathbf{0}|\mathbf{x})$ could be a set rather than a single point: convergence then holds along subsequences of $\mathbf{Q}^{(n)}_{w,\pm}(\,\cdot\,|\mathbf{x})$ towards an element of $\mathbf{Q}_\pm(\mathbf{0}|\mathbf{x})$. This, obviously, has an impact on convergence in terms of the Pompeiu-Hausdorff distance—although it does not affect the control over the asymptotic probability contents of quantile regions. More precisely, the following corollary holds (see Appendix A.1 for the proof).

Corollary 7.3.4. *Under the conditions of Theorem 7.3.3, as n and $N \to \infty$,*

$$\mathbb{P}\left(\mathbf{Y} \in \mathbb{C}^{(n)}_\pm(\tau\,|\mathbf{X})\big|\mathbf{X}\right) \xrightarrow{\mathbb{P}} \tau \quad \text{for all } \tau \in (0, 1). \tag{7.15}$$

Remark 7.3.5. *Under the weaker conditions of Theorem 7.3.2, in view of (7.35), still some asymptotic control of the probability content of the empirical regions can be derived. More precisely, letting $N = N(n)$ be such that $N(n) \to \infty$ as $n \to \infty$, for all $\tau \in (0, 1)$ and every subsequence n_k, there exists a further subsequence n_{k_j} such that*

$$\limsup_m \mathbb{P}\left(\mathbb{C}^{(n_{k_j})}_\pm(\tau\,|\mathbf{x})\big|\mathbf{X} = \mathbf{x}\right) \leq \tau, \;\; \mathbf{x}\text{-a.e. in } \mathbb{R}^m.$$

[6] This is always the case for $d = 2$ and $d = 3$: see Figalli (2018).

The above results, as well as the proposed regularization, are valid for any consistent sequence of weight functions. This includes—along with adequate additional assumptions—most of the classic choices of weight functions. Here are three examples.

(i) The kernel weight function, usually defined (see chapter 5 in GyÃ¶rfi et al. (2002)) as

$$w_i^{(n)}(\mathbf{x}; \mathbf{X}^{(n)}) := K\left(\frac{\mathbf{X_i} - \mathbf{x}}{h_n}\right) / \sum_{j=1}^{n} K\left(\frac{\mathbf{X_j} - \mathbf{x}}{h_n}\right), \quad i = 1, \ldots, n$$

where h_n is the bandwidth, and $K : \mathbb{R}^m \to \mathbb{R}$ the kernel. Sufficient conditions for $w_i^{(n)}$ to form a consistent sequence of weight functions are

(a) $h_n \to 0$,
(b) $c_1 \min\left(\mathbb{1}_{[|\mathbf{x}| \leq r]}, H(|\mathbf{x}|)\right) \leq K(\mathbf{x}) \leq c_2 H(|\mathbf{x}|)$, where c_1, c_2, r are positive constants, and $H : [0, \infty) \to \mathbb{R}$ is bounded, decreasing, and such that $H(t)t^m \to 0$ as $t \to \infty$, and
(c) $\lim_{n\to\infty} n^\alpha h_n^m / \log(n) = \infty$ for any $\alpha \in (0, 1)$;

when the kernel K is compactly supported, the assumptions are are much simpler, and we only need (a) and $\lim_{n\to\infty} h_n^m n = \infty$, see Theorem 5.1 in GyÃ¶rfi et al. (2002). The particular case $K(\mathbf{x}) = e^{-|\mathbf{x}|^2}$ is known as the Gaussian kernel.

(ii) The (classical) $k-$nearest neighbors weight function: the $k-$nearest neighborhood of $\mathbf{x} \in \mathbb{R}^m$ is obtained (Chapter 6 in GyÃ¶rfi et al. (2002)) by ordering $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ according to increasing values of $|\mathbf{X}_j - \mathbf{x}|$. Denoting by $\{\mathbf{X}_{(0,\mathbf{x})}, \ldots, \mathbf{X}_{(n,\mathbf{x})}\}$ the reordered sequence, the set of $k-$nearest neighbors of $\mathbf{x}$ is $\mathcal{N}_k^{(n)}(\mathbf{x}) := \{\mathbf{X}_{(j,\mathbf{x})} : j \leq k\}$ and the k-nearest neighbors weight function is defined as

$$w_i^{(n)}(\mathbf{x}; \mathbf{X}^{(n)}) := \frac{1}{k} \mathbb{1}_{\mathbf{X}_i \in \mathcal{N}_k^{(n)}(\mathbf{x})}, \quad i = 1, \ldots, n.$$

Sufficient conditions for this k-nearest neighbors $w_i^{(n)}$ to form a consistent sequence of weight functions are (see Stone (1977))

$$k \to +\infty \text{ and } k/n \to 0. \tag{7.16}$$

If a k-nearest neighbors weight function $w^{(n)}$ is satisfying (7.16) and Assumption (R) holds, we thus have the convergence—in probability—of the conditional quantile map described in Theorem 7.3.3; without this assumption we still obtain the slightly weaker result of Theorem 7.3.2.

The $k-$nearest neighbors in (ii) were understood in the classical sense of the Euclidean distance in $\mathbb{R}^m$, which does not take into account the distribution $\mathrm{P}_\mathbf{X}$ of $\mathbf{X}$. An alternative $k-$nearest neighbors weight function can be derived from a notion of nearness based on the ordering induced by empirical center-outward distribution functions. This new weight function is obtained as follows.

(iii) An alternative k-nearest neighbors weight function. Fixing $\mathbf{x} \in \mathbb{R}^m$, first compute, as in Hallin et al. (2021), the empirical center-outward distribution function associated with

$$\frac{1}{n+1}\sum_{j=1}^{n}\delta_{\mathbf{X}_j} + \frac{1}{n+1}\delta_{\mathbf{x}} \in \mathcal{P}(\mathbb{R}^m).$$

That distribution function is the solution $T^*_{\mathbf{x}}$ of the minimization problem

$$\min_{T\in\Gamma_{n+1}} \sum_{k=0}^{n} |\mathbf{X}_k - T(\mathbf{X}_k)|^2$$

where $\mathbf{X}_0 = \mathbf{x}$, Γ_{n+1} is the set of all bijections T between $\{\mathbf{x}, \mathbf{X}_1, \ldots, \mathbf{X}_n\}$ and a *regular grid* $\mathfrak{G}^{(n+1)}$ of $\mathbb{S}_m$, of the form described in Section 7.3.1, consisting of $(n+1)$ gridpoints denoted as $\mathfrak{G}_0, \mathfrak{G}_1, \ldots, \mathfrak{G}_n$, obtained via a factorization of n into a product of non-negative integers of the form $n+1 = n_R n_S + n_0$ with $n_0 < \min(n_R, n_S)$. That regular grid is created as the intersection between

- the rays generated by an n_S-tuple $\mathbf{u}_1, \ldots, \mathbf{u}_{n_S} \in \mathcal{S}_{m-1}$ of unit vectors such that ${n_S}^{-1}\sum_{j=1}^{n_S}\delta_{\mathbf{u}_j}$ converges weakly, as $n_S \to \infty$, to the uniform over $\mathcal{S}_{m-1}$ and
- the n_R hyperspheres with center $\mathbf{0}$ and radii $j/(n_R+1)$, $j = 1, \ldots, n_R$,

along with n_0 copies of the origin whenever $n_0 > 0$. Based on this grid, we define the sequence of discrete uniform measures

$$\mathrm{U}_d^{(n+1)} := \frac{1}{n+1}\sum_{j=1}^{n+1}\delta_{\mathfrak{G}_j} \in \mathcal{P}(\mathbb{R}^m) \quad N \in \mathbb{N}$$

This map is defined only at the $(n+1)$ points $\mathbf{x}, \mathbf{X}_1, \ldots, \mathbf{X}_n$, but, as in the previous section, it can be countinuously extended (see also Hallin et al. (2021)) to the whole space $\mathbb{R}^m$—call $\mathbf{F}^{(n)}_{\mathbf{x};\pm} : \mathbb{R}^m \to \mathbb{S}_m$ this extension—with the properties that $\mathbf{F}^{(n)}_{\mathbf{x};\pm}$ coincides, on $\{\mathbf{x}, \mathbf{X}_1, \ldots, \mathbf{X}_n\}$, with $T^*_{\mathbf{x}}$, is the gradient of a differentiable convex function with domain $\mathbb{R}^m$, and satisfies $\mathbf{F}^{(n)}_{\mathbf{x};\pm}(\mathbb{R}^m) \subset \overline{\mathbb{S}}_m$. We then define the *set of* k*−nearest center-outward neighbors* of $\mathbf{x}$ as

$$\mathcal{K}_k^{(n)}(\mathbf{x}) := \{\mathbf{X}_j : \mathbf{F}_{\pm}^{(n)}(\mathbf{X}_j) \in \mathcal{N}_k(\mathbf{F}^{(n)}_{\mathbf{x};\pm}(\mathbf{x}))\}, \tag{7.17}$$

where, for each $\mathbf{a} \in \mathbf{B}_m$ and $k \in \mathbb{N}$, $\mathcal{N}_k(\mathbf{a})$ denotes the set of k-nearest neighbors (in the sense of Euclidean distance) of $\mathbf{a}$. Based on this, define the *center-outward nearest neighbor weight function*

$$w_j^{(n)}(\mathbf{x}; \mathbf{X}^{(n)}) := \frac{1}{k}\mathbb{1}_{\mathbf{X}_j \in \mathcal{K}_k^{(n)}(\mathbf{x})}, \;\; j = 1, \ldots, k \tag{7.18}$$

and proceed as in Section 7.3.1 with the estimation (7.12) of the conditional quantile functions.

The next result shows that, for a suitable choice of $k = k(n)$, center-outward nearest neighbors weight functions form a consistent sequence of weights (see the appendix for a proof).

Lemma 7.3.6. *If $k = k(n)$ is such that $k(n) \to \infty$ and $k(n)/n \to 0$ as $n \to \infty$, the sequence of weight functions defined in (7.18) is consistent in the sense of (7.11).*

This means, in particular, that Theorem 7.3.2 applies when the weight function (7.18) is used under the assimptions of Lemma 7.3.6, and that the resulting estimators are consistent.

Theorems 7.3.2 and 7.3.3 provide *weak* (in probability) *consistency* results under minimal assumptions. For sequences of weights satisfying, as $n \to \infty$,

$$\sum_{j=1}^{n} w_j^{(n)}(\mathbf{X}; \mathbf{X}^{(n)}) Y_j \longrightarrow \mathrm{E}[Y|\mathbf{X}] \ \text{ a.s.} \tag{7.19}$$

(*strongly consistent* sequences), the conclusion in Theorem 7.3.2 can be upgraded to *strong* (almost sure) *consistency*. For the particular case of k-nearest neighbors, (7.19) and strong consistency hold if (7.16) is replaced with

$$k/\log(n) \to \infty \quad \text{and} \quad k/n \to 0, \tag{7.20}$$

see Devroye et al. (1994) and Devroye (1982).

7.4 Numerical results

This section is devoted to a numerical assessment of the performance of the estimation procedures described in Section 7.3. We first analyze (Section 7.4.1) some artificial datasets—including the motivating example of Hallin et al. (2015)—then turn (Section 7.4.2) to real-data cases. These examples showcase three important features of our estimators: their ability to capture heteroskedasticity, to deal (non-parametrically) with highly nonlinear regression, and to adapt to non-convex noise.

7.4.1 Simulated examples

Parabolic trend and periodic heteroskedasticity; spherical conditional densities.

We start with analyzing the motivating example given in Hallin et al. (2015). The model (with $m = 1$, $d = 2$) is

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X \\ X^2 \end{pmatrix} + \left(1 + \frac{3}{2}\sin\left(\frac{\pi}{2}X\right)^2\right)\mathbf{e}, \quad X \sim U_{[-2,2]} \text{ and } \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{Id}), \tag{7.21}$$

with X and $\mathbf{e}$ mutually independent. In this case, the population conditional (on $X = x$) quantile contours are circles with radii depending on x and can be computed exactly; trend is parabolic and heteroskedasticity periodic.

Figure 7.4.1 illustrates the convergence of our estimated contours to the population counterparts. Compared to Figure 1 in Hallin et al. (2015), our method produces less smooth contours, at least for smaller sample sizes. On the other hand, our method is able to capture non-convex contour shapes—something the method in Hallin et al. (2015) cannot,

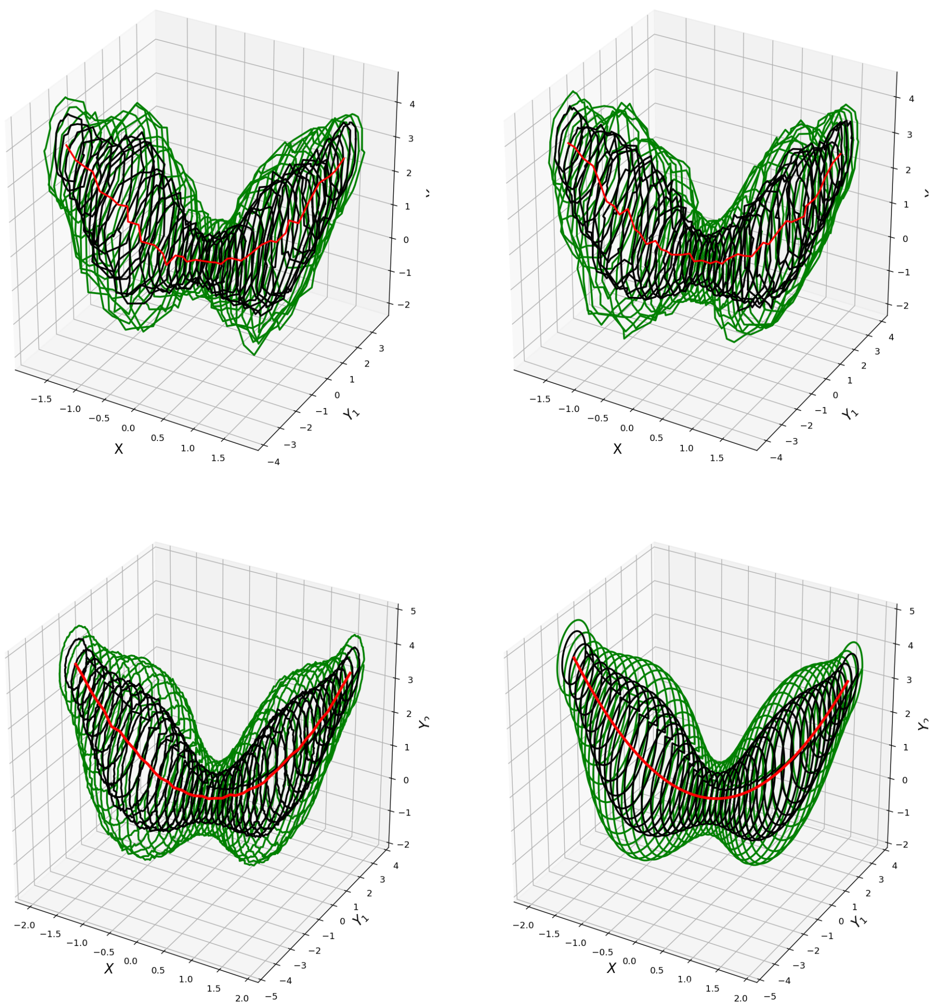

Figure 7.4.1: Estimated (sample sizes $3,601$ in the upper left panel, $10,000$ in the upper right panel, $128,020$ in the lower left panel) and population (lower right panel) quantile contours of order $\tau = 0.2$ (black) and 0.4 (green) for Model (7.21); the (estimated) conditional center-outward medians are shown in red. Estimations are based on the k-nearest neighbors weights (ii) with $N = k = 401, 1,000$, and $6,401$ and $n = 3,601, 10,000$, and $128,020$, respectively.

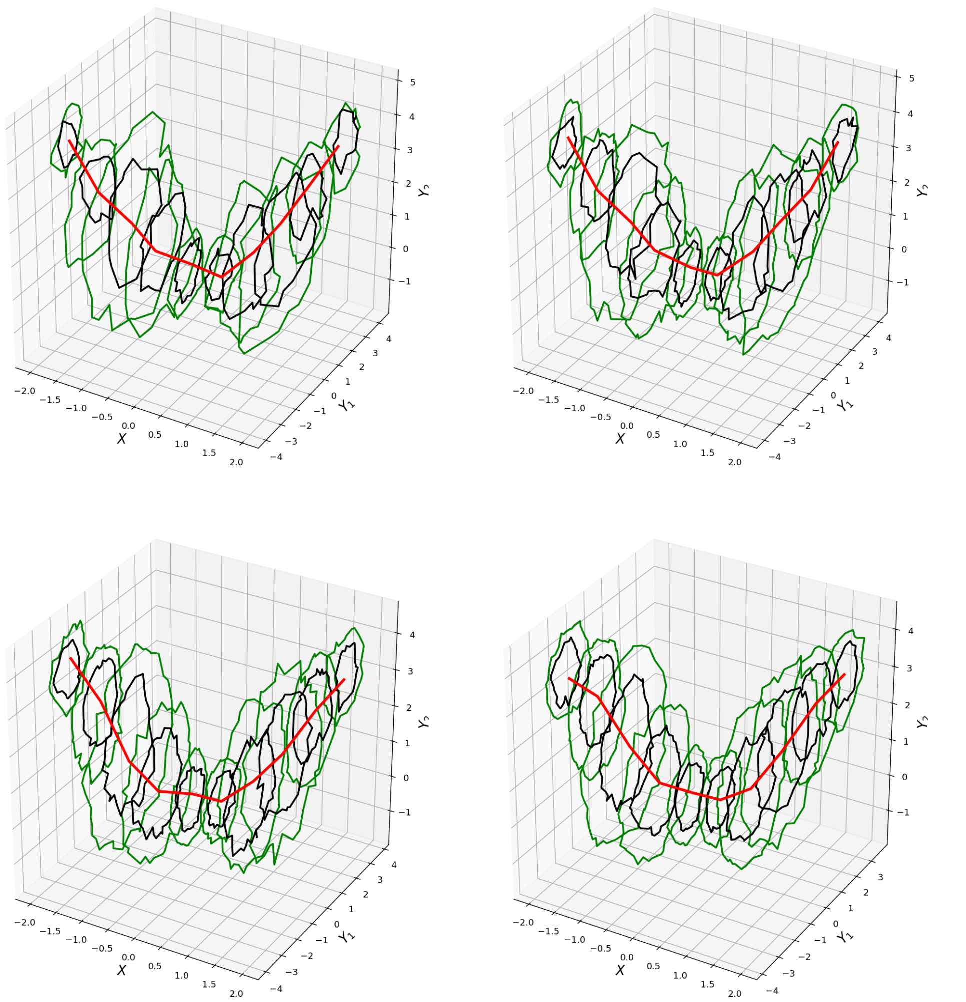

Figure 7.4.2: Estimated quantile contours (Model (7.21)) based on Gaussian kernel weight functions for different choices of the bandwith h. The sample size is $n = 3,601$; bandwidth values are $h = 0.05$ (upper left panel), $h = 0.1$ (upper right panel), $h = 0.2$ (lower left panel) and $h = 0.3$ (lower right panel). The estimated contour orders are $\tau = 0.2$ (black) and 0.4 (green); the estimated conditional center-outward medians are shown in red.

see Section 7.4.1. We also underline that, from a computational point of view, our method is able to handle rather large datasets (in contrast, the R packaged `modQR` cannot handle sample sizes over $10,000$, as explained in the documentation).

Model (7.21), as pointed out in Hallin et al. (2015), allows for testing the capacity of a method to estimate the trend while catching potential heteroskedasticity. A comparison with Figure 1 in Hallin et al. (2015) shows that both methods estimate the parabolic trend quite well, but that our method performs much better at capturing heteroskedasticity. Estimations

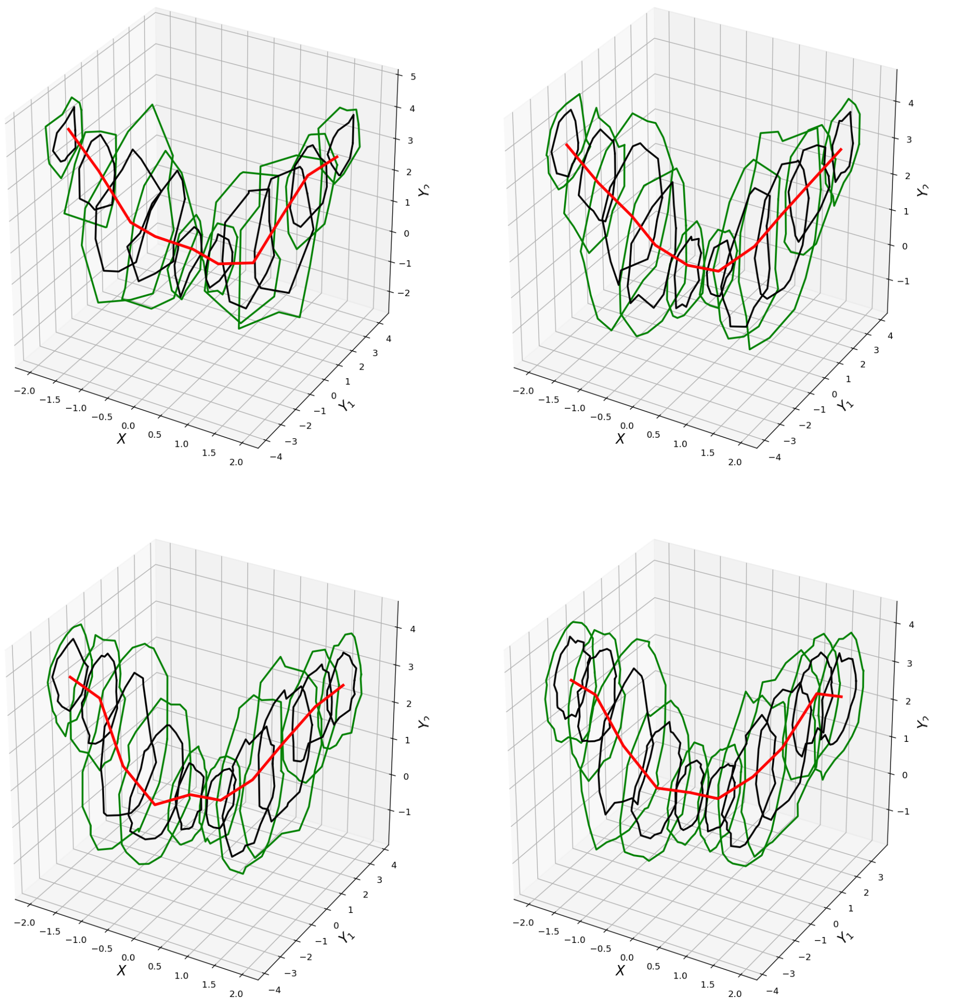

Figure 7.4.3: Estimated quantile contours (Model (7.21)) based on k-nearest neighbors weight functions for different choices of k. The sample size is $n = 3601$; $k = 101$ (upper left panel), $k = 256$ (upper right panel), $k = 401$ (lower left panel) and $k = 625$ (lower left panel). The estimated contour orders are $\tau = 0.2$ (black) and 0.4 (green); the estimated conditional center-outward medians are shown in red.

are based on the k-nearest neighbors weights (ii) of Section 7.3, with $N = k = 401, 1,000$, and $6,401$ and $n = 3,601, 10,000$, and $128,020$, respectively. Note that, in this univariate covariate case, the weight choices (ii) and (iii) coincide.

The performance of the Gaussian kernel (i) (with various bandwidth choices) and classical k-nearest neighbors (ii) (with various choices of k) weight functions are investigated in Figures 7.4.2 and 7.4.3, still for Model 7.21, with sample size $n = 3,601$. More precisely, estimation in Figure 7.4.2 is based on the weight function

$$w_i^{(n)}(\mathbf{x};\mathbf{X}^{(n)}) := e^{-\left(\frac{|\mathbf{X_i}-\mathbf{x}|}{h_n}\right)^2} \Big/ \sum_{j=1}^{n} e^{-\left(\frac{|\mathbf{X_j}-\mathbf{x}|}{h_n}\right)^2}, \quad i = 1,\ldots,n, \tag{7.22}$$

for $h_n = 0.05$, 0.1, 0.2, and 0.3; the discretization of the spherical uniform is based on a grid of size $N = n$. Estimation in Figure 7.4.3 is based on k-nearest neighbors weights with $k = 101$, 256, and 625. Here we chose $N = k$, which yields a one-to-one solution in (7.12). In both figures the conditional contours (of order $\tau = 0.2$ (black) and $\tau = 0.4$ (green)) and the estimated conditional center-outward medians (red) are shown for

$$x \in \{-2, -1.6, -1.1, -0.7, -0.2, 0.2, 0.7, 1.1, 1.6, 2\}.$$

For this sample size, the Gaussian kernel weights—due to the fact that they better exploit the information available on the x's—yield better results than the k-nearest neighbors ones. But the Gaussian kernel has a drawback for large datasets; the optimization problem (7.12) requires the whole dataset and cannot be efficiently computed. For instance, the Gaussian kernel counterpart of Figure 7.4.1 (where $n = 128,020$) cannot be computed on a standard desktop computer: large-sample datasets should be handled either with nearest neighbors or compactly supported kernel weights. On the other hand, the bandwith h_n and the neighborhood size k apparently have little impact on he result.

Parabolic trend and periodic heteroskedasticity; banana-shaped conditional densities.

We now consider a model in which the trend and heteroskedasticity are the same as in Model (7.21), but the quantile contours are non-convex (conditional densities are banana-shaped):

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X \\ X^2 \end{pmatrix} + \begin{pmatrix} \left(1 + \frac{3}{2}\sin\left(\frac{\pi}{2}X\right)^2\right) 1.15\, e_1 \\ \left(1 + \frac{3}{2}\sin\left(\frac{\pi}{2}X\right)^2\right)\left(\frac{e_2}{1.15} + 0.5(e_1^2 + 1.21)\right) \end{pmatrix}, \tag{7.23}$$

with $X \sim U_{[-2,2]}$ and $\mathbf{e} = (e_1, e_2) \sim \mathcal{N}(\mathbf{0}, \mathbf{Id})$, X and $\mathbf{e}$ mutually independent. The dataset shown in Figure 7.1.1 was generated from that model, with $X = 0$ and justifies the termonology "banana-shaped."

Population conditional quantile contours here cannot be computed analytically. Figure 7.4.4 shows their estimations (same method as in Section 7.4.1, with a k-nearest neighbors weight function, $k = 14,401$) for orders $\tau = 0.2$ (black), 0.4 (green), and 0.8 (yellow), along with the conditional medians (red). The sample size $n = 576,040$ is very large, so that, in view of consistency results, these estimations can be considered as close numerical approximations of their theoretical counterparts. This example illustrates further the ability of our method to handle vary large sample sizes. The Hallin et al. (2015) approach produces contours that are convex by construction—hence cannot capture the "banana shape" of the conditional densities. The same comparative analysis of weight functions is performed as for

Model (7.21). Figures 7.4.5 and 7.4.6 show the results for Gaussian kernel weights (various bandwiths) and k-nearest neighbors weights (various values of k), respectively. The sample size is $n = 3601$. In all the examples, for the ease of computation, $N = k$. Note that, for the k-nearest neighbor weights, this choice creates a one-to-one (between the sample and the grid) transport map.

Still for Model (7.23), Figure 7.4.7 shows, for sample size $n = 3601$ and various choices of the bandwidth h and the neighborhood size k, the behavior of the empirical conditional center-outward contours in $X = 0$—and compares them with those of Figure 7.4.4 (considered as the population contours). The empirical conditional quantiles are computed for $\tau = 0.2$, 0.4, and 0.8 with Gaussian kernel weights (bandwidths $h = 0.1$, 0.2, and 0.3) and the $k-$nearest neighbors weights ($k = 226$, 485, and 901). The influence of the choice of h and k is clearly seen here: the bigger h (the bigger k), the smoother the estimation of the shape of the contours but also, unfortunately, the worse the estimation of their location.

7.4.2 Some real-data examples

The CalCOFI oceanographic dataset: depth, temperature, and salinity in the oceans

The dataset "CalCOFI Over 60 Years of Oceanographic Data," available at `https://www.kaggle.com/sohier/calcofi`, contains the longest (1949-present) and most

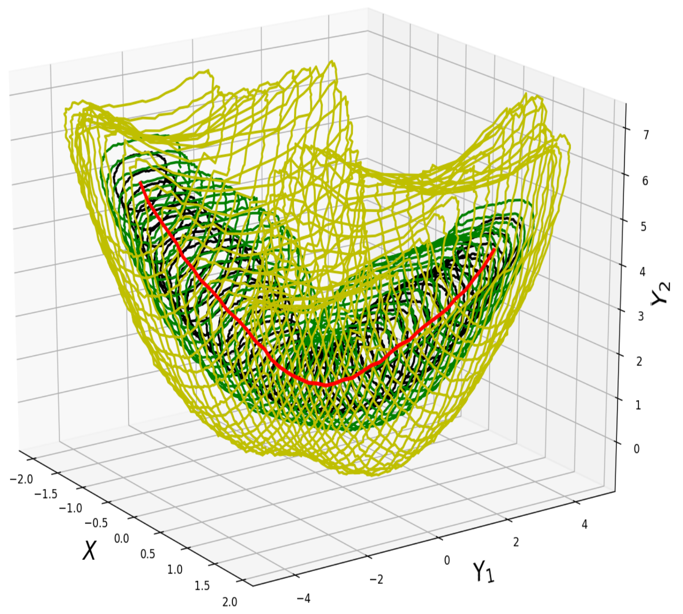

Figure 7.4.4: A numerical approximation (estimation based on a simulated sample of size $n = 576,040$) of the quantile contours of order $\tau = 0.2$ (black), 0.4 (green), and 0.8 (yellow) for Model (7.23); the conditional center-outward medians are shown in red.

complete (more than 50,000 sampling stations; sample size $n = 814,247$) time series of oceanographic and larval fish data worldwide. Data collected at depths down to 500 m include temperature, salinity, oxygen, phosphate, silicate, nitrate and nitrite, chlorophyll, transmissometer, PAR, C14 primary productivity, phytoplankton biodiversity, zooplankton

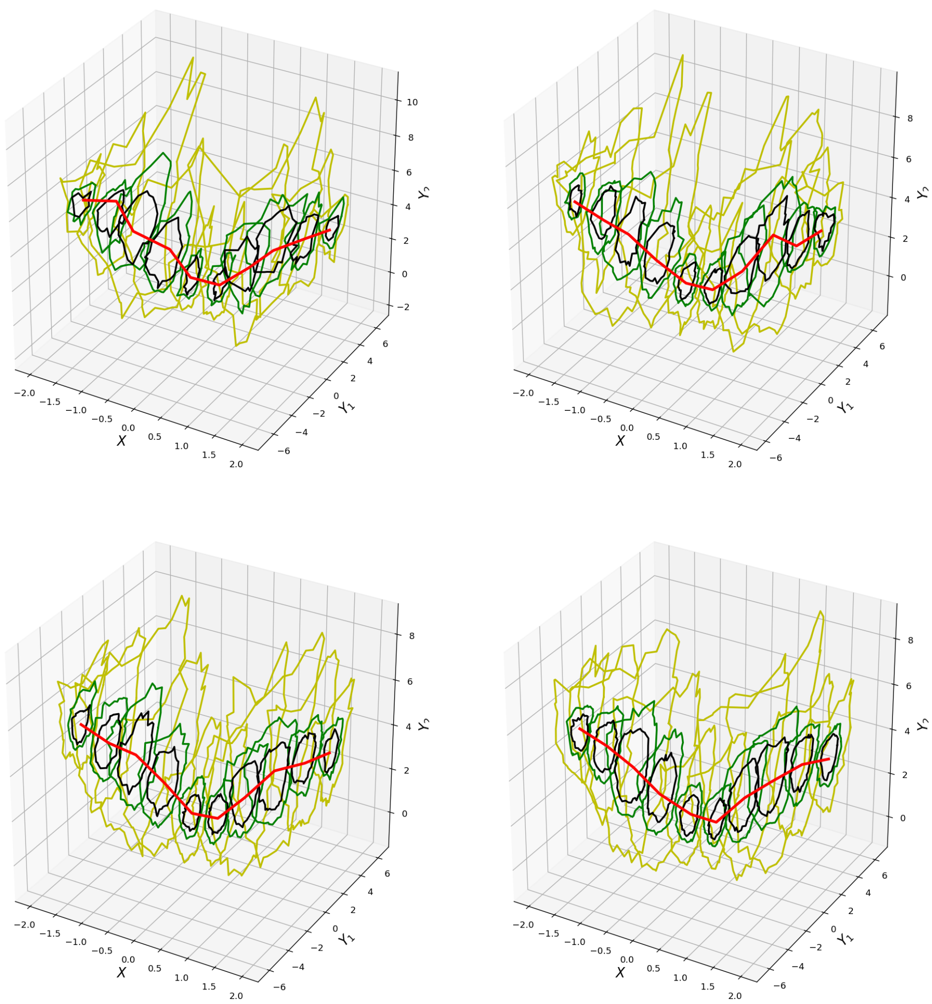

Figure 7.4.5: Performance, in Model (7.23), of Gaussian kernel weight functions-based estimation for various choices of the bandwidth. The sample size is $n = 3,601$, and the bandwidths are $h = 0.05$ (upper left panel), $h = 0.1$ (upper right panel), $h = 0.2$ (lower left panel), and $h = 0.3$ (lower left panel). The empirical contour levels are $\tau = 0.2$ (black) , 0.4 (green) and 0.8 (yellow); the estimated conditional center-outward medians are shown in red.

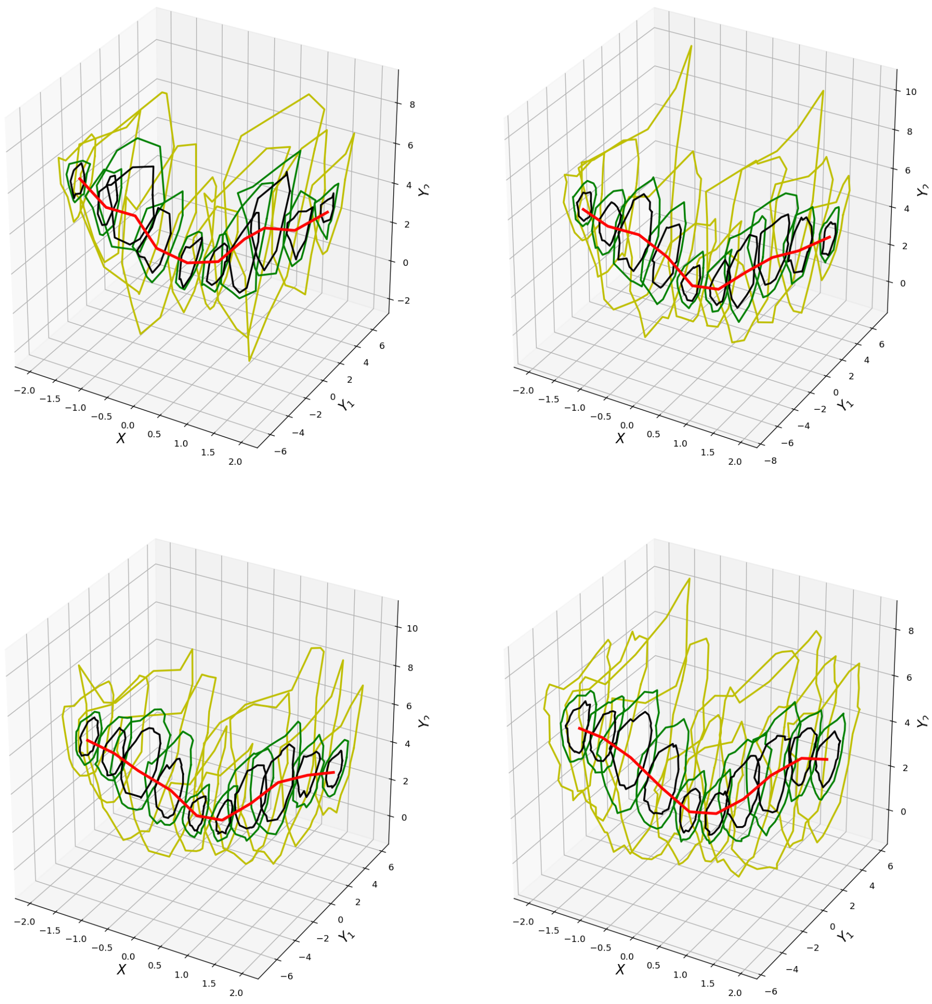

Figure 7.4.6: Performance, in Model (7.23), of the k-nearest neighbors weight functions-based estimation for various choices of k. The sample size is $n = 3601$; $k = 101$ (upper left panel), $k = 256$ (upper right panel), $k = 401$ (lower left panel), and $k = 625$ (lower left panel). The empirical contour levels are $\tau = 0.2$ (black), 0.4 (green), and 0.8 (yellow); the estimated conditional center-outward medians are shown in red.

biomass, and zooplankton biodiversity. We are focusing here on the influence of $X =$ *'depth'* (in meters) on the pair $\mathbf{Y} = ($*'temperature,' 'salinity'*$)'$ (in degrees and grams of salt per kilogram of water, respectively).

Figure 7.4.8 shows the corresponding 3D observations and the estimated conditional center-outward quantile contours obtained from the same method as in Section 7.4.1 (nearest neighbors weight function with $k = 6, 401$); Figure 7.4.9 shows the projections of the same contours on the ('*depth,*' '*salinity*') and ('*depth,*' '*temperature*') axes, respectively. Inspection of these figures reveals a nonlinear center-outward conditional median; heteroskedasticity also appears as the area of the conditional quantile regions clearly decreases as a function of depth, while a positive dependence between temperatures and salinity, which is present at the surface, gradually disappears as depth increases. The projection plots of Figure 7.4.9 also provide clearer views on marginal dependencies. For example, the decrease of temperature as a function of depth is monotone and almost linear, while the dependence on depth of salinity is more complex, high at shallow depths, lower at medium depths, and higher again at greater depths. However, these marginal analyses, to some degree, are hiding the heteroskedasticity effects (in particular, the dependence on depth of the relation between salinity and temperature) which are clearly visible in Figure 7.4.8. Since the dataset is quite large, we used a nearest neighbors weight function, see the comments about the empirical performance of different weights in Section 7.4.1.

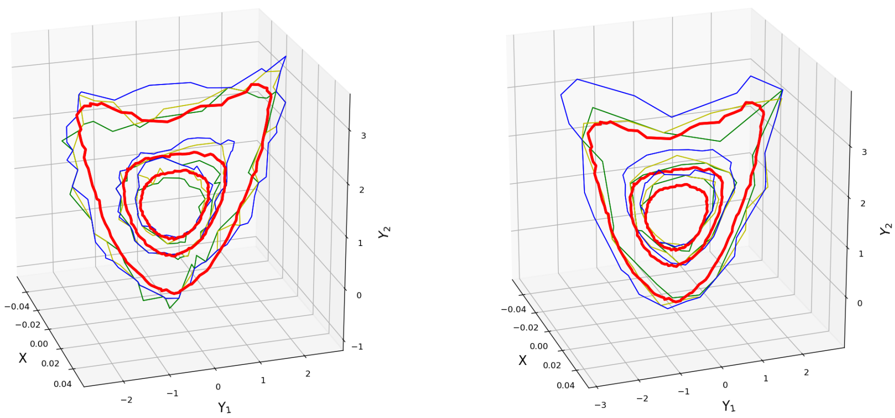

Figure 7.4.7: Comparison between the empirical conditional contours $\mathcal{C}_{\pm}^{(n)}(\tau\,|0)$ (levels $\tau = 0.2,\ 0.4,\ 0.8$; sample size $n = 3,601$) in Model (7.23) based on Gaussian kernel weight functions (bandwidths $h = 0.1$ (green), $h = 0.2$ (yellow), and $h = 0.3$ (blue)) (left panel) and those in Figure 7.4.4, based on $k-$nearest neighbors weight functions ($k = 226$ (green), $k = 485$ (yellow), and $k = 901$ (blue); sample size $n = 576,040$) (right panel). The center-outward quantiles of Figure 7.4.4 (to be considered as population values) are shown in red.

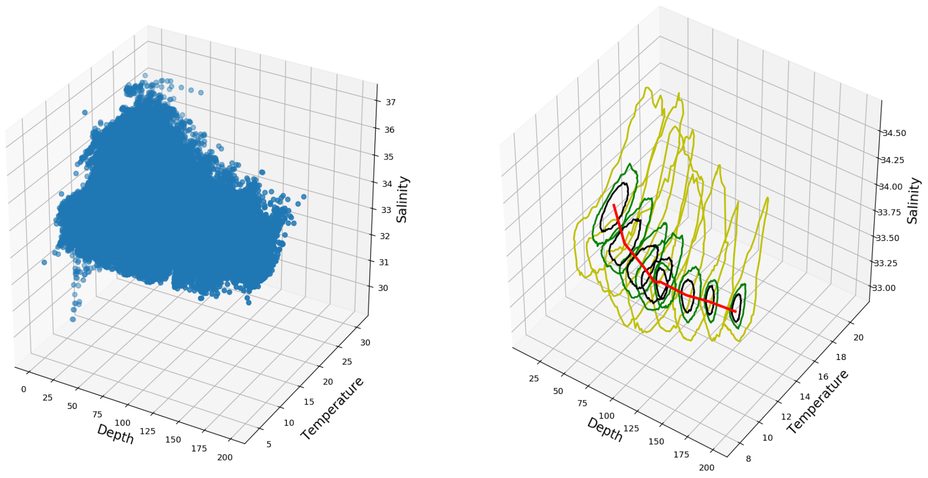

Figure 7.4.8: CalCOFI dataset. Left panel: the original dataset ('depth,' 'temperature,' 'salinity') for *'depth'* $\leq$ 200 (sample size for *'depth'* $\leq$ 200, dropping empty values, is $n = 505,829$). Right panel: the empirical conditional center-outward quantile contours of orders $\tau = 0.2$ (black), $\tau = 0.4$ (green), and $\tau = 0.8$ (yellow) and the empirical conditional center-outward median (red) for the multiple-output regression of $(Y_1, Y_2) =$ (*'temperature,' 'salinity'*) with respect to $X =$ *'depth.'* Estimation based on a k-nearest neighbors weight function with $k = 6,401$.

The Female ANSUR 2 dataset: stature, foot length and tibial height of female US Army personnel .

Our second real-data example involves a smaller sample size n. The Female Anthropometric Survey of US Army Personnel (Female ANSUR 2 or Female ANSUR II) featured in this section consists in ninety-three direct measures and 41 derived ones, as well as three-dimensional head, foot, and whole-body scans of $n = 1,986$ women of the United States' army. These measurements were collected between October 4, 2010 and April 5, 2012, in May 2014 and in May 2015; they are available online at `https://www.openlab.psu.edu/ansur2/`.

We want to analyze the influence of the covariate $X =$ *'stature'* (in centimeters) on the variable of interest $\mathbf{Y} =$ (*'foot length,' 'tibial height'*) (both in centimeters). Figure 7.4.10 provides a 3D view of the center-outward quantile contours/tubes (levels $\tau = 0.2, 0.4, 0.7$), along with the center-outward regression median (red); Figure 7.4.11 shows the projections on the axes of the same contours. Inspecting these two figures reveals the absence of heteroskedasticity, the spherical shape of conditional distributions, and a roughly linear regression. Since the size of the model is not too large, a Gaussian kernel is convenient. The bandwidth was chosen as $h = 15$, which, up to scale changes, corresponds to the choice $h = 0.2$ in Figure 7.4.5.

7.5 Some concluding remarks

7.5.1 Relation to the recent literature on numerical optimal transportation

The estimation of transport maps beyond the sample points currently is a hot topic, and a fastly developing strand of literature is proposing such estimators. However, the objective of most authors is to reach near-optimal convergence rates, for which they typically impose fairly strong assumptions. Some estimators (HÃ¼tter and Rigollet (2021), Manole et al. (2021)) are computationally quite heavy and sometimes numerically almost infeasible; others (Pooladian and Niles-Weed, 2021) use a regularized version (Cuturi (2013)) of the optimal transport problem to provide consistent near-optimal estimators, which needs stringent assumptions on the shape of the underlying distributions—such as being compactly supported, with densities bounded away from 0 and ∞ over their convex supports. This is redhibitory in our case, since the density of U_d (a choice which plays an essential role in the interpretation of transports as quantile functions) is unbounded at $\mathbf{0}$. Other solutions (Makkuva et al. (2020) and GonzÃ¡lez-Sanz et al. (2022)) are based on deep learning and neural network methods; they achieve excellent empirical performance, but the lack of theoretical results for the first one, the Lipschitz constraint on transport maps for the second, preclude their use in this quantile regression context.

We also could estimate the conditional quantiles through the optimal map from the uniform reference measure to $\mathrm{P}_n^{w(\mathbf{x})}$. From a numerical point of view, however, this would

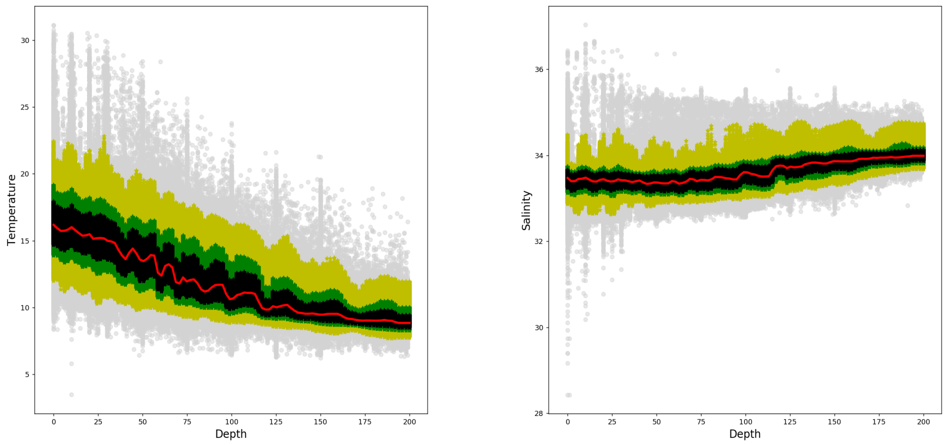

Figure 7.4.9: CalCOFI dataset. Left panel: projection of the empirical center-outward quantile regions shown in Figure 7.4.8 (orders $\tau = 0.2,\ 0.4,\ 0.8$) and median on the axes ('depth,' 'salinity'), for the multiple-output regression of $(Y_1, Y_2) =$ (*'temperature,' 'salinity'*) with respect to $X =$ *'depth.'* Right panel: projection of the same on the axes ('depth,' 'temperature'); see Figure 7.4.8 for the color code.

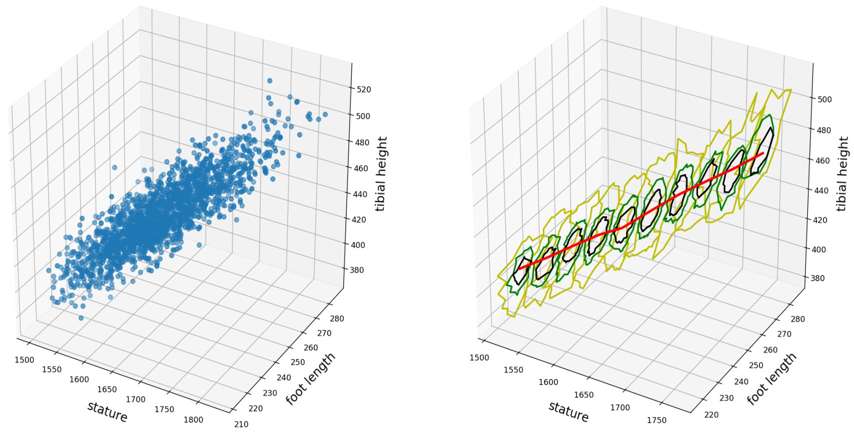

Figure 7.4.10: ANSUR 2 dataset (sample size $n = 1,986$). Left panel: the original dataset of $X =$ *'stature'* and $(Y_1, Y_2) =$ (*'foot length,' 'tibial height'*). Right panel: the empirical conditional center-outward quantile contours of orders $\tau = 0.2$ (black), $\tau = 0.4$ (green), $\tau = 0.7$ (yellow) and the empirical conditional center-outward median (red) for the multiple-output regression of (Y_1, Y_2) with respect to X; estimation based on a Gaussian kernel weight function with bandwidth $h = 15$.

lead to the computation of a semi-discrete optimal transportation plan, which has complexity $O(n^{d/2})$, hence is unfeasible even for moderate d. While the computational complexity of our procedure does not depend on the dimension, its statistical performance does (see Fournier and Guillin (2015)) and, in that sense, we do not escape the curse of dimensionality—up to the case where P is finitely supported, see del Barrio et al. (2021). Despite the fact that the literature on the computation of such maps is growing quite fastly (see LÃ©vy et al. (2020); GallouÃ«t and MÃ©rigot (2018); Meyron (2019); de Goes et al. (2012)), the exisiting methods are restricted to dimension two, sometimes three. A further issue is that the solution of the semi-discrete problem does not produce quantile *contours* but creates a Voronoi tessellation of $\mathbb{S}_d$, each piece of which is mapped to a single sample point.

7.5.2 Conclusions and perspectives for further developments

Building on the concepts of center-outward quantiles recently developed in Hallin et al. (2021), we are proposing here a fully nonparametric solution to the problem of nonparametric multiple-output quantile regression. Contrary to earlier attempts, our solution is enjoying the quintessential property that the (conditional) probability content of its quantile regions is under control irrespective of the underlying distribution. This is only a first step into

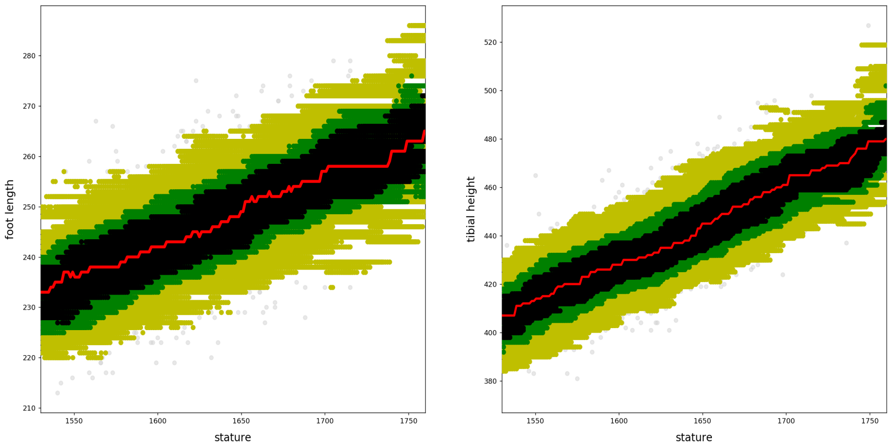

Figure 7.4.11: ANSUR 2 dataset (sample size $n = 1,986$). Left panel: projection of the empirical center-outward quantile regions (orders $\tau = 0.2,\ 0.4,\ 0.8$) and median on the axes ('stature,' 'foot length'), for the multiple-output regression of $(Y_1, Y_2) = ($*'foot length,' 'tibial height'*$)$ with respect to $X =$ *'stature.'* Right panel: projection of the same on the axes ('stature,' 'tibial height'); see Figure 7.4.10 for the color code.

the multifarious applications of multiple-output quantile regression, though. Due to the minimality of the assumptions it requires, a completely agnostic nonparametric approach indeed is attractive, but also comes at a cost: linear or polynomial quantile regression remain justified whenever a priori knowledge of the analytical form of the regression is available and should be taken advantage of. A center-outward version of the results of Carlier et al. (2016), thus, is highly desirable. Single-output quantile regression has been considered in a variety of contexts: survival analysis, longitudinal data, instrumental variable regression, directional, functional, and high-dimensional data, . . . Quantile regression versions of time-series models such as the quantile autoregressive model also have been investigated (Koenker and Xiao, 2006). All these applications call for multiple-output extensions with important real-life consequences; they should and can be based on the concept of center-outward quantile, regions, and contours and are the subject of our ongoing research.

Appendix

Contents

7..3 Proofs for Section 7.3

The convergence described of Theorem 7.3.2 is based on the topology of set-valued maps, in particular the Graphical convergence (or Painlevé-Kuratowski convergence of the graphs). Recall from Rockafellar and Wets (1998) that a sequence of set-valued maps $\{T_n\}_n$ converges *graphically* to another set-valued map T if

- the outer limit $\limsup_n T_n$—which is the set of $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^d$ for which there exists a sequence $\{(\mathbf{x}_n, \mathbf{y}_n)\}$ with $(\mathbf{x}_n, \mathbf{y}_n) \in T_n$ containing a subsequence which converges to $(\mathbf{x}, \mathbf{y})$—exists and coincides with T and
- the inner limit $\liminf_n T_n$—which is the set of $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^d$ for which there exists a sequence $\{(\mathbf{x}_n, \mathbf{y}_n)\}$, with $(\mathbf{x}_n, \mathbf{y}_n) \in T_n$, which converges to $(\mathbf{x}, \mathbf{y})$—exists and coincides with T.

We start with some necessary properties of the population and empirical conditional quantiles and some auxiliary results. The following lemma states that the estimated conditional probability converges weakly in probability to its population counterpart. Recall from Theorem 1.12.4 in van der Vaart and Wellner (1996) that weak convergence can be measured in terms of the bounded Lipschitz norm, which is defined for $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ as

$$d_{BL}(\mu, \nu) := \sup_{f \in \mathcal{F}_{BL}(\mathbb{R}^d)} \left| \mathbb{E}_{\mathbf{Z} \sim \mu}(f(\mathbf{Z})) - \mathbb{E}_{\mathbf{W} \sim \nu}(f(\mathbf{W})) \right|,$$

where the class $\mathcal{F}_{BL}$ is the class of functions $f : \mathbb{R}^d \to \mathbb{R}$ such that $|f(\mathbf{z}_1) - f(\mathbf{z}_2)| \leq |\mathbf{z}_1 - \mathbf{z}_2|$ and $|f(\mathbf{z}_1)| \leq 1$, for all $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^d$.

Lemma 7..1. *For any $\epsilon > 0$, under the assumptions of Theorem 7.3.2, we have*

$$\mathbb{P}\left(d_{BL}(P_n^{w(\mathbf{X})}, \mathrm{P}_{\mathbf{Y}|\mathbf{X}}) > \epsilon \right) \longrightarrow 0.$$

See Section 7..4 for the proof.

Let $\pi^{(n)}(\mathbf{x})$ be a solution of (7.12) and

$$\pi^*(\mathbf{x}) := (\mathbf{Id} \times \mathbf{Q}_{\pm}(\,\cdot\,|\mathbf{x}))\#\mathrm{U}_d.$$

Note that, for each value of $\mathbf{x}$,

(i) there exists a sequence of differentiable convex functions $\psi^{(n)}\left(\,\cdot\,\middle|\mathbf{x}\right) : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, $n \in \mathbb{N}$, such that

$$\mathbf{Q}^{(n)}_{w,\pm}\left(\mathbf{u}_j \middle|\mathbf{x}\right) = \nabla\psi^{(n)}\left(\mathbf{u}_j \middle|\mathbf{x}\right) \text{ for } j = 1, \ldots, k, \text{ and } n \in \mathbb{N};$$

(ii) there exists a convex function $\psi\left(\,\cdot\,\middle|\mathbf{X} = \mathbf{x}\right) : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ such that

$$\mathbf{Q}_{\pm}\left(\mathbf{u} \middle|\mathbf{X} = \mathbf{x}\right) = \nabla\psi\left(\mathbf{u} \middle|\mathbf{X} = \mathbf{x}\right) \text{ for } \mathrm{U}_d\text{-a.e. } \mathbf{u} \in \mathbb{S}_d.$$

Using obvious notation, it holds that $\pi^*(\mathbf{x})$ and $\pi^{(n)}(\mathbf{x})$, for $n \in \mathbb{N}$, have cyclically monotone supports. Moreover, as a consequence of Corollary 7.3.1, we have

$$\mathrm{supp}(\pi^{(n)}(\mathbf{x})) \subset \partial\tilde{\psi}^{(n)}\left(\,\cdot\,\middle|\mathbf{x}\right) \text{ and } \mathrm{supp}(\pi^*(\mathbf{x})) \subset \partial\psi\left(\,\cdot\,\middle|\mathbf{X} = \mathbf{x}\right), \tag{7.24}$$

possibly for some other sequence $\tilde{\psi}^{(n)}\left(\,\cdot\,\middle|\mathbf{x}\right) : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ of convex functions such that

$$\mathbf{Q}^{(n)}_{w,\pm}\left(\mathbf{u}_j \middle|\mathbf{x}\right) \in \partial\tilde{\psi}^{(n)}\left(\mathbf{u}_j \middle|\mathbf{x}\right) \text{ for } j = 1, \ldots, k, \text{ and } n \in \mathbb{N}. \tag{7.25}$$

The following result then follows from Lemma 9 and Corollary 14 in McCann (1995).

Lemma 7..2. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ be such that $\mu \ll \ell_d$ is supported on a convex set. Let $\{\mu_n\}_{n\in\mathbb{N}}$ and $\{\nu_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(\mathbb{R}^d)$ converge weakly as $n \to \infty$ to μ and ν, respectively. Suppose, moreover, that there exists a sequence of probability measures $\{\pi_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with marginals μ_n and ν_n such that, for some sequence of convex functions $\{\phi_n\}_{n\in\mathbb{N}}$, it holds that, for all $n \in \mathbb{N}$, $\mathrm{supp}(\pi_n) \subset \partial\phi_n$. Then,*

(i) $\{\pi_n\}_n$ converges weakly as $n \to \infty$ to $\pi^* = (\mathbf{Id} \times \nabla\phi)\#\mu$, where $\nabla\phi$ is the gradient of a convex function ϕ pushing μ forward to ν, and

(ii) there exists a sequence $\{a_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ such that, μ-a.e., $\phi_n + a_n \to \phi$ and $\partial\phi_n \to \partial\phi$ graphically as $n \to \infty$.

See Section 7..4 for the proof.

Proof of Theorem 7.3.2. Suppose that there exist $\mathbf{u}_0 \in \mathbb{S}_d$ and $\epsilon_0 > 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}\left(\mathbf{Q}^{(n)}_{w,\pm}(\mathbf{u}_0|\mathbf{x}) \notin \mathbf{Q}_{\pm}(\mathbf{u}_0|\mathbf{X}) + \epsilon_0\mathbb{S}_d\right) = \delta > 0.$$

We can find a subsequence n_k such that

$$\lim_{k\to\infty} \mathbb{P}\left(\mathbf{Q}^{(n_k)}_{w,\pm}(\mathbf{u}_0|w(\mathbf{X};\mathbf{X}^{[n_k]})) \notin \mathbf{Q}_{\pm}(\mathbf{u}_0|\mathbf{X}) + \epsilon_0\mathbb{S}_d\right) = \delta > 0. \tag{7.26}$$

Note that the space of probability distributions on $\mathbb{R}^d$, endowed with the bounded Lipschitz metric, is separable and complete, see Theorem 1.12.4 in van der Vaart and Wellner (1996).

Therefore, the convergence described in Lemma 7..1 implies that, for the subsequence n_k, there exists a further sub sequence n_{k_i} such that the event

$$\Omega_0 = \left(\sup_{f \in \mathcal{F}_{BL}(\mathbb{R}^d)} \left| \sum_{j=1}^{n_{k_i}} w_j(\mathbf{X}; \mathbf{X}^{(n_{k_i})}) f(\mathbf{Y}_j) - \mathbb{E}(f(\mathbf{Y})|\mathbf{X}) \right| \longrightarrow 0 \right) \tag{7.27}$$

has probability one. Let $\mathbf{x} = \mathbf{X}(\omega)$ with $\omega \in \Omega_0$. Then, by Lemma 7..2, we have, as $i \to \infty$,

$$\partial \tilde{\psi}^{(n_{k_i})} \left(\,\cdot\, |\mathbf{x} \right) \longrightarrow \mathbf{Q}_{\pm}(\,\cdot\, |\mathbf{X} = \mathbf{x}) \ \text{ graphically}.$$

This implies, by Theorem 8.3 in Bagh and Wets (1996), that there exists $i_0 \in \mathbb{N}$ such that, for all $i > i_0$,

$$\partial \tilde{\psi}^{(n_{k_i})} \left(\mathbf{u}_0 |\mathbf{x} \right) \subset \mathbf{Q}_{\pm}(\mathbf{u}_0|\mathbf{X} = \mathbf{x}) + \epsilon_0 \mathbb{S}_d. \tag{7.28}$$

From (7.25) and the fact that (7.28) holds with probability one, we deduce that

$$\mathbb{P}\left(\bigcap_{i_0 \in \mathbb{N}} \bigcup_{i \geq i_0} \mathbf{Q}_{\pm}^{(n_{k_i})}(\mathbf{u}_0|\mathbf{X}) \not\in \mathbf{Q}_{w,\pm}(\mathbf{u}_0|\mathbf{X}) + \epsilon_0 \mathbb{S}_d \right) = 0, \tag{7.29}$$

which implies that

$$\limsup_{i \to \infty} \mathbb{P}\left(\mathbf{Q}_{\pm}^{(n_{k_i})}(\mathbf{u}_0|\mathbf{X}) \not\in \mathbf{Q}_{w,\pm}(\mathbf{u}_0|\mathbf{X}) + \epsilon_0 \mathbb{S}_d \right) \leq 0.$$

This contradicts (7.26).

The rest of the proof follows from compactness arguments and a refined use of Theorem 8.3. in Bagh and Wets (1996). Suppose that, for some $q_0 \in (0, 1)$ and $\epsilon_0 > 0$, and along a subsequence n_k, we have

$$\lim_{k \to \infty} \mathbb{P}\left(\mathbb{C}_{\pm}^{(n_k)}(q_0 \,|\mathbf{X}) \not\subset \mathbb{C}_{\pm}(q_0 \,|\mathbf{X}) + \epsilon_0 \mathbb{S}_d \right) = \delta > 0. \tag{7.30}$$

Theorem 8.3. in Bagh and Wets (1996) and Lemma 7..2 jointly imply the existence of a further subsequence n_{k_i}, such that, for every $\bar{\mathbf{u}} \in \mathbb{S}_d$, there exists $I_{\bar{\mathbf{u}}} \in \mathbb{N}$ and $\lambda_{\bar{\mathbf{u}}} > 0$ satisfying

$$\mathbf{Q}_{\pm}^{(n_{k_i})}(\mathbf{u}\,|\mathbf{x}) \in \mathbf{Q}_{\pm}(\bar{\mathbf{u}}\,|X = \mathbf{x}) + \epsilon_0 \mathbb{S}_d \text{ for all } \mathbf{u} \in \bar{\mathbf{u}} + \lambda_{\bar{\mathbf{u}}} \mathbb{S}_d \text{ and all } i > I_{\bar{\mathbf{u}}}.$$

Note that $q\,\mathbb{S}_d \subset \bigcup_{\bar{\mathbf{u}}:\ |\bar{\mathbf{u}}| \leq q} \bar{\mathbf{u}} + \lambda_{\bar{\mathbf{u}}} \mathbb{S}_d$. Since $q\,\overline{\mathbb{S}}_d$ is compact, there exists a finite covering $q\,\mathbb{S}_d \subset \bigcup_{k=1}^{K_\epsilon} \bar{\mathbf{u}}_k + \delta_{\bar{\mathbf{u}}_k} \mathbb{S}_d$. Set $I_0 = \max(I_{\bar{\mathbf{u}}_1}, \ldots, I_{\bar{\mathbf{u}}_{K_\epsilon}})$: then for all $i > I_{\bar{\mathbf{u}}}$, we have

$$\mathbb{C}_{\pm}^{(n_{k_i})}(q_0 \,|\mathbf{x}) \subset \mathbb{C}_{\pm}(q_0 \,|\mathbf{x}) + \epsilon_0 \mathbb{S}_d,$$

which holds with probability one and, using the same argument as for (7.28), contradicts (7.30). The same reasoning holds also for the contours. $\square$

Proof of Theorem 7.3.3. Under Assumption (R), it follows from del Barrio et al. (2020) that, for $\omega \in \Omega_0 \subseteq \Omega$ where $\mathbb{P}(\Omega_0) = 1$, the center-outward quantile function $\mathbf{Q}_\pm(\mathbf{u}\,|\mathbf{X} = \mathbf{x} := \mathbf{X}(\omega))$ is a singleton for U_d-almost all values of $\mathbf{u} \in \mathbb{S}_d$. Therefore, we adopt here the slight abuse of notation commented before Theorem 7.3.3 and write $\mathbf{Q}_\pm(\mathbf{u}\,|\mathbf{X} = \mathbf{x}) = \{\mathbf{Q}_\pm(\mathbf{u}\,|\mathbf{X} \text{ for } \mathbf{x})\}$. Set $\mathbf{u} \in \mathbb{S}_d \setminus \{\mathbf{0}\}$ and note that, since $\mathbf{Q}_\pm(\mathbf{u}\,|\mathbf{X})$ and $\mathbf{Q}^{(n)}_{w,\pm}(\mathbf{u}\,|\mathbf{x})$ are a.s. singletons, we have, as n and N tend to infinity,

$$\mathbb{P}\left(\left|\mathbf{Q}^{(n)}_{w,\pm}(\mathbf{u}\,|\mathbf{x}) - \mathbf{Q}_\pm(\mathbf{u}\,|\mathbf{X})\right| > \epsilon\right) \to 0.$$

Let K be a compact subset of $\mathbb{S}_d \setminus \{\mathbf{0}\}$. In order to establish uniform convergence in K, suppose that the contrary holds. Since the space of continuous functions from K to $\mathbb{R}^d$ (endowed with the topology of uniform convergence) is complete and separable, there exists a subsequence n_k such that, for some $\delta > 0$, the probability of

$$\Omega' = \left(\sup_{\mathbf{u}\in K}\left|\mathbf{Q}^{(n_k)}_{w,\pm}(\mathbf{u}\,|\mathbf{X}) - \mathbf{Q}_\pm(\mathbf{u}\,|\mathbf{X})\right| \to \delta\right), \tag{7.31}$$

is one. Set $\omega \in \Omega'$ and consider $\mathbf{x} = \mathbf{X}(\omega)$. There exists a sequence $\{\mathbf{u}_{n_k}\} \subset K$ such that

$$\left|\mathbf{Q}^{(n_k)}_{w,\pm}(\mathbf{u}_{n_k}\,|\mathbf{x}) - \mathbf{Q}_\pm(\mathbf{u}_{n_k}\,|\mathbf{X} = \mathbf{x})\right| \to \delta,$$

for a possibly different $\delta > 0$. Since the sequence $\{\mathbf{u}_{n_k}\}_k$ is in K, it admits at least one point of accumulation $\bar{\mathbf{u}} \in K$. Hence,

$$\begin{aligned}\liminf_{k\to\infty}\Big(&\left|\mathbf{Q}^{(n_k)}_{w,\pm}(\mathbf{u}_{n_k}\,|\mathbf{x}) - \mathbf{Q}_\pm(\bar{\mathbf{u}}\,|\mathbf{X} = \mathbf{x})\right| \\ &+ \left|\mathbf{Q}_\pm(\bar{\mathbf{u}}\,|\mathbf{X} = \mathbf{x}) - \mathbf{Q}_\pm(\mathbf{u}_{n_k}\,|\mathbf{X} = \mathbf{x})\right|\Big) \geq \delta,\end{aligned}$$

where the second term tends to 0 by the continuity of $\mathbf{u} \mapsto \mathbf{Q}_\pm(\mathbf{u}\,|\mathbf{X} = \mathbf{x})$. This implies that

$$\liminf_{k\to\infty}\left|\mathbf{Q}^{(n_k)}_{w,\pm}(\mathbf{u}_{n_k}\,|\mathbf{x}) - \mathbf{Q}_\pm(\bar{\mathbf{u}}\,|\mathbf{X} = \mathbf{x})\right| \geq \delta. \tag{7.32}$$

But Lemma 7..2 entails the Graphical convergence to $\mathbf{Q}_\pm(\,\cdot\,|\mathbf{X} = \mathbf{x})$ of $\mathbf{Q}^{(n_k)}_{w,\pm}(\,\cdot\,|\mathbf{x})$, which contradicts (7.32). The desired uniformity over K follows.

Finally, the convergence of the contours is a consequence of the previous result on the regions. □

Proof of Corollary 7.3.4. To prove (7.15), let us show that, for any subsequence n_k, there exists a further subsequence converging a.s. To avoid repetitions, assume that $N = N(n)$ is such that $N \to \infty$ as $n \to \infty$.

Leaving aside the singular points, let us consider the empirical and population *quantile rings*

$$\mathbb{C}^{(n)}_\pm(\epsilon, \tau\,|\mathbf{x}) := \mathbb{C}^{(n)}_\pm(\tau\,|\mathbf{x}) \setminus \mathbb{C}^{(n)}_\pm(\epsilon\,|\mathbf{x})$$

and

$$\mathbb{C}_\pm(\epsilon, \tau\,|\mathbf{x}) := \mathbb{C}_\pm(\tau\,|\mathbf{x}) \setminus \mathbb{C}_\pm(\epsilon\,|\mathbf{x}),$$

respectively. Theorem 3.3 yields, for all $0 < \epsilon < \tau$, as $n \to \infty$,

$$\mathbb{P}\left(\mathbf{Y} \in \mathbb{C}_{\pm}^{(n)}(\epsilon,\, \tau \,|\mathbf{X}) \text{ and } \mathbf{Y} \notin \mathbb{C}_{\pm}(\epsilon,\, \tau \,|\mathbf{X}) \Big| \mathbf{X}\right) \xrightarrow{\mathbb{P}} 0. \tag{7.33}$$

Let $\{\epsilon_j\}_{j\in\mathbb{N}}$ be a monotone sequence tending to 0. For $j = 1$ there exists a further subsequence n_k^1, say, and a subset Ω_1 of Ω such that $\mathbb{P}(\Omega_1) = 1$ such that, for every $\mathbf{x} = \mathbf{X}(\omega)$ with $\omega \in \Omega_1$,

$$\mathbb{P}\left(\mathbf{Y} \in \mathbb{C}_{\pm}^{(n_k^1)}(\epsilon_1,\, \tau \,|\mathbf{x}) \text{ and } \mathbf{Y} \notin \mathbb{C}_{\pm}(\epsilon_1,\, \tau \,|\mathbf{x}) \Big| \mathbf{X} = \mathbf{x}\right) \longrightarrow 0. \tag{7.34}$$

By definition, $\mathbb{P}\left(\mathbb{C}_{\pm}(\epsilon_1,\, \tau \,|\mathbf{x}) \Big| \mathbf{X} = \mathbf{x}\right) = \tau - \epsilon_1$; therefore, in view of (7.34),

$$\mathbb{P}\left(\mathbb{C}_{\pm}^{(n_k^1)}(\epsilon_1,\, \tau \,|\mathbf{x}) \Big| \mathbf{X} = \mathbf{x}\right) \longrightarrow \tau - \epsilon_1.$$

Repeating the argumentat for $j = 2$, there exists a set $\Omega_2 \subset \Omega$, with $\mathbb{P}(\Omega_2) = 1$, and a subsequence (call it n_k^2) of n_k^1 such that, for every $\omega \in \Omega_2$ and $\mathbf{x} = \mathbf{X}(\omega)$,

$$\mathbb{P}\left(\mathbb{C}_{\pm}^{(n_k^2)}(\epsilon_2,\, \tau \,|\mathbf{x}) \Big| \mathbf{X} = \mathbf{x}\right) \longrightarrow \tau - \epsilon_2.$$

This argument can be repeated for each $j \in \mathbb{N}$ and the set $\Omega_0 = \bigcap_{j\in\mathbb{N}} \Omega_j$ has probability one. Set $\omega \in \Omega_0 \subset \Omega$ and $\mathbf{x} = \mathbf{X}(\omega)$: then there exists k_1 such that

$$\left|\mathbb{P}\left(\mathbb{C}_{\pm}^{(n_{k_1})}(\epsilon_1,\, \tau \,|\mathbf{x}) \big| \mathbf{X} = \mathbf{x}\right) - \tau\right| < 2\,\epsilon_1.$$

Analogously, for each m, there exist k_j such that

$$\left|\mathbb{P}\left(\mathbb{C}_{\pm}^{(n_{k_j})}(\epsilon_j,\, \tau \,|\mathbf{x}) \big| \mathbf{X} = \mathbf{x}\right) - \tau\right| < 2\,\epsilon_j.$$

Therefore, we obtain the limit

$$\mathbb{P}\left(\mathbb{C}_{\pm}^{(n_{k_j})}(\epsilon_j,\, \tau \,|\mathbf{x}) \big| \mathbf{X} = \mathbf{x}\right) \longrightarrow \tau$$

and, noticing that

$$\mathbb{P}\left(\mathbb{C}_{\pm}^{(n_{k_j})}(\epsilon_j,\, \tau \,|\mathbf{x}) \big| \mathbf{X} = \mathbf{x}\right) \leq \mathbb{P}\left(\mathbb{C}_{\pm}^{(n_{k_j})}(\tau \,|\mathbf{x}) \big| \mathbf{X} = \mathbf{x}\right),$$

also the asymptotic upper bound

$$\liminf \mathbb{P}\left(\mathbb{C}_{\pm}^{(n_{k_j})}(\tau \,|\mathbf{x}) \big| \mathbf{X} = \mathbf{x}\right) \geq \tau.$$

Now, Theorem 7.3.2 implies, for every subsequence (for which we keep the notation n) of n and $j \in \mathbb{N}$, there existence of some $n_j \in \mathbb{N}$ such that

$$\mathbb{P}\left(\mathbb{C}_{\pm}^{(n)}(\tau\,|\mathbf{X}) \not\subset \mathbb{C}_{\pm}(\tau\,|\mathbf{X}) + \frac{1}{2^j}\mathbb{S}_d \right) \leq \frac{1}{2^j}$$

for all $n \geq n_j$. The sum

$$\sum_{j=1}^{\infty} \mathbb{P}\left(\mathbb{C}_{\pm}^{(n_j)}(\tau\,|\mathbf{X}) \not\subset \mathbb{C}_{\pm}(\tau\,|\mathbf{X}) + \frac{1}{2^j}\mathbb{S}_d \right)$$

thus is finite, and the Borel-Cantelli lemma yields

$$\mathbb{P}\left(\bigcap_{J\in\mathbb{N}} \bigcup_{j\geq J} \mathbb{C}_{\pm}^{(n_j)}(\tau\,|\mathbf{X}) \not\subset \mathbb{C}_{\pm}(\tau\,|\mathbf{X}) + \frac{1}{2^j}\mathbb{S}_d \right) = 0,$$

or, equivalently, $\mathbb{P}(\Omega^*) = 1$, where

$$\Omega^* := \left(\bigcup_{J\in\mathbb{N}} \bigcap_{j\geq J} \mathbb{C}_{\pm}^{(n_j)}(\tau\,|\mathbf{X}) \subset \mathbb{C}_{\pm}(\tau\,|\mathbf{X}) + \frac{1}{2^j}\mathbb{S}_d \right).$$

Setting $\omega \in \Omega^* \subset \Omega$ and $\mathbf{x} = \mathbf{X}(\omega)$, there exists $J \in \mathbb{N}$ such that

$$\mathbb{P}\left(\mathbb{C}_{\pm}^{(n_j)}(\tau\,|\mathbf{x}) \big| \mathbf{X} = \mathbf{x} \right) \leq \mathbb{P}\left(\mathbb{C}_{\pm}(\tau\,|\mathbf{x}) + \frac{1}{2^j}\mathbb{S}_d \big| \mathbf{X} = \mathbf{x} \right),$$

for all $j \geq J$. Since

$$\mathbb{C}_{\pm}(\tau\,|\mathbf{x}) + \frac{1}{2^{j+1}}\mathbb{S}_d \subset \mathbb{C}_{\pm}(\tau\,|\mathbf{x}) + \frac{1}{2^j}\mathbb{S}_d$$

and

$$\bigcap_{j\in\mathbb{N}} \mathbb{C}_{\pm}(\tau\,|\mathbf{x}) + \frac{1}{2^j}\mathbb{S}_d = \mathbb{C}_{\pm}(\tau\,|\mathbf{x}),$$

we obtain

$$\limsup \mathbb{P}\left(\mathbb{C}_{\pm}^{(n_j)}(\tau\,|\mathbf{x}) \big| \mathbf{X} = \mathbf{x} \right) \leq \tau, \tag{7.35}$$

which concludes the proof. □

Proof of Lemma 7.3.6 To prove this lemma, we show that the following five conditions of Stone's theorem (Theorem 1 in Stone (1977)) are satisfied (convergence for $n \to \infty$):

(a) *there exists $C \geq 1$ such that, for any non-negative measurable function f,*

$$\mathbb{E}\left(\frac{1}{k} \sum_{j=1}^{n} \mathbb{1}_{\mathbf{X}_j \in K_n^k(\mathbf{X})} f(\mathbf{X}_j) \right) \leq C\mathbb{E}\left(f(\mathbf{X})\right);$$

(b) $\mathbb{P}\left(\frac{1}{k}\sum_{j=1}^{n}\mathbb{1}_{\mathbf{X}_j\in K_n^k(\mathbf{X})}\leq 1\right)=1$ for all $n\in\mathbb{N}$;

(c) $\frac{1}{k}\sum_{j=1}^{n}\mathbb{1}_{\mathbf{X}_j\in K_n^k(\mathbf{X})}\mathbb{1}_{|\mathbf{X}_j-X|>a}\longrightarrow 0,$ in probability, for all $a>0$;

(d) $\frac{1}{k}\sum_{j=1}^{n}\mathbb{1}_{\mathbf{X}_j\in K_n^k(\mathbf{X})}\longrightarrow 1$ in probability;

(e) $\max_{i=j,\cdots,n}\frac{1}{k}\mathbb{1}_{\mathbf{X}_j\in K_n^k(\mathbf{X})}\longrightarrow 0$ in probability.

Proof of (a). Let $\mathbf{X}$ be a random variable independent of $\mathbf{X}^{(n)}$, with the same distribution as $\mathbf{X}_i$. Set $\mathbf{u}_0=\mathbf{F}_{\pm}^{(n)}(\mathbf{X})$ and $\mathbf{u}_j=\mathbf{F}_{\pm}^{(n)}(\mathbf{X}_j)$, $j=1,\dots,n$. Defining

$$K_n^{k,(i)}(\mathbf{X}):=\{\mathbf{X}_j:\ i\neq j,\ \mathbb{F}_{\pm}^{n}(\mathbf{X}_j)\in N_k(\mathbf{F}_{\pm}^{(n)}(\mathbf{X}_i))\},$$

note that $\mathbb{1}_{\mathbf{X}_i\in K_n^k(\mathbf{X})}f(\mathbf{X}_i)$ and $\mathbb{1}_{X\in K_n^{k,(i)}(\mathbf{X}_i)}f(\mathbf{X})$ have the same distribution. It follows that

$$\mathbb{E}\left(\frac{1}{k}\sum_{j=1}^{n}\mathbb{1}_{\mathbf{X}_j\in K_n^k(\mathbf{X})}f(\mathbf{X}_j)\right)=\frac{1}{k}\sum_{j=1}^{n}\mathbb{E}\left(\mathbb{1}_{\mathbf{X}_j\in K_n^k(\mathbf{X})}f(\mathbf{X}_j)\right)=\frac{1}{k}\sum_{j=1}^{n}\mathbb{E}\left(\mathbb{1}_{X\in K_n^{k,(j)}(\mathbf{X}_j)}f(\mathbf{X})\right),$$

which implies that

$$\mathbb{E}\left(\frac{1}{k}\sum_{j=1}^{n}\mathbb{1}_{\mathbf{X}_j\in K_n^k(\mathbf{X})}f(\mathbf{X}_j)\right)\leq\mathbb{E}\left(f(\mathbf{X})\frac{1}{k}\sum_{j=1}^{n}\mathbb{1}_{X\in K_n^{k,(j)}(\mathbf{X}_j)}\right). \tag{7.36}$$

Since $\sum_{j=1}^{n}\mathbb{1}_{X\in K_n^k(\mathbf{X}_j)}=\sum_{j=1}^{n}\mathbb{1}_{\mathbf{u}_0\in N_k(\mathfrak{G}_j)}$, we can apply Corollary 6.1. in GyÃ¶rfi et al. (2002) and conclude that there exists $\lambda_d\in\mathbb{R}$ such that $\sum_{j=1}^{n}\mathbb{1}_{X\in K_n^{k,(j)}(\mathbf{X}_j)}\leq k\lambda_d$. This and (7.36) complete the proof. □

Proof of (b), (d), and (e). Conditions (b) and (d) are direct consequences of the properties of the weight function, see Corollary 1 in Stone (1977). As for (e), it follows from the fact that $k\to\infty$. □

Proof of (c). The following lemma (see Section 7..4 for a proof) is a corollary of Lemma 6.1 in GyÃ¶rfi et al. (2002).

Lemma 7..3. *Let $k/n\to 0$ as $n\to\infty$. Then, as $n\to\infty$*

$$\sup_{\mathfrak{G}_i\in N_k(\mathfrak{G}_0)}|\mathfrak{G}_i-\mathfrak{G}_0|\longrightarrow 0,\quad a.s.$$

Since $\mathbf{Q}_\pm$ is a singleton with probability one, let us assume, without loss of generality, that $\mathbf{Q}_\pm(\mathfrak{G}_0)$ and $\mathbf{Q}_\pm^{(n)}(\mathfrak{G}_k)$ are singletons. Actually, the set

$$\bigcap_{k=1}^{\infty}\bigcap_{n=0}^{\infty}\left\{\mathbf{Q}_\pm^{(n)}(\mathfrak{G}_k) \text{ is a singleton }\right\}$$

also has probability one. Within that set,

(i) $\mathbf{Q}_\pm^{(n)} \longrightarrow \mathbf{Q}_\pm$ graphically as $n \to \infty$,

(ii) for every $\mathbf{X}_i \in K_n^k(\mathbf{X})$, there exists some $\mathfrak{G}_i \in N_k(\mathfrak{G}_0)$ such that $\mathbf{Q}_\pm^{(n)}(\mathfrak{G}_i) = \{\mathbf{X}_i\}$, and

(iii) Lemma 7..3 yields $\mathfrak{G}_i \to \mathfrak{G}_0$ as $n \to \infty$.

This, in view of Proposition 5.33 in Rockafellar and Wets (1998), implies that

$$\sup_{\mathbf{X}_i\in K_n^k(\mathbf{X})} |\mathbf{X}_i - \mathbf{X}\,| \longrightarrow 0, \quad a.s. \tag{7.37}$$

We are ready now to prove (c). Set $a > 0$. Since there are k elements in $K_n^k(\mathbf{X})$, we have

$$\mathbb{E}\left(\frac{1}{k}\sum_{j=1}^{n}\mathbb{1}_{\mathbf{X}_j\in K_n^k(\mathbf{X})}\mathbb{1}_{|\mathbf{X}_j-X|>a}\right) = \mathbb{E}\left(\frac{1}{k}\sum_{\mathbf{X}_j\in K_n^k(\mathbf{X})}\mathbb{1}_{|\mathbf{X}_j-\mathbf{X}\,|>a}\right)$$
$$= \mathbb{P}\left(\sup_{\mathbf{X}_i\in K_n^k(\mathbf{X})}|\mathbf{X}_j - \mathbf{X}\,| > a\right)$$

which, owing to (7.37), tends to 0. The desired result follows as a direct consequence of Markov's inequality. □

7..4 Proofs of Lemmas 7..1, 7..2, and 7..3

Proof of Lemma 7..1. Let $\delta > 0$ and $\epsilon > 0$ be arbitrary. Denote by $K \subset \mathbb{R}^d$ a compact set such that $P(\mathbf{Y} \in K) \geq 1 - \delta\epsilon/18$. Suppose that $\mathbf{0} \in K$ and define $\mathcal{F}_K$ as the class of 1-Lipschitz functions f supported on K such that $f(\mathbf{0}) = 0$. Such a class, by the Arzelà–Ascoli theorem, is relatively compact for $\mathcal{C}(K)$ and the uniform convergence. Then, there exists a sequence $f_1, \ldots, f_{N_\epsilon}$, such that $\sup_{f\in\mathcal{F}_K}\inf_{k=1\cdots N_\epsilon}||f - f_k||_\infty \leq \epsilon/8$. Therefore, for every $f \in \mathcal{F}_K$, we have

$$\left|\sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})f(\mathbf{Y}_j) - \mathbb{E}(f(\mathbf{Y})|\mathbf{X})\right| \leq \sup_{k=1\cdots N_\epsilon}\left|\sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})f_k(\mathbf{Y}_j) - \mathbb{E}(f_k(\mathbf{Y})|\mathbf{X})\right| + \frac{\epsilon}{4}.$$

In consequence, there exists n_0 such that, for $n \geq n_0$, we have

$$\begin{aligned} &\mathbb{P}\left(\sup_{f\in\mathcal{F}_K}\left|\sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})f(\mathbf{Y}_j)-\mathbb{E}(f(\mathbf{Y})|\mathbf{X})\right|>\frac{\epsilon}{2}\right) \\ &\leq \mathbb{P}\left(\sup_{k=1\cdots N_\epsilon}\left|\sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})f(\mathbf{Y}_j)-\mathbb{E}(f(\mathbf{Y})|\mathbf{X})\right|>\frac{\epsilon}{4}\right) \\ &\leq \sum_{k=1}^{N_\epsilon}\mathbb{P}\left(\left|\sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})f(\mathbf{Y}_j)-\mathbb{E}(f(\mathbf{Y})|\mathbf{X})\right|>\frac{\epsilon}{4}\right)\leq\frac{\delta}{3}, \end{aligned} \tag{7.38}$$

where the last inequality follows from the fact that the weight function is consistent, Note that every $f \in \mathcal{F}_{BL}$ can be approximated by $f\mathbb{1}_K$. This yields

$$\begin{aligned} &\left|\sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})f(\mathbf{Y}_j)-\mathbb{E}(f(\mathbf{Y})|\mathbf{X})\right| \\ &\leq \sup_{f\in\mathcal{F}_K}\left|\sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})f(\mathbf{Y}_j)-\mathbb{E}(f(\mathbf{Y})|\mathbf{X})\right| \\ &+\left|\sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})f(\mathbf{Y}_j)\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y}_j)-\mathbb{E}(f(\mathbf{Y})\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y})|\mathbf{X})\right|. \end{aligned} \tag{7.39}$$

Inequality (7.38) provides an upper bound for the first term. The second one, denoted as A_n, is bounded by

$$\left|\sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})f(\mathbf{Y}_j)\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y}_j)\right|+\left|\mathbb{E}(f(\mathbf{Y})\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y})|\mathbf{X})\right|.$$

Since the weights are positive and $\sup_{\mathbf{x}\in\mathbb{R}^d}|f(\mathbf{x})|\leq 1$,

$$\begin{aligned} A_n &\leq \sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y}_j)+\mathbb{E}(\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y})|\mathbf{X}) \\ &\leq \left|\sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y}_j)-\mathbb{E}(\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y})|\mathbf{X})\right|+2\mathbb{E}(\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y})|\mathbf{X}). \end{aligned}$$

Note that the bound does not depend on the function f. Consequently,

$$\sup_{f\in\mathcal{F}_{BL}}\left|\sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})f(\mathbf{Y}_j)\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y}_j)-\mathbb{E}(f(\mathbf{Y})\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y})|\mathbf{X})\right|$$
$$\leq\left|\sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y}_j)-\mathbb{E}(\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y})|\mathbf{X})\right|+2\mathbb{E}(\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y})|\mathbf{X}).$$

Taking expectations on both sides we obtain

$$\mathbb{E}\left(\sup_{f\in\mathcal{F}_{BL}}\left|\sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})f(\mathbf{Y}_j)\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y}_j)-\mathbb{E}(f(\mathbf{Y})\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y})|\mathbf{X})\right|\right)$$
$$\leq\mathbb{E}\left|\sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y}_j)-\mathbb{E}(\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y})|\mathbf{X})\right|+2\mathbb{P}\left(\mathbf{Y}\notin K\right).$$

Since the weights are consistent, there exists n_1 such that

$$\mathbb{E}\left|\sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y}_j)-\mathbb{E}(\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y})|\mathbf{X})\right|\leq\frac{\delta\epsilon}{12}$$

for all $n>n_1$. Since $\mathbb{P}\left(\mathbf{Y}\notin K\right)\leq\delta\epsilon/12$, if $n>n_1$, then

$$\mathbb{E}\left(\sup_{f\in\mathcal{F}_{BL}}\left|\sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})f(\mathbf{Y}_j)\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y}_j)-\mathbb{E}(f(\mathbf{Y})\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y})|\mathbf{X})\right|\right)\leq\frac{\delta\epsilon}{3}. \tag{7.40}$$

Using Markov's inequality in (7.40), we obtain

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}_{BL}}\left|\sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})f(\mathbf{Y}_j)\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y}_j)-\mathbb{E}(f(\mathbf{Y})\mathbb{1}_{\mathbb{R}^d\setminus K}(\mathbf{Y})|\mathbf{X})\right|>\frac{\epsilon}{2}\right)\leq\frac{2\delta}{3} \tag{7.41}$$

for all $n>n_1$. Finally using (7.38), (7.41) and (7.39), we conclude that

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}_{BL}}\left|\sum_{j=1}^{n} w_j(X;\mathbf{X}^{(n)})f(\mathbf{Y}_j)-\mathbb{E}(f(\mathbf{Y})|\mathbf{X})\right|>\epsilon\right)\leq\delta \tag{7.42}$$

for all $n>N=\max(n_0,n_1)$. □

Proof of Lemma 7..2. Due to the fact that finite second-order moments are not required in our setting, Theorem 2.8 in del Barrio and Loubes (2019) does not directly apply. Their proof, however, relies on the weak convergence of the couplings (the joint measures solving the Kantorovich problem). In our case, we can prove a similar result using Lemma 9 in McCann (1995). Indeed, since the sequences $\{\mu_n\}_n$ and $\{\nu_n\}_n$ are tight with respect to weak convergence, the same result holds also for the class $\Gamma(\mu_n, \nu_n)$ of probabilities on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals $\{\mu_n\}_n$ and $\{\nu_n\}_n$, see Lemma 4.4 in Villani (2008). Note that all the measures π_n, $n \in \mathbb{N}$, belong to $\Gamma(\mu_n, \nu_n)$. Denote by $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ the weak limit of $\{\pi_n\}_n$ along a subsequence; for simplicity, we keep the index n for the subsequence. Lemma 9 (ii) in McCann (1995) implies that the marginals of π are μ and ν. Moreover, the support of π_n is cyclically monotone: indeed, it is contained in the subdifferential of ϕ_n. Therefore, using Lemma 9 (i) in McCann (1995), we know that also π is supported on a cyclically monotone set. Corollary 14 in McCann (1995) yields—since μ is uniformly continuous with respect to the Lebesgue measure—the existence of a unique measure with cyclically monotone support and marginals μ and U_d. As a consequence, $\pi = (\mathbf{Id} \times \phi)\#\mu$. Since this holds along all possible subsequence, we have the weak convergence of $(\mathbf{Id} \times \phi_n)\#\mu_n$ to $(\mathbf{Id} \times \phi)$. At this point, to conclude the proof of Lemma 7..2, we can repeat verbatim the rest of the proof of Theorem 2.8 in del Barrio and Loubes (2019). □

Proof of Lemma 7..3. Note that, for all $\epsilon > 0$, $\sup_{\mathfrak{G}_i \in N_k(\mathfrak{G}_0)} |\mathfrak{G}_i - \mathfrak{G}_0| > \epsilon$ if and only if $\sum_{i=1}^n \mathbb{1}_{|\mathfrak{G}_i - \mathfrak{G}_0| < \epsilon} < k/n$. Since

$$\inf_{\mathbf{u} \in \mathbb{S}_m} \sum_{i=1}^n \mathbb{1}_{|\mathfrak{G}_i - \mathfrak{G}_0| < \epsilon} \longrightarrow \inf_{\mathbf{u} \in \mathbb{S}_m} \mathrm{U}_m(\mathbf{u} + \epsilon\, \mathbb{S}_m) > 0$$

and $k/n \to 0$ as $n \to \infty$,

$$\sup_{\mathfrak{G}_0 \in \mathbb{S}_m} \sup_{\mathfrak{G}_i \in N_k(\mathfrak{G}_0)} |\mathfrak{G}_i - \mathfrak{G}_0| \longrightarrow 0 \quad \text{as } n \to \infty.$$

□

Bibliography

Adib Bagh and Roger J. Wets. Convergence of set-valued mappings: equi-outer semicontinuity. *Set-Valued Anal.*, 4:333–360, 1996.

J. Breckling and R. Chambers. M-Quantiles. *Biometrika*, 75:761–771, 1988.

Guillaume Carlier, Victor Chernozhukov, and Alfred Galichon. Vector quantile regression: An optimal transport approach. *The Annals of Statistics*, 44(3):1165–1192, 2016. doi: 10.1214/15-AOS1401. URL https://doi.org/10.1214/15-AOS1401.

Victor Chernozhukov, Alfred Galichon, Marc Hallin, and Marc Henry. Monge-Kantorovich depth, quantiles, ranks and signs. *The Annals of Statistics*, 45(1):223–256, 2017. doi: 10.1214/16-AOS1450. URL https://doi.org/10.1214/16-AOS1450.

Marco Cuturi. Sinkhorn distances: lightspeed computation of optimal transport. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper/2013/file/af21d0c97db2e27e13572cbf59eb343d-Paper.pdf.

Fernando de Goes, Katherine Breeden, Victor Ostromoukhov, and Mathieu Desbrun. Blue noise through optimal transport. *ACM Trans. pn Graphics*, 31:1–11, 2012. ISSN 0730-0301. doi: 10.1145/2366145.2366190. URL https://doi.org/10.1145/2366145.2366190.

Nabarun Deb and Bodhisattva Sen. Multivariate rank-based distribution-free nonparametric testing using measure transportation. *Journal of the American Statistical Association*, 0:1–16, 2021. doi: 10.1080/01621459.2021.1923508. URL https://doi.org/10.1080/01621459.2021.1923508.

Eustasio del Barrio and Jean-Michel Loubes. Central limit theorems for empirical transportation cost in general dimension. *Ann. Probab.*, 47(2):926–951, 03 2019. doi: 10.1214/18-AOP1275. URL https://doi.org/10.1214/18-AOP1275.

Eustasio del Barrio, Alberto González-Sanz, and Marc Hallin. A note on the regularity of optimal-transport-based center-outward distribution and quantile functions. *J. Multivar. Anal.*, 180:104671, 2020.

Eustasio del Barrio, Alberto GonzÃ¡lez-Sanz, and Jean-Michel Loubes. A central limit theorem for semidiscrete wasserstein distances. arXiv:2105.11721, 2021.

Eustasio del Barrio, Alberto Gonzalez Sanz, and Marc Hallin. Nonparametric multiple-output center-outward quantile regression. *arXiv: Probability*, 2022.

Luc Devroye. Necessary and sufficient conditions for the pointwise convergence of nearest neighbor regression function estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 61:467–481, 1982.

Luc Devroye, Laszlo Gyorfi, Adam Krzyzak, and Gabor Lugosi. On the strong universal consistency of nearest neighbor regression function estimates. *The Annals of Statistics*, 22 (3):1371–1385, 1994. ISSN 00905364. URL http://www.jstor.org/stable/2242230.

Shai Feldman, Stephen Bates, and Yaniv Romano. Calibrated multiple-output quantile regression with representation learning. arXiv:2110.00816, 2021.

Alessio Figalli. On the continuity of center-outward distribution and quantile functions. *Nonlinear Anal.*, 177 :413–21, 2018. ISSN 0362-546X. doi: 10.1016/j.na.2018.05.008. URL https://doi.org/10.1016/j.na.2018.05.008.

Nicolas Fournier and Arnaud Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3):707–738, 2015.

Thomas GallouÃ«t and Quentin MÃ©rigot. A lagrangian scheme à la brenier for the incompressible euler equations. *Found. Comput. Math.*, 18:835–865, 2018. ISSN 1935-8237. URL https://doi.org/10.1007/s10208-017-9355-y.

Promit Ghosal and Bodhisattva Sen. Multivariate ranks and quantiles using optimal transportation and applications to goodness-of-fit testing. arXiv:1905.05340, 2019.

Alberto GonzÃ¡lez-Sanz, Lucas de Lara, Louis BÃ©thune, and Jean-Michel Loubes. Gan estimation of lipschitz optimal transport maps. arXiv:2202.07965, 2022.

LÃ¡szlÃ³ GyÃ¶rfi, Michael Kohler, Adam Krzyzak, and Harro Walk. *A Distribution-Free Theory of Nonparametric Regression.* Springer Series in Statistics. Springer, 2002. ISBN 978-0-387-95441-7.

M. Hallin, D. La Vecchia, and H. Liu. Center-outward r-estimation for semiparametric varma models. *Journal of the American Statistical Association*, 0(0):1–14, 2020. doi: 10.1080/01621459.2020.1832501. URL https://doi.org/10.1080/01621459.2020.1832501.

M. Hallin, D. La Vecchia, and H. Liu. Rank-based testing for semiparametric VAR models: a measure transportation approach. *Bernoulli*, 0(0):1–14, 2022a.

Marc Hallin. On distribution and quantile functions, ranks and signs in $\mathbb{R}^d$: a measure transportation approach. Available at https://ideas.repec.org/p/eca/wpaper/2013-258262.html, 2017.

Marc Hallin. Measure transportation and statistical decision theory. *Annual Review of Statistics and its Application*, 9:401–424, 2022. doi: 10.1146/annurev-statistics-040220-105948. URL https://doi.org/10.1146/annurev-statistics-040220-105948.

Marc Hallin and Miroslav Šiman. Multiple-output quantile regression. In Roger Koenker, Victor Chernozhukov, Xuming He, and L. Peng, editors, *Handbook of Quantile Regression*, pages 185–207. CRC Press, 2018.

Marc Hallin, Davy Paindaveine, and Miroslav Šiman. Multivariate quantiles and multiple-output regression quantiles: from l_1 optimization to halfspace depth. *The Annals of Statistics*, 38(2):635–703, Apr 2010. ISSN 0090-5364. doi: 10.1214/09-aos723. URL http://dx.doi.org/10.1214/09-AOS723.

Marc Hallin, Zudi Lu, Davy Paindaveine, and Miroslav Šiman. Local bilinear multiple-output quantile/depth regression. *Bernoulli*, 21(3):1435–1466, Aug 2015. ISSN 1350-7265. doi: 10.3150/14-bej610. URL http://dx.doi.org/10.3150/14-BEJ610.

Marc Hallin, Eustasio del Barrio, Juan Cuesta-Albertos, and Carlos MatrÃ¡n. Distribution and quantile functions, ranks and signs in dimension d: a measure transportation approach. *The Annals of Statistics*, 49:1139–1165, 2021. doi: 10.1214/20-AOS1996.

Marc Hallin, Daniel Hlubinka, and Š. HudecovÃ¡. Efficient fully distribution-free center-outward rank tests for multiple-output regression and MANOVA. *Journal of the American Statistical Association*, 0:1–43, 2022b. doi: 10.1080/01621459.2021.2021921. URL https://doi.org/10.1080/01621459.2021.2021921.

Jan-Christian HÃ¼tter and Philippe Rigollet. Minimax estimation of smooth optimal transport maps. *The Annals of Statistics*, 49(2):1166–1194, 2021. doi: 10.1214/20-AOS1997. URL https://doi.org/10.1214/20-AOS1997.

R. Koenker and Z. Xiao. Quantile autoregression [with comments, rejoinder]. *Journal of the American Statistical Association*, 101:980–1006, 2006.

Roger Koenker. *Quantile Regression*. Econometric Society Monographs. Cambridge University Press, 2005. doi: 10.1017/CBO9780511754098.

Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1913643.

Roger Koenker, Victor Chernozhukov, Xuming He, and L. Peng, editors. *Handbook of Quantile Regression*. CRC Press, 2018.

D. Konen and D. Paindaveine. Multivariate ρ-quantiles: a spatial approach. *Bernoulli*, 2022.

Linglong Kong and Ivan Mizera. Quantile tomography: using quantiles with multivariate data. *Statistica Sinica*, 22:1589–1610, 2012. ISSN 1017-0405. doi: 10.5705/ss.2010.224. URL http://dx.doi.org/10.5705/ss.2010.224.

E.L. Lehmann and J.P. Romano. *Testing Statistical Hypotheses*. Springer, 2005.

Bruno LÃ©vy, Roya Mohayaee, and Sebastian von Hausegger. A fast semi-discrete optimal transport algorithm for a unique reconstruction of the early universe. arXiv:2012.09074, 2020.

Ashok Makkuva, Amirhossein Taghvaei, Sewoong Oh, and Jason Lee. Optimal transport mapping via input convex neural networks. In *International Conference on Machine Learning*, pages 6672–6681. PMLR, 2020.

Tudor Manole, Sivaraman Balakrishnan, Jonathan Niles-Weed, and Larry Wasserman. Plugin estimation of smooth optimal transport maps. arXiv:2107.12364, 2021.

Robert J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80:309–323, 1995.

Luca Merlo, Lea Petrella, Nicola Salvati, and Nikos Tzavidis. Marginal m-quantile regression for multivariate dependent data. *Computational Statistics & Data Analysis*, page 107500, 2022. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2022.107500. URL https://www.sciencedirect.com/science/article/pii/S0167947322000809.

Jocelyn Meyron. Initialization procedures for discrete and semi-discrete optimal transport. *Computer-Aided Design*, 115:13–22, 2019. ISSN 0010-4485. doi: https://doi.org/10.1016/j.cad.2019.05.037. URL https://www.sciencedirect.com/science/article/pii/S0010448519301770.

Gabriel PeyrÃ© and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and TrendsÂ® in Machine Learning*, 11(5-6):355–607, 2019. ISSN 1935-8237. doi: 10.1561/2200000073. URL http://dx.doi.org/10.1561/2200000073.

Aram-Alexandre Pooladian and Jonathan Niles-Weed. Entropic estimation of optimal transport maps. arXiv preprint arXiv:2109.12004, 2021.

R. Tyrrell Rockafellar. *Convex Analysis*. Princeton Mathematical Series. Princeton University Press, Princeton, N. J., 1970.

R. Tyrrell Rockafellar and Roger J.-B. Wets. *Variational Analysis*. Springer Verlag, Heidelberg, Berlin, New York, 1998.

Robert Serfling. Quantile functions for multivariate analysis: approaches and applications. *Statistica Neerlandica*, 56(2):214–232, 2002. doi: https://doi.org/10.1111/1467-9574.00195. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9574.00195.

Robert Serfling. Depth functions on general data spaces, i. perspectives, with consideration of "density" and "local" depths". Available at https://www.utdallas.edu~serfling, 2019a.

Robert Serfling. Depth functions on general data spaces, ii. formulation and maximality, with consideration of the tukey, projection, spatial, and "contour" depths. Available at https://www.utdallas.edu~serfling, 2019b.

Robert Serfling and Yijun Zuo. General notions of statistical depth function. *The Annals of Statistics*, 28(2):461–482, 2000. doi: 10.1214/aos/1016218226. URL https://doi.org/10.1214/aos/1016218226.

Hongjian Shi, Mathias Drton, Marc Hallin, and Fang Han. On universally consistent and fully distribution-free rank tests of vector independence. *The Annals of Statistics*, 0(0): 1–14, 2021a.

Hongjian Shi, Mathias Drton, Marc Hallin, and Fang Han. Center-outward sign- and rank-based quadrant, spearman, and kendall tests for multivariate independence. arXiv:2111.15567, 2021b.

Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper/2015/file/8d55a249e6baa5c06772297520da2051-Paper.pdf.

Charles J. Stone. Consistent nonparametric regression. *Ann. Statist.*, 5(4):595–620, 07 1977. doi: 10.1214/aos/1176343886. URL https://doi.org/10.1214/aos/1176343886.

J. W. Tukey. Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians, ver, 1975*, 2:523–531, 1975. URL https://ci.nii.ac.jp/naid/10029477185/en/.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer, New York, NY, 1996.

C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, Providence, Rhode Island, 2003.

C. Villani. *Optimal Transport: Old and New*. Springer-Verlag Berlin Heidelberg, 2008.

Part III

Conclusion and final remarks

Concluding remarks

Throughout this thesis the central topic has been optimal transport, seen, in particular, from a statistical point of view. The work has beed divided into two distinct parts; one dedicated to the asymptotic study of the problem and the other dedicated to the properties of the center-outward distribution function.

The content of the first part has been presented in four chapters. The first of them, Chapter 2, has been devoted to the study of the fluctuations of the Wasserstein distance in Euclidian spaces and strictly convex costs. Motivated by the obtention of a central limit theorem of the fluctuations we derived stability results on the potentials and transport maps. Those results were based on the set-valued mapping topology.

The initial aim of Chapter 3 was to study, as a case where the bias converges faster than the parametric rate, the semidiscrete optimal transport problem. However, instead of limiting the study to bounding the bias and applying the central limit theorem of the fluctuations, a new proof technique was proposed. That covers the cases where the fluctuations are not Gaussians. Moreover, we found that, under some regularity assumptions, the potentials (solutions of the dual problem) are well separated, so that a central limit theorem could be derived for them. As a consequence, the weak limit of the Voronoi tessellation was obtained.

Chapter 4 was motivated by the paper of Mena and Niles-Weed (2019) where they proved that fluctuations of the Sinkhorn cost were always asymptotically Gaussians. The first result was to show that the rate of convergence of the bias (difference between the expected value of the Sinkhorn cost and the population value) towards 0 is faster than that of its variance. That enables to obtain the central limit theorem of the cost centered at the population value. The proof took advantage of the regularity of the potentials and the fact that they are well separated. The major mathematical innovation was the use of operator theory. That is, if we know that the norm in $\mathcal{C}^s$ of the difference between the empirical and the population potential $f_n - f^*$ follows the rate $n^{\frac{1}{2}}$ and –by using empirical process theory– that the empirical process $\sqrt{n}(P_n - P)$ is bounded as an element of the dual of $\mathcal{C}^s$. Then the basic inequality $(P_n - P)(f_n - f^*) \leq \|f_n - f^*\|_{\mathcal{C}^s} \|P_n - P\|_{(\mathcal{C}^s)'}$ leads to this faster rate. This study of the regularized optimal transport problem through the potentials gave also a faster rate of the divergence.

Chapter 5 was a continuation of Chapter 4. Once we knew the correct rate of the plans, potentials and divergences, the question was about their weak limits. We differentiated the potentials with respect to the empirical process, giving the first order Taylor development. The limit of the potentials followed directly from the application of the delta-method. As a consequence, we showed that the limits of plans and divergences can be derived from that of the potentials.

The content of the second part has been presented in two chapters. The first of them, Chapter 6, has been devoted to the study of the regularity of the center-outward distribution and quantile functions. We observed that the same strategies of Caffarelli's theory can be applied to this problem. Of course, taking care of the singularity at 0 of the spherical reference .

Chapter 7 proposed a method to deal with the problem of multiple output quantile regression. This had been an open problem due to the lack of satisfactory enough concept of multivariate quantile. That is, one satisfying the property of probability control—which, in one dimension gives meaning to the median, as a central point that leaves half (and only half) of the probabilistic mass on both sides of it—, here derived from the distribution freeness of the multivariate distribution function. We proposed a feasible estimator that is consistent under weak assumptions on the shape and smoothness of the probability distribution. Real and synthetic examples showed its performance. In particular, its capacity to catch the heterocedasticity and trend of a model.

To conclude, this thesis applied several techniques of mathematical analysis to the statistical framework of the optimal transport problem. In particular, convex analysis was the main tool to show the convergence of the optimal transport maps and potentials (see Chapter 2) and the consistency of the estimator of the conditional center-outward distribution function (see Chapter 7). Operator theory was the core of the proofs of the weak limits of the Sinkhorn regulaized optimal transport. The arguments that lead to find the limits of the limits given in the semidiscrete optimal transport were developed under the solid foundations of the theory of empirical processes. Finally, partial differential equations and convex analysis provided the rationale for the proof of the regularity of the center-outward distribution.

Future work, consequences and open problems

This manuscript closes several open questions of optimal transport but also opens several doors for future works. This section examines some of these open questions and possible lines of upcoming research that can be drawn from this thesis.

With respect to the weak limits of optimal transport, after this thesis and the parallel works of Hundrieser et al. (2022); Goldfeld et al. (2022b); Klatt et al. (2020); Goldfeld et al. (2022a), the open problems are scarce and quite difficult. One of those is in the two dimensional case, as stated by Hundrieser et al. (2022) the limit of the transport cost follows the parametric rate. However, if the cost is null—i.e. both probability measures are equal—the limit is degenerated, the convergence is faster, and the rate can be improved. Ambrosio et al. (2016) proposed a new way to deal with this problem by using a linearization of the problem through the heat kernel. That gives the limit of the expectation with a rate n^{-1} up to logarithm factors. Then it could be possible to obtain a weak limit of the optimal transport cost via a suitable control of the fluctuations, see the conjecture of Ledoux (2019).

Another open problem comes from the possibility of approximate fast enough the Wasserstein distance from the Sinkhorn divergence by making the ϵ small at the same time as the sample size increases. As the reader can see, the bounds of the Sinkhorn divergence (see Chapter 4) are exponential in ϵ. Then even if the rate is $1/n$, a choice of ϵ_n such that the Sinkhorn divergence approximates the Wasserstein distance is not possible. Up to my knowledge, the literature has not offered an answer to this fact yet.

The stability results of Chapter 2 and that of the ranks and quantiles are consequence of the point-wise convergence of sub-gradients. Indeed, the techniques of Segers (2022); Hallin et al. (2021); del Barrio et al. (2022) are based on the Fell topology, which does not have nice properties for non-locally compact spaces. An open problem is the generalization of those results to Banach spaces. However, this is a challenging topic (Segers, 2022). Moreover, these result would open the possibility to generalize Hallin's distribution function to non-locally compact spaces. This definition of infinite-variate distribution function would make possible to make rank-based inference —which includes applications to goodness-of-fit or independence testing, outlier detection or quantile regression— for functional data or, via some embeddings to reproducing kernel Hilbert spaces, to even more abstract spaces.

The methodology and results emanating from this thesis, apart from being interesting in themselves and maybe important for the optimal transportation community, could be significant for the statistical community. The techniques that have been presented are reproducible in other areas of machine learning and statistics. In general, a learning problem can be stated as follows; let $\mathcal{X}$ be a Polish space with metric d, and X be a r.v. taking values on $\mathcal{X}$, the goal is to solve $f^* \in \arg\min_{f \in \mathcal{F}} E(\ell(f(X)))$, where $\mathcal{F}$ is a class of functions and ℓ a loss function. The risk is defined as $\mathcal{R} = E(\ell(f^*(X)))$. In machine learning, the function class is quite common parametrized by a neural network. Recently, in Béthune et al. (2022) we observed that the class of Lipschitz neural networks–a class using a sorting activation function with bounded weights (see, eg. Anil et al. (2019) and references therein)–behaves better against adversarial attacks than a classical neural network. That is, its robustness can be certified. Moreover, the same technique used to bound the fluctuations (or a control of the variance) of the empirical optimal transport cost bounds here the variance of the empirical risk $\mathcal{R}_n$. More precisely, we can obtain

$$\text{Var}(\mathcal{R}_n) \leq \frac{L}{n} E(d(X_1', X_1)^2),$$

where L is the Lipschitz constant of the loss and X_1' an independent copy of X. The variance is thus controlled. This complements the exhaustive analysis developed in Béthune et al. (2022), and *a fortiori* bounds the fluctuations of the empirical risk of the transport-based classifier proposed in Serrurier et al. (2021).

In a complete different framework, the minimization of the variance of the different clusters is the core of the well-known K-means procedure—undoubtedly one of the oldest and simplest method of clustering. As conjectured in Chapter 3, the arguments that lead to obtaining the distributional limits of the sets corresponding to the Hotelling problem can be applied to the clusters produced by K-means. This would provide asymptotically valid confidence intervals for the clusters obtained by the K-means procedure over a sample.

Talk of reproducible techniques may seem intangible—even fatuous—to the practitioner. But let us now see how the results (and not only the techniques) can be applied to different problems related to statistics. One as old as important is that of nonparametric hypothesis testing. In the context of a goodness-of-fit test—i.e. $H_0 : P = Q$ vs $H_0 : P \neq Q$—the Wasserstein distance, which metrizes the weak convergence while taking into account the geometry of space, appears to be a successful option. In fact, we found in González-Delgado et al. (2021) that in the flat torus of dimension 2, the statistic based on this distance is asymptotically consistent under fixed alternatives, while capturing the metric structure of the flat torus. This work has had important applications in structural biology, where the comparison of protein structures is a crucial (see eg. González-Delgado et al. (2022) and references therein).

However, in higher dimension, the curse of the dimension affects the optimal transport, so that the Wasserstein distance is not convenient to derive a goodness-of-fit testing procedure. Then the Sinkhorn divergence, whose statistical complexity is inherent to the dimension, may offer a more satisfactory option. But in order to do so, it is necessary to have knowledge of the weak limit

$$\frac{n\,m}{n+m} D_1(P_n, Q_m) \xrightarrow{w} \mathbb{Z}_{P,\lambda},$$

under $H_0 : P = Q$, given in Chapter 5. Even if the limit $\mathbb{Z}_{P,\lambda}$ is not free –i.e. it may depend on P–, it can be made free by means of center-outward ranks, see Shi et al. (2022); Deb and Sen (2019). Similarly, an independence test can be created.

The Sinkhorn divergence is, as the name indicates, a divergence and not a distance. The triangular inequality is not satisfied. This means that many of the historically known results for metric spaces cannot be applied here. However, setting a reference distribution $\mathcal{U}$, the function defined for $\mathrm{P}, \mathrm{Q} \in \mathcal{P}(\Omega)$, as the $L^2(\mathcal{U})$ norm between the Stinkhorn potentials

$$\|g^{\epsilon}_{P,\mathcal{U}} - g^{\epsilon}_{Q,\mathcal{U}}\|_{L^2(\mathcal{U})},$$

is a distance. Moreover, this distance defines an embedding of the space of probability measures to the Hilbert space $L^2(\mathcal{U})$. That can be used, as in Bachoc et al. (2022), to define a kernel K on the space of probability distributions. The exhaustive analysis of the Stinkhorn potentials of Chapter 4 creates the theoretical framework to derive the properties of this distance (and thus the kernel). The study of further properties of this kernel will be the scope of future research.

Bibliography

Luigi Ambrosio, Federico Stra, and Dario Trevisan. A pde approach to a 2-dimensional matching problem. *Probability Theory and Related Fields*, 173:433–477, 2016.

Cem Anil, James Lucas, and Roger Baker Grosse. Sorting out lipschitz function approximation. In *ICML*, 2019.

Franccois Bachoc, Louis B'ethune, Alberto González-Sanz, and Jean-Michel Loubes. Gaussian processes on distributions based on regularized optimal transport. *ArXiv*, abs/2210.06574, 2022.

Louis Béthune, Alberto González-Sanz, Franck Mamalet, and Mathieu Serrurier. Pay attention to your loss: understanding misconceptions about 1-lipschitz neural networks. *Advances in Neural Information Processing Systems.*, 2022.

Nabarun Deb and Bodhisattva Sen. Multivariate rank-based distribution-free nonparametric testing using measure transportation. *J. Amer. Statist. Assoc.*, 2019. To appear.

Eustasio del Barrio, Alberto Gonzalez Sanz, and Marc Hallin. Nonparametric multiple-output center-outward quantile regression. *arXiv: Probability*, 2022.

Ziv Goldfeld, Kengo Kato, Gabriel Rioux, and Ritwik Sadhu. Limit theorems for entropic optimal transport maps and the sinkhorn divergence. *arXiv preprint*, 2022a.

Ziv Goldfeld, Kengo Kato, Gabriel Rioux, and Ritwik Sadhu. Statistical inference with regularized optimal transport. *arXiv: Probability*, 2022b.

Javier González-Delgado, Alberto González-Sanz, Juan Cortés, and Pierre Neuvial. Two-sample goodness-of-fit tests on the flat torus based on wasserstein distance and their relevance to structural biology. *arXiv: Probability*, 2021.

Javier González-Delgado, Amin Sagar, Christophe Zanon, Kresten Lindorff-Larsen, Pau Bernado, Pierre Neuvial, and Juan Cortés. Wasco: A wasserstein-based statistical tool to compare conformational ensembles of intrinsically disordered proteins, 12 2022.

Marc Hallin, Eustasio del Barrio, Juan Cuesta-Albertos, and Carlos Matrán. Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *The Annals of Statistics*, 49(2):1139 – 1165, 2021. doi: 10.1214/20-AOS1996.

Shayan Hundrieser, Marcel Klatt, Thomas Staudt, and Axel Munk. A unifying approach to distributional limits for empirical optimal transport. *Pre-Print*, 02 2022. URL `https://arxiv.org/pdf/2202.12790.pdf`.

Marcel Klatt, Carla Tameling, and Axel Munk. Empirical regularized optimal transport: Statistical theory and applications. *SIAM J. Math. Data Sci.*, 2:419–443, 2020.

Michel Ledoux. A fluctuation result in dual sobolev norm for the optimal matching problem. *preprint*, 2019.

Gonzalo Mena and Jonathan Niles-Weed. Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/5acdc9ca5d99ae66afdfe1eea0e3b26b-Paper.pdf.

Johan Segers. Graphical and uniform consistency of estimated optimal transport plans. *arXiv*, 2022.

Mathieu Serrurier, Franck Mamalet, Alberto Gonzalez-Sanz, Thibaut Boissin, Jean-Michel Loubes, and Eustasio del Barrio. Achieving robustness in classification using optimal transport with hinge regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 505–514, June 2021.

Hongjian Shi, Mathias Drton, and Fang Han. Distribution-free consistent independence tests via center-outward ranks and signs. *J. Amer. Statist. Assoc.*, 117(537):395–410, 2022. ISSN 0162-1459. doi: 10.1080/01621459.2020.1782223. URL https://doi.org/10.1080/01621459.2020.1782223.

APPENDIX A

Other collaborations

The optimal transport problem as a statistical tool is nowadays a hot research topic. The speed at which research is advancing in this field makes collaborations the main pillar for successful scientific production. A researcher's own skills are complemented by those of his co-workers. This manuscript does not include the full collaboration of the author's three years of scientific adventure. The decision to omit certain results (and papers) in this document became indispensable for its readability, maintenance of a consistent line of argument and drawing of conclusive statements. However, these works have been, in a way, indispensable in the course of acquiring knowledge. Therefore, this final section contains the list of those omitted works.

Publications

- François Bachoc, Louis Béthune, Alberto González Sanz, Jean-Michel Loubes. Gaussian Processes on Distributions based on Regularized Optimal Transport. To appear in Proceedings of The 26th International Conference on Artificial Intelligence and Statistics. (2023+)

- Lucas De Lara, Alberto Gonzalez Sanz, and Jean-Michel Loubes. Diffeomorphic Registration using Sinkhorn Divergences. To appear in SIAM Journal on Imaging Sciences. (2022).

- Louis Béthune, Thibaut Boissin, Mathieu Serrurier, Franck Mamalet, Corentin Friedrich and Alberto González Sanz. Pay attention to your loss: understanding misconceptions about 1-Lipschitz neural networks. Advances in Neural Information Processing Systems. (2022).

- Laurent Risser, Alberto González Sanz, Quentin Vincenot and Jean-Michel Loubes. Tackling Algorithmic Bias in Neural-Network Classifiers using Wasserstein-2 Regularization. J Math Imaging Vis. (2022).

- Serrurier, Mathieu, Franck Mamalet, Alberto González Sanz, Thibaut Boissin, Jean-Michel Loubes and Eustasio del Barrio. Achieving robustness in classification using optimal transport with hinge regularization. CVPR. pp. 505-514. (2021).

Preprints

- Alberto González Sanz, Lucas De Lara, Louis Béthune, Jean-Michel Loubes. GAN Estimation of Lipschitz Optimal Transport Maps. (2022).
- Javier González-Delgado, Alberto González Sanz, Juan R. Cortés and Pierre Neuvial. Two-sample goodness-of-fit tests on the flat torus based on Wasserstein distance and their relevance to structural biology. (2021).
- Lucas de Lara, Alberto González Sanz, Nicholas Asher and Jean-Michel Loubes. Transport-based Counterfactual Models. (2021).