



**HAL**  
open science

# A Central Limit Theorems for Semidiscrete Wasserstein Distances

Eustasio del Barrio, Alberto González-Sanz, Jean-Michel Loubes

► **To cite this version:**

Eustasio del Barrio, Alberto González-Sanz, Jean-Michel Loubes. A Central Limit Theorems for Semidiscrete Wasserstein Distances. 2023. hal-03232450v3

**HAL Id: hal-03232450**

**<https://hal.science/hal-03232450v3>**

Preprint submitted on 16 Feb 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Central Limit Theorems for semi-discrete Wasserstein Distances

EUSTASIO DEL BARRIO<sup>1,a</sup>, ALBERTO GONZÁLEZ SANZ<sup>2,b</sup> and JEAN-MICHEL LOUBES<sup>2,d</sup>

<sup>1</sup>IMUVA, Universidad de Valladolid, Spain. <sup>a</sup>[tasio@eio.uva.es](mailto:tasio@eio.uva.es), <sup>b</sup>[alberto.gonzalez.sanz@math.univ-toulouse.fr](mailto:alberto.gonzalez.sanz@math.univ-toulouse.fr)

<sup>2</sup>IMT, Université de Toulouse, France. <sup>b</sup>[alberto.gonzalez.sanz@math.univ-toulouse.fr](mailto:alberto.gonzalez.sanz@math.univ-toulouse.fr),

<sup>d</sup>[loubes@math.univ-toulouse.fr](mailto:loubes@math.univ-toulouse.fr)

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  $p$ -Wasserstein 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.

*Keywords:* Central limit theorem; Laguerre cells; Optimal transport; Optimal transport potentials; Semi-discrete optimal transport

## 1. Introduction

A large number of problems in statistics or computer science require the comparison between histograms or, more generally, measures. Optimal transport (OT) has proven to be an important tool to compare probability measures since it enables to define a metric over the set of distributions which conveys their geometric properties, see [Verdinelli and Wasserman \(2019\)](#). Moreover, together with the convergence of the moments, it metrizes the weak convergence, see Chapter 7.1. in [Villani \(2003\)](#). It is nowadays used in a large variety of fields, in probability and statistics. In particular in machine learning, OT based methods have been developed to tackle problems in fairness as in [de Lara et al. \(2021\)](#), [Gordaliza et al. \(2019\)](#), [Jiang et al. \(2020\)](#), [?](#), in domain adaptation [Shen et al. \(2018\)](#), or transfer learning [Gayraud, Rakotomamonjy and Clerc \(2017\)](#). Hence, there is a growing need for theoretical results to support such applications and provide theoretical guarantees on the asymptotic distribution.

The most general formulation of the optimal transport problem considers  $X, \mathcal{Y}$  both Polish spaces. We use the notation  $\mathcal{P}(X)$  (resp.  $\mathcal{P}(\mathcal{Y})$ ) for the set of Borel probability measures on  $X$  (resp.  $\mathcal{Y}$ ). The optimal transport problem between  $P \in \mathcal{P}(X)$  and  $Q \in \mathcal{P}(\mathcal{Y})$  for the cost  $c : X \times \mathcal{Y} \rightarrow [0, \infty)$  is formulated as minimization problem

$$\mathcal{T}_c(P, Q) := \inf_{\gamma \in \Pi(P, Q)} \int_{X \times \mathcal{Y}} c(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y}), \quad (1)$$

where  $\Pi(P, Q)$  is the set of probability measures  $\pi \in \mathcal{P}(X \times \mathcal{Y})$  such that  $\pi(A \times \mathcal{Y}) = P(A)$  and  $\pi(\mathcal{Y} \times B) = Q(B)$  for all  $A, B$  measurable sets.

If  $c$  is continuous and there exist two continuous functions  $a \in L^1(P)$  and  $b \in L^1(Q)$  such that

$$\text{for all } (\mathbf{x}, \mathbf{y}) \in \text{supp}(P) \times \text{supp}(Q), \quad c(\mathbf{x}, \mathbf{y}) \geq a(\mathbf{x}) + b(\mathbf{y}), \quad (2)$$

then the Kantorovich problem (1) can be formulated in a dual form, as

$$\mathcal{T}_c(P, Q) = \sup_{(f, g) \in \Phi_c(P, Q)} \int f(\mathbf{x}) dP(\mathbf{x}) + \int g(\mathbf{y}) dQ(\mathbf{y}), \quad (3)$$

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 for instance Theorem 5.10 in Villani (2008). 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  $(\psi, \varphi)$  solves (3).

This work focuses on the semi-discrete optimal transport, i.e. when one of both probabilities, let us assume  $P$ , is supported on a discrete set  $\mathbb{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ . Lemma 2.1 shows that  $\mathcal{T}_c(P, Q)$  is equivalent to the maximization of

$$g_c(P, Q, \mathbf{z}) = \sum_{i=1}^N z_i p_i + \int \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\} dQ(\mathbf{y}), \quad (4)$$

The solution  $\mathbf{z}^*$  (unique up to additive constant under some assumptions) of (4) defines set of Laguerre cells

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

which are generalizations of the Voronoi cells –equivalent to  $\text{Lag}_k(\mathbf{0})$  for the quadratic cost. From an economical perspective semi-discrete optimal transport plays an important role. We refer to Galichon’s monograph Galichon (2016) and references therein. Let us consider the example contained in chapter 5.1, “*Hotelling’s location model*”, 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 which enjoys the properties to be 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 prize of the fountain  $i$ . The fraction of population that prefers to consume from the fountain  $i$  is then  $\text{Lag}_k(\mathbf{z})$ . Under market clearing –supply equals demand –each fountain is used to its full capacity and the problem of determining the prizes reduces to (4).

Other important applications of the semi-discrete optimal transport are quantization and sampling –where the goal of both is to reduce a probability distribution  $Q$  into a finitely supported one. It is well known (cf. Lemma 3.4 in Graf and Luschgy (2000) eg.) that the quantization error is the semi-discrete optimal transport between  $Q$  and its quantized version and the optimal quantizer minimizes, among all probability measures supported in a set of  $N$  elements, the optimal transport cost. In the quadratic case this gives the  $k$ -means clustering. Further applications include the resolution of the incompressible Euler equation using Lagrangian methods (Gallouët and Mérigot, 2018), non-imaging optics; matching between a point cloud and a triangulated surface; seismic imaging (Meyron, 2019), astronomy (Lévy, Mohayaee and von Hausegger, 2020). From a statistical point of view, Goodness-of-fit-tests based on semi-discrete optimal transport enable to detect deviations from a density map to have  $P \neq Q$ , by using the fluctuations of  $\mathcal{W}(P_n, Q)$ , see Hartmann and Schuhmacher (2020).

## 1.1. Contributions

In this paper we consider observations drawn from two mutually independent samples  $\mathbf{X}_1, \dots, \mathbf{X}_n$  and  $\mathbf{Y}_1, \dots, \mathbf{Y}_m$  i.i.d. with laws  $P$  and  $Q$ . Let  $P_n = \frac{1}{n} \sum_{k=1}^n \delta_{\mathbf{X}_k}$  and  $Q_m = \frac{1}{m} \sum_{k=1}^m \delta_{\mathbf{Y}_k}$  be the corresponding empirical measures. The optimal transport cost between the empirical distributions  $\mathcal{T}_c(P_n, Q_m)$  defines a random variable.

We are concerned with the asymptotic behaviour of  $\sqrt{\frac{nm}{n+m}} \{\mathcal{T}_c(P_n, Q_m) - \mathcal{T}_c(P, Q)\}$  in a general semi-discrete setting. We propose a new proof that relies on the framework introduced in [Cárcamo, Cuevas and Rodríguez \(2020\)](#) and considers the Hadamard derivative of the sup of the process with respect to  $\ell^\infty$  topology. The link to the CLT for optimal transport cost comes from the fact that the dual formulation of the transport problem is, in fact, a supremum of functions. Hence if such functions live in a Donsker class  $\mathcal{B}$  (see [Vaart and Wellner \(1996\)](#)), then we can obtain the CLT by proving differentiability of the supremum in  $\ell^\infty(\mathcal{B})$  and applying then the general delta-method.

Hence this paper first covers and generalizes the results of [Sommerfeld and Munk \(2018\)](#) for a semi-discrete  $P$  approximated by  $P_n$  and a general probability distribution  $Q$  to handle all cases of the semi-discrete framework. Moreover, the computation of  $\mathcal{T}_c(P_n, Q)$  is not easy in general, see for instance [Gallouët and Mérigot \(2018\)](#). Consequently, an interesting problem, also for applications, becomes its approximation by a  $Q_m$ , an estimation of  $Q$ . Hence we also provide the asymptotic behaviour of  $\sqrt{m} \{\mathcal{T}_c(P, Q_m) - \mathcal{T}_c(P, Q)\}$ . Surprisingly, Theorem 2.4 yields that it tends to

$$\sup_{\mathbf{z} \in \text{Opt}_c(P, Q)} \mathbb{G}_Q \left( \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\} \right),$$

where  $\text{Opt}_c(P, Q)$  is the set of optimal transport potentials and  $\mathbb{G}_Q$  is the Brownian bridge in  $\mathcal{F}_c^K$  (both will be defined more precisely later) with mean zero and covariance

$$(f, g) \mapsto \int f(\mathbf{y})g(\mathbf{y})dQ(\mathbf{y}) - \int f(\mathbf{y})dQ(\mathbf{y}) \int g(\mathbf{y})dQ(\mathbf{y}).$$

Finally we provide in Section 2 a unified general result that describes the asymptotic distribution of the empirical transport cost between a probability  $P$  supported in the finite set  $\mathbb{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subset \mathcal{X}$  and  $Q \in \mathcal{P}(\mathcal{Y})$  under the minimal assumption  $\int c(\mathbf{x}_i, \mathbf{y})dQ(\mathbf{y}) < \infty$ , for all  $i = 1, \dots, N$ , for all cases:

- (One sample case for empirical discrete distribution  $P_n$ )

$$\sqrt{n} (\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q)) \xrightarrow{w} \sup_{\mathbf{z} \in \text{Opt}_c^0(P, Q)} \sum_{i=1}^N z_i \mathbb{U}_i,$$

and 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 empirical distribution  $Q_m$ )

$$\sqrt{m} (\mathcal{T}_c(P, Q_m) - \mathcal{T}_c(P, Q)) \xrightarrow{w} \sup_{\mathbf{z} \in \text{Opt}_c^0(P, Q)} \mathbb{G}_Q^c(\mathbf{z}).$$

- (Two sample case) If  $n, m \rightarrow \infty$ , with  $\frac{m}{n+m} \rightarrow \lambda \in (0, 1)$ , then

$$\sqrt{\frac{nm}{n+m}} (\mathcal{T}_c(P_n, Q_m) - \mathcal{T}_c(P, Q)) \xrightarrow{w} \sup_{\mathbf{z} \in \text{Opt}_c^0(P, Q)} \left( \sqrt{\lambda} \sum_{i=1}^N z_i \mathbb{U}_i + (\sqrt{1-\lambda}) \mathbb{G}_Q^c(\mathbf{z}) \right).$$

The fact that the curse of dimensionality does not seem to affect the semi-discrete case for both probabilities is quite astonishing. But it is partially hidden in the assumption that the set  $\mathbb{X}$  has a fixed size. For a better understanding, Theorem 2.6 provides, in the context of quadratic costs, a bound which studies the effect of the choice of a discretization with size the one of the set  $\mathbb{X}$ . It highlights a natural trade-off between the discretization scheme of the distribution and the sampling of the distribution.

Moreover, in cases where  $P, Q \in \mathcal{P}(\mathbb{R}^d)$  with  $Q$  having a Lebesgue negligible boundary and being absolutely continuous with respect to  $\ell_d$ , and the cost  $c$  satisfying conditions (A1) through (A3), the limits can be made more explicit and the supremum in the previous limits can be computed. These results are presented in Section 3, where the transport potential is unique up to additive constants. Under some regularity assumptions on the cost and on  $Q$ , the limit is no longer a supremum but instead a centered Gaussian random variable.

The contributions of Section 4 are two-fold. The first part studies the semi-discrete OT in manifolds and provides, to the best of our knowledge, the first Central Limit Theorem for the solutions of the dual problem (3). The second part studies the asymptotic distribution of the Laguerre cells. It should be noted that these results cannot be generalized to continuous distributions. If both probabilities are continuous and the space is not one-dimensional, we cannot expect such a central limit for the potentials, since the expected value of the transport cost estimate converges at a rate of  $O(n^{-\frac{1}{d}})$  and not  $O(n^{-\frac{1}{2}})$ . In the case where the two samples are discrete, even if such a rate is  $O(n^{-\frac{1}{2}})$ , the lack of uniqueness in the dual problem does not allow for the proof of such results. As a result, the semi-discrete case is the only one where such results for the potentials of the OT problem in general dimension can be expected. Regarding the Laguerre cells, we obtain the weak limits of the  $L^p$  metric (see Vitale (1985) for definition) between the empirical and population cells. Additionally, we provide asymptotic confidence intervals in terms of the Hausdorff distance. In all cases, the parametric ratio is achieved. We conjecture that the same approach would provide similar bounds for the empirical Voronoi cells, the solutions of the  $k$ -means clustering. Finally, Remark 4.9 bounds the  $L^1(\ell_d)$  norm between empirical and population transport maps, opening new doors of research in this direction.

The last section is dedicated to applying some of the results discussed earlier to the Hotelling's location model. We present asymptotic confidence intervals for prices and demand sets in a simulated location model in Brooklyn (NYC).

## 1.2. Prior and current works in weak limits of optimal transport

The asymptotic distribution of the empirical transport cost  $\mathcal{T}_c(P_n, Q_m)$  has been explored in several works. For the general dimensional case, a Central Limit Theorem (CLT) for the centered process has been established using the Efron-Stein inequality. In particular, del Barrio and Loubes (2019), del Barrio, Sanz and Loubes (2021), González-Delgado et al. (2021) have proven that  $\sqrt{\frac{nm}{n+m}}(\mathcal{T}_c(P_n, Q_m) - \mathbb{E}\mathcal{T}_c(P, Q))$  converges to a Gaussian distribution. Mena and Niles-Weed (2019) also proves a similar result for the regularized optimal transport cost. Under additional conditions, this result can be extended to the semi-discrete framework. In particular, when  $P$  is finitely supported, but  $Q$  is absolutely continuous with respect to the Lebesgue measure and has a convex support, del Barrio and Loubes (2019) shows that for the quadratic cost, the limit  $\sqrt{n}(\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q))$  is Gaussian. Their argument relies on the differentiability properties of the optimal transport problem, but it is restricted by the assumption of uniqueness of the optimal transport potential, which is not always guaranteed for general costs or Polish spaces. Similar results have been established for the case where both  $P$

and  $Q$  are supported on a finite or countable set. [Sommerfeld and Munk \(2018\)](#) and [Taming, Sommerfeld and Munk \(2019\)](#) respectively prove that, in these settings,  $\sqrt{\frac{nm}{n+m}}(\mathcal{T}_c(P_n, Q_m) - \mathcal{T}_c(P, Q))$  converges to a weak limit  $X$ , which is the supremum of a Gaussian process. Their proof relies on the identification of the space of distributions supported in a finite set  $\mathbb{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  with a subset of  $\mathbb{R}^N$ , and then on a proof based on the directional Hadamard differentiability of the functional  $(\mathbf{p}, \mathbf{q}) \mapsto \mathcal{T}_c(\sum_{i=1}^N p_i \delta_{\mathbf{x}_i}, \sum_{i=1}^N q_i \delta_{\mathbf{x}_i})$ . The result of [Sommerfeld and Munk \(2018\)](#) establishes that, if  $P$  and  $Q$  are both supported in a finite set, then

$$\sqrt{n}(\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q)) \xrightarrow{w} \sup_{\mathbf{z} \in \text{Opt}_c^0(P, Q)} \mathbb{G}(\mathbf{z}), \text{ where } \mathbb{G}(\mathbf{z}) := \sum_{i=1}^N z_i U_i.$$

Lately [Taming, Sommerfeld and Munk \(2019\)](#) extended the same result for probabilities supported in countable spaces. Yet this approach does not hold for probabilities non supported on finite or countable sets.

The parametric rate obtained for the limits in the semi-discrete case is consequence of *the adaptability of the optimal transport problem to lower complexities*. This was first observed by [Weed and Bach \(2019\)](#). During the review process of our work, two new papers by the same authors underlined this fact. [Hundrieser, Staudt and Munk \(2022\)](#) observes that the covering numbers of a class of functions are invariant after conjugation, which is exactly what happens here for the class  $\mathcal{F}_K^c$ . As a consequence, [Hundrieser et al. \(2022\)](#) focuses on the cases where the complexity is small enough to obtain the parametric rate and therefore weak limits. However, since they assume bounded costs, the semi-discrete case is partially covered. We avoid this assumption by a control of the quantity  $K_{n,m}$ , described above. We believe that this technique could successfully address their case, where the bound of the potentials could be the more accurate given in [del Barrio, Sanz and Loubes \(2021\)](#), eq. (2.5).

The Hadamard derivative of supremum type functionals as a way to provide weak limits of the Wasserstein distance has been proposed also in [Goldfeld et al. \(2022a\)](#), [Sadhu, Goldfeld and Kato \(2021\)](#) for Gaussian-convoluted measures. However, although related, this is not the Wasserstein distance—they will coincide when the variance of the convolution tends to zero. This supremum functional derivative argument has proven to be valid in other simplifications of the transport problem; such as sliced or Sinkhorn [Goldfeld et al. \(2022b\)](#). However, in these cases the same results can be obtained by other methods, e.g. [Manole, Balakrishnan and Wasserman \(2022\)](#) uses the quantile representation in the sliced model and [del Barrio et al. \(2022\)](#) the centered-in-expectation CLT of [Mena and Niles-Weed \(2019\)](#) and a bound of the expectation.

Regarding weak limits of the optimal transport potentials, two recent pre-prints ([Goldfeld et al., 2022c](#), [González Sanz, Loubes and Niles-Weed, 2022](#)) obtain the weak limits of the potentials of the Sinkhorn regularized transport problem.

### 1.3. Notation

For convenience, we list here the main notation that will be used throughout the document. Unless otherwise stated, we denote by  $(\tilde{\Omega}, \mathcal{A}, \mathbb{P})$  the underlying probability space. The the probability  $P$  is always assumed to be supported in the set  $\mathbb{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ . The notation  $\ell_d$  is reserved for the Lebesgue measure in  $\mathbb{R}^d$ , whereas  $\mathcal{H}^{d-1}$  denotes the  $d - 1$ -dimensional Hausdorff measure. We use boldface for the vectors, a notation that is not used, however, for their components, which will be expressed by a sub-index. For instance,  $\mathbf{z} = (z_1, \dots, z_N)$ . For the norm in a Polish space we use always the notation

$|\cdot|$ , which, when there is no room for confusion and we are in  $\mathbb{R}^d$ , will denote the Euclidean norm. The rest of the norms will be specified in each case, i.e.  $|\mathbf{x}|_\infty = \sup_i |x_i|$  and  $|\mathbf{x}|_p = \left(\sum_{i=1}^d |x_i|^p\right)^{\frac{1}{p}}$ . For a set  $A$ , we denote its border by  $\partial A$ . For a measure  $\mu \in \mathcal{P}(\mathcal{Y})$  and a function  $f : \mathcal{Y} \rightarrow \mathbb{R}^d$ , we denote  $\|f\|_{L^p(\mu)} = \left(\int \|f\|^p d\mu\right)^{\frac{1}{p}}$ . The functions where this last value is finite are said to belong to  $L^p(\mu)$ .

## 2. Central Limit Theorems for semi-discrete distributions

### 2.1. Semi-discrete 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,  $\mathbb{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subset \mathcal{X}$  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

$$P := \sum_{k=1}^N p_k \delta_{\mathbf{x}_k}, \text{ where } p_i > 0, \text{ for all } i = 1, \dots, N, \text{ and } \sum_{k=1}^N p_k = 1. \quad (5)$$

In consequence  $P$  is characterized by the vector  $\mathbf{p} = (p_1, \dots, p_N) \in \mathbb{R}^N$ .

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

$$g_c(P, Q, \mathbf{z}) = \sum_{i=1}^N z_i p_i + \int \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\} dQ(\mathbf{y}), \quad \mathbf{z} \in \mathbb{R}^N \quad (6)$$

**Lemma 2.1.** *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*

$$\mathcal{T}_c(P, Q) = \sup_{\mathbf{z} \in \mathbb{R}^N, |\mathbf{z}| \leq K^*} g_c(P, Q, \mathbf{z}). \quad (7)$$

for  $K^* = \frac{1}{\inf_i p_i} \left( \sup_{i=1, \dots, N} \int c(\mathbf{y}, \mathbf{x}_i) dQ(\mathbf{y}) \right)$ . Moreover we can assume that  $z_1 = 0$ .

**Remark 2.2.** *Consider the dual expression for  $\mathcal{T}_c(P, Q)$  and let  $\varphi$  denote an optimal transport potential from  $P$  to  $Q$  for the cost  $c$ , then*

$$\mathcal{T}_c(P, Q) = g_c(P, Q, (\varphi(\mathbf{x}_1), \dots, \varphi(\mathbf{x}_N))).$$

Hence the optimal transport potentials and optimal values for (7) are linked through the expression  $\mathbf{z} = (\varphi(\mathbf{x}_1), \dots, \varphi(\mathbf{x}_N))$ .

Note that  $g_c(P, Q, \cdot)$  is a continuous function, a fact that follows from the next lemma. Therefore, the sup in (7) is attained and the the class of optimal values

$$\text{Opt}_c(P, Q) := \{\mathbf{z} \in \mathbb{R}^N : \mathcal{T}_c(P, Q) = g_c(P, Q, \mathbf{z})\} \quad (8)$$

and its restriction

$$\text{Opt}_c^0(P, Q) := \{\mathbf{z} \in \mathbb{R}^N : \mathcal{T}_c(P, Q) = g_c(P, Q, \mathbf{z}), z_1 = 0\}. \quad (9)$$

are both non-empty.

**Lemma 2.3.** *If  $f(\mathbf{y}) = \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\}$  and  $g(\mathbf{y}) = \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - s_i\}$ , then*

$$|f(\mathbf{y}) - g(\mathbf{y})| \leq \sup_{i=1, \dots, N} \{|z_i - s_i|\} \leq |\mathbf{z} - \mathbf{s}|. \quad (10)$$

## 2.2. Main results : Central Limit Theorems for semi-discrete optimal transport cost

Our aim is to study the empirical semi-discrete optimal transport cost. Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  and  $\mathbf{Y}_1, \dots, \mathbf{Y}_m$  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  $k = 1, \dots, n$ , 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  $p_1^n, \dots, p_N^n$  are real random variables such that  $p_i^n \geq 0$ , for all  $i = 1, \dots, N$ , 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

$$\{\sqrt{n}(\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q))\}_{n \in \mathbb{N}}, \quad \{\sqrt{n}(\mathcal{T}_c(P, Q_m) - \mathcal{T}_c(P, Q))\}_{m \in \mathbb{N}},$$

and the two sample case

$$\left\{ \sqrt{\frac{nm}{n+m}} (\mathcal{T}_c(P_n, Q_m) - \mathcal{T}_c(P, Q)) \right\}_{m, n \in \mathbb{N}},$$

under the assumption  $\frac{m}{n+m} \rightarrow \lambda \in (0, 1)$ .

To state the asymptotic behaviour we introduce first a centered Gaussian vector,  $(\mathbf{X}_1, \dots, \mathbf{X}_N)$  with covariance matrix  $\Sigma(\mathbf{p})$  with entries

$$\Sigma(\mathbf{p})_{i,j} = -p_i p_j, \quad i \neq j \quad \text{and} \quad \Sigma(\mathbf{p})_{i,i} = p_i(1 - p_i). \quad (11)$$

We also define a centered Gaussian process  $\mathbb{G}_Q^c$  in  $\mathbb{R}^N$  with covariance function

$$\begin{aligned} \Xi_Q^c(\mathbf{z}, \mathbf{s}) := & \int \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\} \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - s_i\} dQ(\mathbf{y}) \\ & - \int \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\} dQ(\mathbf{y}) \int \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - s_i\} dQ(\mathbf{y}). \end{aligned} \quad (12)$$

We can now state our main theorem.

**Theorem 2.4.** Let  $P \in \mathcal{P}(\mathbb{X})$ ,  $Q \in \mathcal{P}(\mathcal{Y})$ ,  $c$  be non-negative and

$$\int c(\mathbf{x}_i, \mathbf{y}) dQ(\mathbf{y}) < \infty, \text{ for all } i = 1, \dots, m, \quad (13)$$

then the following limits hold.

- (One sample case for empirical discrete distribution  $P_n$ ) If  $n \rightarrow \infty$ ,

$$\sqrt{n} (\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q)) \xrightarrow{w} \sup_{\mathbf{z} \in \text{Opt}_c^0(P, Q)} \sum_{i=1}^N z_i \mathbb{U}_i.$$

Suppose that

$$\int c(\mathbf{x}_i, \mathbf{y})^2 dQ(\mathbf{y}) < \infty, \text{ for all } i = 1, \dots, N. \quad (14)$$

- (One sample case for empirical distribution  $Q_m$ ) If  $m \rightarrow \infty$ ,

$$\sqrt{m} (\mathcal{T}_c(P, Q_m) - \mathcal{T}_c(P, Q)) \xrightarrow{w} \sup_{\mathbf{z} \in \text{Opt}_c^0(P, Q)} \mathbb{G}_Q^c(\mathbf{z}).$$

- (Two sample case) If  $n, m \rightarrow \infty$ , with  $\frac{m}{n+m} \rightarrow \lambda \in (0, 1)$ ,

$$\sqrt{\frac{nm}{n+m}} (\mathcal{T}_c(P_n, Q_m) - \mathcal{T}_c(P, Q)) \xrightarrow{w} \sup_{\mathbf{z} \in \text{Opt}_c^0(P, Q)} \left( \sqrt{\lambda} \sum_{i=1}^N z_i \mathbb{U}_i + (\sqrt{1-\lambda}) \mathbb{G}_Q^c(\mathbf{z}) \right).$$

Here  $(\mathbb{U}_1, \dots, \mathbb{U}_N) \sim \mathcal{N}(\mathbf{0}, \Sigma(\mathbf{p}))$ , with  $\Sigma(\mathbf{p})$  as in (11), and  $\mathbb{G}_Q^c$  is a centered Gaussian process with covariance function  $\Xi^c(Q)$  defined in (12). 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 2.4 is  $d^p$ , for all  $p \geq 1$ . Then applying Theorem 2.4 to the empirical estimators 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 2.5.** 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  $p \geq 1$ , the following limits hold.

- (One sample case for  $P$ ) If  $n \rightarrow \infty$ ,

$$\sqrt{n} (\mathcal{W}_p(P_n, Q) - \mathcal{W}_p(P, Q)) \xrightarrow{w} \frac{1}{p (\mathcal{W}_p(P, Q))^{p-1}} \sup_{\mathbf{z} \in \text{Opt}_{d^p}^0(P, Q)} \sum_{i=1}^N z_i \mathbb{U}_i.$$

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  $Q$ ) If  $m \rightarrow \infty$ ,

$$\sqrt{m} (\mathcal{W}_p(P, Q_m) - \mathcal{W}_p(P, Q)) \xrightarrow{w} \frac{1}{p (\mathcal{W}_p(P, Q))^{p-1}} \sup_{\mathbf{z} \in \text{Opt}_c^0(P, Q)} \mathbb{G}_Q^{d^p}(\mathbf{z}).$$

- **(Two sample case)** If  $n, m \rightarrow \infty$ , with  $\frac{m}{n+m} \rightarrow \lambda \in (0, 1)$ ,

$$\sqrt{\frac{nm}{n+m}} (\mathcal{W}_p(P_n, Q_m) - \mathcal{W}_p(P, Q)) \xrightarrow{w} \frac{\sup_{\mathbf{z} \in \text{Opt}_c^0(P, Q)} \left( \sqrt{\lambda} \sum_{i=1}^N z_i \mathbb{U}_i + (\sqrt{1-\lambda}) \mathbb{G}_Q^{d_P}(\mathbf{z}) \right)}{p(\mathcal{W}_p(P, Q))^{p-1}}.$$

Here  $(\mathbb{U}_1, \dots, \mathbb{U}_N)$  follows a Gaussian distribution  $\mathcal{N}(\mathbf{0}, \Sigma(\mathbf{p}))$ , with  $\Sigma(\mathbf{p})$  defined in (11), and  $\mathbb{G}_Q^{d_P}$  is a centered Gaussian process with covariance function  $\Xi^{d_P}(Q)$ , defined in (12). Moreover,  $\mathbb{G}_Q^{d_P}$  and  $(\mathbb{U}_1, \dots, \mathbb{U}_N)$  are independent.

The proof of Theorem 2.4 is provided in the supplementary material, and is restricted to the case of two samples. 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.

### 2.3. An upper-bound on the expectation for the Wasserstein distance

Theorem 2.4 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 2.6.** 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

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

where

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

and  $\mathcal{W}_1$  is the 1-Wasserstein distance for the Euclidean distance. Moreover, if  $\text{diam}(\mathcal{Y}) < \infty$ , then

$$E |\mathcal{W}_p^P(P, Q_m) - \mathcal{W}_p^P(P, Q)| \leq \frac{4\sqrt{2N}}{\sqrt{m}} K(4 \text{diam}(\mathbb{X}) \text{diam}(\mathcal{Y})^{p-1}, Q).$$

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 2.6 and the triangle inequality give the following upper bound

$$E |\mathcal{W}_1(P^N, Q_m) - \mathcal{W}_1(P, Q)| \leq \frac{8\sqrt{2N}}{\sqrt{m}} K(\Omega, Q) + |\mathcal{W}_1(P^N, Q) - \mathcal{W}_1(P, Q)|.$$

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.

We point out that previous upper bound is optimal, up to logarithmic factors, as specified in (Bing, Bunea and Niles-Weed, 2022, Theorem 5), where is proved the faster convergence rate at which a function of the sample can approximate the population Wasserstein distance between discrete measures. The lower bound they provide is

$$\text{diam}(\mathbb{X}) \max \left( \frac{\min_{\mathbf{x} \neq \mathbf{x}' \in \mathbb{X}} |\mathbf{x} - \mathbf{x}'|^2}{\max_{\mathbf{x}, \mathbf{x}' \in \mathbb{X}} |\mathbf{x} - \mathbf{x}'|^2} \sqrt{\frac{N}{m \log(N)^2}}, \sqrt{\frac{1}{m}} \right).$$

It follows directly from Theorem 2.6 that the empirical Wasserstein distance is an estimator that achieves an optimal convergence rate up to logarithmic factors, even in cases where one of the underlying probabilities is continuous.

### 3. Asymptotic Gaussian distribution optimal transport cost

Theorem 2.4 is valid for generic Polish spaces. When  $X, \mathcal{Y}$  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  $h : \mathbb{R}^d \rightarrow [0, \infty)$  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

$$K(r, \theta, \mathbf{z}, \mathbf{p}) := \{ \mathbf{x} \in \mathbb{R}^d : |\mathbf{x} - \mathbf{p}| |\mathbf{z}| \cos(\theta/2) \leq \langle \mathbf{z}, \mathbf{x} - \mathbf{p} \rangle \leq r |\mathbf{z}| \}, \quad (15)$$

with vertex at  $\mathbf{p}$  on which  $h(\mathbf{x})$  attains its maximum at  $\mathbf{p}$ .

- (A3):  $\lim_{|\mathbf{x}| \rightarrow 0} \frac{h(\mathbf{x})}{|\mathbf{x}|} = \infty$ .

Under such assumptions, Gangbo and McCann (1996) shows the existence of an optimal transport map  $T$  solving

$$\mathcal{T}_c(P, Q) := \inf_T \int c(\mathbf{y}, T(\mathbf{y})) dQ(\mathbf{y}), \quad \text{and } T_{\#}Q = P, \quad (16)$$

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 (16) is an *optimal transport map* from  $P$  to  $Q$ . Moreover, it is defined as the unique Borel function satisfying

$$T(\mathbf{x}) = \mathbf{x} - \nabla h^*(\nabla \varphi(\mathbf{x})), \quad \text{where } \varphi \text{ solves (3)}. \quad (17)$$

Here  $h^*$  denotes the convex conjugate of  $h$ , see Rockafellar (1970). Such uniqueness enabled del Barrio, Sanz and Loubes (2021) to deduce the uniqueness, under additive constants, of the solutions of (3) in  $\varphi$ . They assumed (A1)-(A3) to show that if two solutions of (3) have the same gradient almost everywhere for  $\ell_d$  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,  $\ell_d(\partial \text{supp}(Q)) = 0$ , the uniqueness, up to additive constants, of the solutions of (3) holds. The proof of the main theorem in this section is a direct consequence of Lemma 3.1, which proves that there exists a unique, up to an additive constant,  $\mathbf{z} \in \text{Opt}(P, Q)$ . We use within this section the notation  $\mathbf{1} := (1, \dots, 1)$ .

**Lemma 3.1.** *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*

$$\int c(\mathbf{x}_i, \mathbf{y}) dQ(\mathbf{y}) < \infty, \text{ for all } i = 1, \dots, N.$$

Then the set  $\text{Opt}_c^0(P, Q)$  is a singleton.

The following theorem states, under the previous assumptions, that the limit distribution described in Theorem 2.4 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  $\{\mathbf{z}\} = \text{Opt}_c^0(P, Q)$ . Note that  $\sum_{i=1}^N z_i \mathbb{U}_i$  is Gaussian and centered, with variance

$$\sigma^2(P, \mathbf{z}) = \text{Var}\left(\sum_{i=1}^N z_i \mathbb{U}_i\right) \text{ and } (\mathbb{U}_1, \dots, \mathbb{U}_N) \sim \mathcal{N}(\mathbf{0}, \Sigma(\mathbf{p})), \quad (18)$$

where  $\Sigma(\mathbf{p})$  is defined in (11). On the other side  $\mathbb{G}_Q^c(\mathbf{z})$  follows the distribution  $\mathcal{N}(0, \sigma_c^2(Q, \mathbf{z}))$ , where

$$\sigma_c^2(Q, \mathbf{z}) = \int \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\}^2 dQ(\mathbf{y}) - \left(\int \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\} dQ(\mathbf{y})\right)^2. \quad (19)$$

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.** *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 the following limits hold.*

- **(One sample case for  $P$ )** If  $n \rightarrow \infty$ ,

$$\sqrt{n} (\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q)) \xrightarrow{w} X \sim \mathcal{N}(0, \sigma^2(P, \mathbf{z})).$$

If, additionally,  $\int c(\mathbf{x}_i, \mathbf{y})^2 dQ(\mathbf{y}) < \infty$ , for all  $i = 1, \dots, N$ , then the following limits hold.

- **(One sample case for  $Q$ )** If  $n \rightarrow \infty$ ,

$$\sqrt{m} (\mathcal{T}_c(P, Q_m) - \mathcal{T}_c(P, Q)) \xrightarrow{w} Y \sim \mathcal{N}(0, \sigma_c^2(Q, \mathbf{z})).$$

- **(Two sample case)** If  $n, m \rightarrow \infty$ , with  $\frac{m}{n+m} \rightarrow \lambda \in (0, 1)$ ,

$$\sqrt{\frac{nm}{n+m}} (\mathcal{T}_c(P_n, Q_m) - \mathcal{T}_c(P, Q)) \xrightarrow{w} \sqrt{\lambda} X + (\sqrt{1-\lambda}) Y.$$

Here,  $\sigma^2(P, \mathbf{z})$  and  $\sigma_c^2(Q, \mathbf{z})$  are defined in (18) and (19) 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  $p > 1$ , satisfy (A1)-(A3), then the following result follows immediately from Theorem 3.2 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 3.3.** *Let  $P \in \mathcal{P}(\mathbb{X})$  be as in (5) 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  $p > 1$ , the following limits hold.*

- **(One sample case for  $P$ )** If  $n \rightarrow \infty$ ,

$$\sqrt{n} (\mathcal{T}_p(P_n, Q) - \mathcal{T}_p(P, Q)) \xrightarrow{w} \mathcal{N}(0, \sigma^2(P, \mathbf{z})),$$

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, as  $m \rightarrow \infty$ ,

- **(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, moreover,  $n, m \rightarrow \infty$ , with  $\frac{m}{n+m} \rightarrow \lambda \in (0, 1)$ ,

$$\sqrt{\frac{nm}{n+m}} (\mathcal{T}_p(P_n, Q_m) - \mathcal{T}_p(P, Q)) \xrightarrow{w} \mathcal{N}(0, \lambda \sigma^2(P, \mathbf{z}) + (1 - \lambda) \sigma_p^2(Q, \mathbf{z})),$$

and

$$\sqrt{\frac{nm}{n+m}} (\mathcal{W}_p(P_n, Q_m) - \mathcal{W}_p(P, Q)) \xrightarrow{w} \mathcal{N}\left(0, \frac{\lambda \sigma^2(P, \mathbf{z}) + (1 - \lambda) \sigma_p^2(Q, \mathbf{z})}{p^2 \mathcal{W}_p(P, Q)^{2p-2}}\right).$$

Here,  $\sigma^2(P, \mathbf{z})$  and  $\sigma_p^2(Q, \mathbf{z})$  are defined in (18) and (19) for  $\mathbf{z} \in \text{Opt}_{|\cdot|^p}(P, Q)$  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 3.3 is always well defined. We note also that Corollary 3.3 is a particular case of Corollary 2.5 in the cases where the optimal transport potential is unique –the hypotheses of Theorem 3.2 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, Sanz and Loubes \(2021\)](#) or [Gangbo and McCann \(1996\)](#).

**Remark 3.4.** 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  $\text{Opt}_c^0(P_n, Q_m)$ . [del Barrio, Sanz and Loubes \(2021\)](#) proves that

$$\sigma_{n,m}^2(P, Q) = \frac{nm}{n+m} \left( \int (z_i^{n,m})^2 P_n(\mathbf{x}_i) - \left( \int z_i^{n,m} P_n(\mathbf{x}_i) \right)^2 + \int \left( \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i^{n,m}\} \right)^2 dQ_m(\mathbf{y}) - \left( \int \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i^{n,m}\} dQ_m(\mathbf{y}) \right)^2 \right)$$

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.

## 4. Central Limit theorems for the potentials and Laguerre cells

### 4.1. 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 of the empirical version of the dual formulation of the Monge-Kantorovich problem (3). In the semi-discrete case the potentials are pairs formed by  $\mathbf{z} = (z_1, \dots, z_N) \in \text{Opt}_c(P, Q)$  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  $(\psi, \varphi)$  solves (3) then  $(\psi + C, \varphi - C)$  also solves (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  $\mathbf{1} = (1, \dots, 1)$

$$\begin{aligned} \mathcal{M}_{\mathbf{p}} : \langle \mathbf{1} \rangle^\perp &\longrightarrow \mathbb{R} \\ \mathbf{z} &\mapsto g_c(P, Q, \mathbf{z}), \end{aligned}$$

where  $g_c(P, Q, \mathbf{z})$  is defined as in (6).

In this section we will use some framework developed in [Kitagawa, Mérigot and Thibert \(2019\)](#). 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  $C(\mathcal{Y})$ ,  $C^1(\mathcal{Y})$  and  $C^{1,1}(\mathcal{Y})$  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, Mérigot and Thibert \(2019\)](#), we assume that the cost satisfies the following assumptions

$$c(\mathbf{x}_i, \cdot) \in C^{1,1}(\mathcal{Y}), \text{ for all } i = 1, \dots, N, \quad (\text{Reg})$$

$$D_{\mathbf{y}}c(\mathbf{x}_i, \mathbf{y}) : \mathcal{Y} \rightarrow T_{\mathbf{y}}^*(\mathcal{Y}) \text{ is injective as a function of } \mathbf{y}, \text{ for all } i = 1, \dots, N, \quad (\text{Twist})$$

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  $C^{1,1}$  diffeomorphism  $\exp_i^c :$

$\mathcal{Y}_i \rightarrow \mathcal{Y}$  such that the functions

$$\mathcal{Y}_i \ni \mathbf{p} \mapsto f_{i,j}(\mathbf{p}) := c(\mathbf{x}_i, \exp_i^c \mathbf{p}) - c(\mathbf{x}_j, \exp_j^c \mathbf{p}) \text{ are quasi-convex for all } j = 1, \dots, N. \quad (\text{QC})$$

Here quasi-convex, according to [Kitagawa, Mériqot and Thibert \(2019\)](#), 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 (5) and suppose that

$$Q \in \mathcal{P}(\mathcal{Y}) \text{ satisfies } Q \ll \mathcal{V}_g^d \text{ with density } q \in C(\mathcal{Y}). \quad (\text{Cont})$$

The last required assumption in [Kitagawa, Mériqot and Thibert \(2019\)](#) 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  $f \in C^1(\mathcal{Y})$  we have that for  $Y \sim Q$

$$E(|f(Y) - E(f(Y))|) \leq C_{PW} E(|\nabla f(Y)|). \quad (\text{PW})$$

In order to clarify the feasibility of such assumptions, we will provide some insights on them at the end of the section. [Kitagawa, Mériqot and Thibert \(2019\)](#) proved the following assertions.

1. Under assumptions [\(Reg\)](#) and [\(Twist\)](#) the function  $\mathcal{M}_{\mathbf{p}}(\mathbf{z})$  is concave and differentiable with derivative

$$\nabla_{\mathbf{z}} \mathcal{M}_{\mathbf{p}}(\mathbf{z}) = (-Q(\text{Lag}_1(\mathbf{z}) + p_1, \dots, -Q(\text{Lag}_N(\mathbf{z}) + p_N)), \quad (20)$$

where

$$\text{Lag}_k(\mathbf{z}) := \{\mathbf{y} \in \mathbb{R}^d : c(\mathbf{x}_k, \mathbf{y}) - z_k < c(\mathbf{x}_i, \mathbf{y}) - z_i, \text{ for all } i \neq k\}. \quad (21)$$

2. Under assumptions [\(Reg\)](#), [\(Twist\)](#) and [\(QC\)](#), the function  $\mathcal{M}_{\mathbf{p}}$  is twice continuously differentiable with Hessian matrix  $D_{\mathbf{z}}^2 \mathcal{M}_{\mathbf{p}}(\mathbf{z}) = \left( \frac{\partial^2 \mathcal{M}_{\mathbf{p}}}{\partial z_i \partial z_j}(\mathbf{z}) \right)_{i,j=1,\dots,N}$  and partial derivatives

$$\begin{aligned} \frac{\partial^2}{\partial z_i \partial z_j} \mathcal{M}_{\mathbf{p}}(\mathbf{z}) &= \int_{\text{Lag}_k(\mathbf{z}) \cap \text{Lag}_k(\mathbf{z})} \frac{q(\mathbf{y})}{|\nabla_{\mathbf{y}} c(\mathbf{x}_i, \mathbf{y}) - \nabla_{\mathbf{y}} c(\mathbf{x}_j, \mathbf{y})|} d\mathcal{V}_g^{d-1}(\mathbf{y}), \text{ if } i \neq j, \\ \frac{\partial^2}{\partial z_i^2} \mathcal{M}_{\mathbf{p}}(\mathbf{z}) &= - \sum_{j \neq i} \frac{\partial^2}{\partial z_i \partial z_j} \mathcal{M}_{\mathbf{p}}(\mathbf{z}). \end{aligned} \quad (22)$$

3. Under assumptions [\(Reg\)](#), [\(Twist\)](#) and [\(QC\)](#), and if  $Q$  satisfies [\(PW\)](#), there exists a positive constant  $C$  such that

$$\langle D_{\mathbf{z}}^2 \mathcal{M}_{\mathbf{p}}(\mathbf{z}) \mathbf{v}, \mathbf{v} \rangle \leq -C \epsilon^3 |\mathbf{v}|^2, \text{ for all } \mathbf{z} \in \mathcal{K}^\epsilon \text{ and } \mathbf{v} \in \langle \mathbf{1} \rangle^\perp, \quad (23)$$

where

$$\mathcal{K}^\epsilon := \{\mathbf{z} \in \mathbb{R}^d : Q(\text{Lag}_i(\mathbf{z})) > \epsilon, \text{ for all } i = 1, \dots, N\}.$$

Under this assumptions we can state the main result of the section: a CLT for the OT potentials.

**Theorem 4.1.** *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  $P$ ) If  $n \rightarrow \infty$ ,

$$\sqrt{n}(\hat{\mathbf{z}}^n - \mathbf{z}^*) \xrightarrow{w} -(D^2 \mathcal{M}_{\mathbf{p}}(\mathbf{z}^*))^{-1}((\mathbb{U}_1, \dots, \mathbb{U}_N)).$$

- (One sample case for  $Q$ ) If  $m \rightarrow \infty$ ,

$$\sqrt{m}(\hat{\mathbf{z}}^m - \mathbf{z}^*) \xrightarrow{w} -(D^2 \mathcal{M}_{\mathbf{p}}(\mathbf{z}^*))^{-1}((\mathbb{U}_1, \dots, \mathbb{U}_N)).$$

- (Two sample case) If  $n, m \rightarrow \infty$ , with  $\frac{m}{n+m} \rightarrow \lambda \in (0, 1)$ ,

$$\sqrt{\frac{nm}{n+m}}(\hat{\mathbf{z}}^{n,m} - \mathbf{z}^*) \xrightarrow{w} -(D^2 \mathcal{M}_{\mathbf{p}}(\mathbf{z}^*))^{-1}((\mathbb{U}_1, \dots, \mathbb{U}_N)).$$

Here  $\mathbf{z}^* \in \langle \mathbf{1} \rangle^\perp \cap \text{Opt}_c(P, Q)$ ,  $\hat{\mathbf{z}}^n \in \langle \mathbf{1} \rangle^\perp \cap \text{Opt}_c(P_n, Q)$ ,  $\hat{\mathbf{z}}^m \in \langle \mathbf{1} \rangle^\perp \cap \text{Opt}_c(P, Q_m)$ ,  $\hat{\mathbf{z}}^{n,m} \in \langle \mathbf{1} \rangle^\perp \cap \text{Opt}_c(P_n, Q_m)$  and  $(\mathbb{U}_1, \dots, \mathbb{U}_N) \sim \mathcal{N}(\mathbf{0}, \Sigma(\mathbf{p}))$ , for  $\Sigma(\mathbf{p})$  defined in (11).

For  $\mathbf{z}^*$  defined as in Theorem 4.1, set

$$\varphi(\mathbf{y}) := \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i^*\} \quad (24)$$

and note that it is an optimal transport map from  $Q$  to  $P$ . Set also the value  $i(\mathbf{y}) \in \{1, \dots, N\}$ , where the infimum of (24) is attained. As before, we can define their empirical counterparts

$$\begin{aligned} \varphi_{n,m}(\mathbf{y}) := \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - \hat{z}_i^{n,m}\}, \quad \varphi_n(\mathbf{y}) := \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - \hat{z}_i^n\}, \\ \text{and } \varphi_m(\mathbf{y}) := \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - \hat{z}_i^m\}, \end{aligned} \quad (25)$$

which are the other empirical OT potentials, and  $i_{n,m}(\mathbf{y})$  denotes the index where the infimum of (25) is attained for the two sample case. Then we have

$$\sqrt{\frac{nm}{n+m}}(\hat{z}_{i_{n,m}(\mathbf{y})}^{n,m} - z_{i_{n,m}(\mathbf{y})}^*) \leq \sqrt{\frac{nm}{n+m}}(\varphi_{n,m}(\mathbf{y}) - \varphi(\mathbf{y})) \leq \sqrt{\frac{nm}{n+m}}(\hat{z}_{i(\mathbf{y})}^{n,m} - z_{i(\mathbf{y})}^*). \quad (26)$$

We can take supremum over  $\mathbf{y}$  in both sides of (26) and obtain

$$\sqrt{\frac{nm}{n+m}} \sup_{i=1, \dots, N} (\hat{z}_i^{n,m} - z_i^*) = \sqrt{\frac{nm}{n+m}} \sup_{\mathbf{y} \in \mathcal{Y}} (\varphi_{n,m}(\mathbf{y}) - \varphi(\mathbf{y})).$$

By symmetry we have that

$$\sqrt{\frac{nm}{n+m}} |\hat{z}_i^{n,m} - z_i^*| = \sqrt{\frac{nm}{n+m}} \sup_{\mathbf{y} \in \mathcal{Y}} |\varphi_{n,m}(\mathbf{y}) - \varphi(\mathbf{y})|,$$

which implies the following corollary.

**Corollary 4.2.** Under the hypotheses and notation of Theorem 4.1, for  $\varphi$ ,  $\varphi_n$ ,  $\varphi_m$  and  $\varphi_{n,m}$  defined in (24) and (25), we have the following limits.

- (One sample case for  $P$ ) If  $n \rightarrow \infty$

$$\sqrt{n} \sup_{\mathbf{y} \in \mathcal{Y}} |\varphi_n(\mathbf{y}) - \varphi(\mathbf{y})| \xrightarrow{w} |(D^2 \mathcal{M}_{\mathbf{p}}(\mathbf{z}^*))^{-1}((\mathbb{U}_1, \dots, \mathbb{U}_N))|_\infty.$$

- *(One sample case for  $Q$ )*  $m \rightarrow \infty$

$$\sqrt{m} \sup_{\mathbf{y} \in \mathcal{Y}} |\varphi_m(\mathbf{y}) - \varphi(\mathbf{y})| \xrightarrow{w} |(D^2 \mathcal{M}_{\mathbf{p}}(\mathbf{z}^*))^{-1}((\mathbb{U}_1, \dots, \mathbb{U}_N))|_{\infty}.$$

- *(Two sample case)* If  $n, m \rightarrow \infty$ , with  $\frac{m}{n+m} \rightarrow \lambda \in (0, 1)$ ,

$$\sqrt{\frac{nm}{n+m}} \sup_{\mathbf{y} \in \mathcal{Y}} |\varphi_{n,m}(\mathbf{y}) - \varphi(\mathbf{y})| \xrightarrow{w} |(D^2 \mathcal{M}_{\mathbf{p}}(\mathbf{z}^*))^{-1}((\mathbb{U}_1, \dots, \mathbb{U}_N))|_{\infty}.$$

We will conclude by some comments on the assumptions made in this section.

1. Under the hypotheses of Theorem 4.1, the optimal potential is unique once we set its value at a given point. Then Corollary 4.2 provides a uniform confidence band for this optimal potential, namely,

$$\left[ \varphi_n(\mathbf{y}) \pm \frac{\Delta_{\alpha}}{\sqrt{n}} \right]_{\mathbf{y} \in \mathcal{Y}},$$

where  $\Delta_{\alpha}$  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  $C^{\infty}(\mathbb{R}^d)$  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.

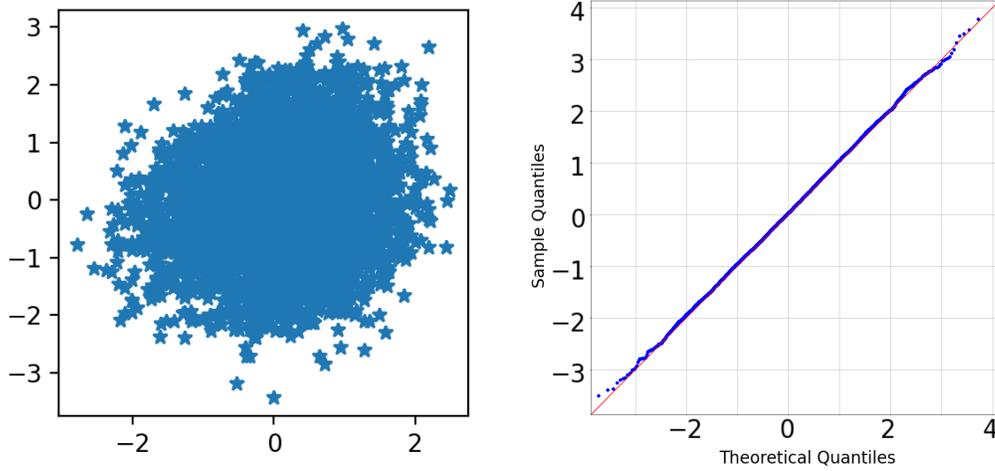
3. Assumption **(PW)** on the probability  $Q$  has been widely studied in the literature for its implications in PDEs, see [Acosta and Durán \(2004\)](#). They proved that **(PW)** holds for a uniform distribution on a convex set  $\mathcal{Y}$ . ([Rathmair, 2019](#), Lemma 1) claims that **(PW)** is equivalent to the bound of  $\inf_{t \in \mathbb{R}} E(|f(Y) - t|)$ , for every  $f \in C^1(\mathcal{Y})$ . Let  $Y \sim Q$  be such that there exists a  $C^1(\mathcal{Y})$  map  $T$  satisfying the relation  $T(U) = Y$ , where  $U$  follows a uniform distribution on a compact convex set  $A$ . Since  $f \circ T \in C^1(A)$ , 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, Mérigot and Thibert \(2019\)](#), more general probabilities can satisfy that assumption such as radial functions on  $\mathbb{R}^d$  with density

$$\frac{p(|\mathbf{x}|)}{|\mathbf{x}|^{d-1}}, \text{ for } |\mathbf{x}| \leq R, \text{ with } p = 0 \text{ in } [0, r] \text{ and concave in } [r, R].$$

Moreover the spherical uniform  $\mathbb{U}_d$ , used in [del Barrio, González Sanz and Hallin \(2022\)](#), [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 4.1 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-Ampère equation, see [del Barrio, González Sanz and Hallin \(2020\)](#).



**Figure 1.** Bootstrap approximation of  $N(\mathbf{0}, \Sigma(\mathbf{z}^*))$ . Here  $P$  is supported in three 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.

4. The limit distribution described in 4.1 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 procedure can be used to approximate the limit distribution. The approximation will be consistent as in [Fang and Santos \(2018\)](#). In Figure 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  $(0, 1)^6$ .

## 4.2. 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

$$\text{Lag}_i^R(\mathbf{z}) = \overline{\text{Lag}_i(\mathbf{z}) \cap R\mathbb{B}_d},$$

which is compact. As a consequence, 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

$$\mathbb{S}^{d-1} \ni \mathbf{v} \rightarrow h_A(\mathbf{v}) := \sup_{\mathbf{y} \in A} \langle \mathbf{v}, \mathbf{y} \rangle.$$

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 ([Bansil and Kitagawa, 2020](#), Remark 5.1)).

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  $d_p(A, B) := (\int |h_A - h_B|^p d\mathcal{H}^{d-1})^{\frac{1}{p}}$ , where  $\mathcal{H}^{d-1}$  is the Hausdorff measure in  $\mathbb{S}^{d-1}$ . The case  $p = \infty$  is  $d_\infty(A, B) = \sup_{\mathbf{v} \in \mathbb{S}^{d-1}} |h_A - h_B|$ , 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  $\text{Lag}_k^R(\mathbf{z}^*)$ , we set the notation

$$\text{Sol}(\mathbf{z}^*, \mathbf{v}, \mathbf{y}^0) := \arg \min_{t_j > 0} \left\{ \sum_{j \neq k} t_j \psi_j + R|\mathbf{v} - \sum_{j \neq k} t_j(\mathbf{x}_k - \mathbf{x}_j)| - \langle \mathbf{v} - \sum_{j \neq k} t_j(\mathbf{x}_k - \mathbf{x}_j), \mathbf{y}^0 \rangle \right\},$$

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  $\mathbf{z}^* \in \langle \mathbf{1} \rangle^\perp \cap \text{Opt}_{|\cdot|^2}(P, Q)$ , 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 4.3.** *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  $P$ ) If  $n \rightarrow \infty$

$$\sqrt{n} d_p(\text{Lag}_k^R(\hat{\mathbf{z}}^n), \text{Lag}_k^R(\mathbf{z}^*)) \xrightarrow{w} \left( \int \left| \inf_{\mathbf{t} \in \text{Sol}(\mathbf{z}^*, \mathbf{v}, \mathbf{y}^0)} \left\{ \sum_{j \neq k} t_j (\mathbb{M}_k - \mathbb{M}_j) \right\} \right|^p d\mathcal{H}^{d-1} \right)^{\frac{1}{p}}.$$

- (One sample case for  $Q$ ) If  $m \rightarrow \infty$ ,

$$\sqrt{m} d_p(\text{Lag}_k^R(\hat{\mathbf{z}}^m), \text{Lag}_k^R(\mathbf{z}^*)) \xrightarrow{w} \left( \int \left| \inf_{\mathbf{t} \in \text{Sol}(\mathbf{z}^*, \mathbf{v}, \mathbf{y}^0)} \left\{ \sum_{j \neq k} t_j (\mathbb{M}_k - \mathbb{M}_j) \right\} \right|^p d\mathcal{H}^{d-1} \right)^{\frac{1}{p}}.$$

- **(Two sample case)** If  $n, m \rightarrow \infty$ , with  $\frac{m}{n+m} \rightarrow \lambda \in (0, 1)$ ,

$$\sqrt{\frac{nm}{n+m}} d_P(\text{Lag}_k^R(\hat{\mathbf{z}}^{n,m}), \text{Lag}_k^R(\mathbf{z}^*)) \xrightarrow{w} \left( \int \left| \inf_{\mathbf{t} \in \text{Sol}(\mathbf{z}^*, \mathbf{v}, \mathbf{y}^0)} \left\{ \sum_{j \neq k} t_j (\mathbb{M}_k - \mathbb{M}_j) \right\} \right|^p d\mathcal{H}^{d-1} \right)^{\frac{1}{p}}.$$

Here  $\mathbf{z}^* \in \langle \mathbf{1} \rangle^\perp \cap \text{Opt}_{|\cdot|^2}(P, Q)$ , for the quadratic cost  $|\cdot|^2$  and

$$(\mathbb{M}_1, \dots, \mathbb{M}_N) = \left( D^2 \mathcal{M}_{\mathbf{p}}(\mathbf{z}^*) \right)^{-1} (\mathbb{U}_1, \dots, \mathbb{U}_N),$$

where  $(\mathbb{U}_1, \dots, \mathbb{U}_N) \sim \mathcal{N}(\mathbf{0}, \Sigma(\mathbf{p}))$ , for  $\Sigma(\mathbf{p})$ , is defined in (11).

The proof is based on the following description of the support functions of Laguerre's cells

$$h_{\text{Lag}_k^R(\mathbf{z})}(\mathbf{v}) = \min_{t_j > 0} \left\{ \sum_{j \neq k} t_j (|\mathbf{x}_k|^2 - |\mathbf{x}_j|^2 - z_k + z_j) + R|\mathbf{v} - \sum_{j \neq k} t_j (\mathbf{x}_k - \mathbf{x}_j)| \right\}, \quad (27)$$

which allows to apply standard arguments to handle Hadamard derivative of the infimum. However, the compactness of the solution-set of (27) is not guaranteed. Note 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 4.5 below implies that

$$h_{\text{Lag}_k^R(\hat{\mathbf{z}}^{n,m})}(\mathbf{v}) - h_{\text{Lag}_k^R(\mathbf{z}^*)}(\mathbf{v}) = h_{\text{Lag}_k^R(\hat{\mathbf{z}}^{n,m) + \{-\mathbf{y}^0\}}(\mathbf{v}) - h_{\text{Lag}_k^R(\mathbf{z}^*) + \{-\mathbf{y}^0\}}(\mathbf{v}).$$

In such a case

$$h_{\text{Lag}_k^R(\mathbf{z}^*) + \{-\mathbf{y}^0\}}(\mathbf{v}) = \min_{t_j > 0} \left\{ \sum_{j \neq k} t_j \psi_j(\mathbf{z}^*) + R|\mathbf{v} - \sum_{j \neq k} t_j (\mathbf{x}_k - \mathbf{x}_j)| - \langle \mathbf{v} - \sum_{j \neq k} t_j (\mathbf{x}_k - \mathbf{x}_j), \mathbf{y}^0 \rangle \right\}, \quad (28)$$

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  $\text{Sol}(\mathbf{z}^*, \mathbf{v}, \mathbf{y}^0)$  and gives, as an intermediate step, the point-wise limit of the support functions.

**Lemma 4.4.** *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)** If  $n \rightarrow \infty$ ,

$$\sqrt{n} (h_{\text{Lag}_k^R(\hat{\mathbf{z}}^n)}(\mathbf{v}) - h_{\text{Lag}_k^R(\mathbf{z}^*)}(\mathbf{v})) \xrightarrow{w} \inf_{\mathbf{t} \in \text{Sol}(\mathbf{z}^*, \mathbf{v}, \mathbf{y}^0)} \left\{ \sum_{j \neq k} t_j (\mathbb{M}_k - \mathbb{M}_j) \right\}.$$

- **(One sample case for Q)** If  $m \rightarrow \infty$ ,

$$\sqrt{m} (h_{\text{Lag}_k^R(\hat{\mathbf{z}}^m)}(\mathbf{v}) - h_{\text{Lag}_k^R(\mathbf{z}^*)}(\mathbf{v})) \xrightarrow{w} \inf_{\mathbf{t} \in \text{Sol}(\mathbf{z}^*, \mathbf{v}, \mathbf{y}^0)} \left\{ \sum_{j \neq k} t_j (\mathbb{M}_k - \mathbb{M}_j) \right\}.$$

- **(Two sample case)** If  $n, m \rightarrow \infty$ , with  $\frac{m}{n+m} \rightarrow \lambda \in (0, 1)$ ,

$$\sqrt{\frac{nm}{n+m}} (h_{\text{Lag}_k^R(\hat{\mathbf{z}}^{n,m})}(\mathbf{v}) - h_{\text{Lag}_k^R(\mathbf{z}^*)}(\mathbf{v})) \xrightarrow{w} \inf_{\mathbf{t} \in \text{Sol}(\mathbf{z}^*, \mathbf{v}, \mathbf{y}^0)} \left\{ \sum_{j \neq k} t_j (\mathbb{M}_k - \mathbb{M}_j) \right\}.$$

Here  $(\mathbb{M}_1, \dots, \mathbb{M}_N) = \left( D^2 \mathcal{M}_{\mathbf{p}}(\mathbf{z}^*) \right)^{-1} ((\mathbb{U}_1, \dots, \mathbb{U}_N))$ , where  $(\mathbb{U}_1, \dots, \mathbb{U}_N) \sim \mathcal{N}(\mathbf{0}, \Sigma(\mathbf{p}))$ , for  $\Sigma(\mathbf{p})$ , is defined in (11).

**Remark 4.5.** Let us recall some basic properties of the support function (see 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 4.5 is important for the proof of Theorem 4.3. 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:

$$d_p \left( \sqrt{\frac{nm}{n+m}} \text{Lag}_k^R(\hat{\mathbf{z}}^{n,m}), \sqrt{\frac{nm}{n+m}} \text{Lag}_k^R(\mathbf{z}^*) \right) \xrightarrow{w} \left( \int \left| \inf_{\mathbf{t} \in \text{Sol}(\mathbf{z}^*, \mathbf{v}, \mathbf{y}^0)} \left\{ \sum_{j \neq k} t_j (\mathbb{M}_k - \mathbb{M}_j) \right\} \right|^p d\mathcal{H}^{d-1} \right)^{\frac{1}{p}}. \quad (29)$$

To go from point-wise convergence to  $L^p$  convergence, the proof of Theorem 4.3 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, Sanz and Loubes (2021)*).

**Lemma 4.6.** Under the hypotheses and notation of Theorem 4.3, we have

$$t_n d_\infty(\text{Lag}_k^R(\mathbf{z}^* + t_n \mathbf{h}_n), \text{Lag}_k^R(\mathbf{z}^*)) \leq \frac{4R}{\delta_n} |\mathbf{h}_n|_\infty,$$

for any  $t_n \searrow 0$  and  $\mathbf{d}_n \rightarrow \mathbf{d}$ , with

$$\delta_n = \sup_{\mathbf{y} \in \text{Lag}_k^R(\mathbf{z}^* + t_n \mathbf{h}_n) \cap \text{Lag}_k^R(\mathbf{z}^*)} \min \left( d_\infty(\{\mathbf{y}\}, \partial \text{Lag}_k^R(\mathbf{z}^*)), d_\infty(\{\mathbf{y}\}, \partial \text{Lag}_k^R(\mathbf{z}^* + t_n \mathbf{h}_n)) \right).$$

Moreover, if  $\mathbf{u} + 2a\mathbb{B}_d \subset \text{Lag}_k^R(\mathbf{z}^*)$ , then  $\liminf_n \delta_n \geq a$ .

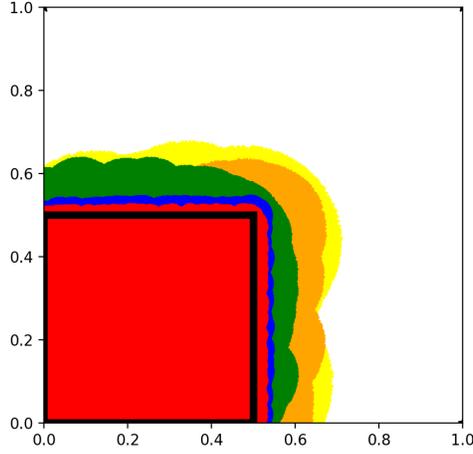
Lemma 4.6 gives confidence intervals for the Hausdorff distance between the cells –uncovered in Theorem 4.3. Note that Lemma (4.6) implies that

$$\begin{aligned} \mathbb{P} \left( \text{Lag}_k^R(\mathbf{z}^*) \subset \text{Lag}_k^R(\hat{\mathbf{z}}^m) + \frac{4R\Psi^{-1}(\alpha)}{a\sqrt{m}} \mathbb{B}_d \right) &\geq \mathbb{P} \left( d_\infty(\text{Lag}_k^R(\mathbf{z}^*), \text{Lag}_k^R(\hat{\mathbf{z}}^m)) \leq \frac{4R\Psi^{-1}(\alpha)}{a\sqrt{m}} \right) \\ &\geq \mathbb{P} \left( \frac{4R|(\mathbb{M}_1, \dots, \mathbb{M}_N)|_\infty}{\delta_m} \leq \frac{4R\Psi^{-1}(\alpha)}{a\sqrt{m}} \right) = \mathbb{P} \left( \frac{a\sqrt{m}|(\mathbb{M}_1, \dots, \mathbb{M}_N)|_\infty}{\delta_m} \leq \Psi^{-1}(\alpha) \right), \end{aligned}$$

where, taking inferior limits, we obtain the following result.

**Remark 4.7.** Let  $(\mathbb{M}_1, \dots, \mathbb{M}_N) = \left( D^2 \mathcal{M}_{\mathbf{p}}(\mathbf{z}^*) \right)^{-1} ((\mathbb{U}_1, \dots, \mathbb{U}_N))$  be as in Theorem 4.3 and  $\Psi^{-1}$  be the quantile function of  $|(\mathbb{M}_1, \dots, \mathbb{M}_N)|_\infty$ . Under the assumptions of Theorem 4.3, if  $\mathbf{u} + 2a\mathbb{B}_d \subset \text{Lag}_k^R(\mathbf{z}^*)$ , then

$$\liminf_m \mathbb{P} \left( \text{Lag}_k^R(\mathbf{z}^*) \subset \text{Lag}_k^R(\hat{\mathbf{z}}^m) + \frac{4R\Psi^{-1}(\alpha)}{a\sqrt{m}} \mathbb{B}_d \right) \geq \alpha.$$



**Figure 2.** Estimated upper confidence intervals for the set  $\text{Lag}_1^R(\mathbf{z}^*)$ , 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  $\text{Lag}_1^R(\mathbf{z}^*)$ .

We notice that we need to approximate the distribution  $\Psi$ . In the previous section we justified the consistency of the parametric bootstrap to approximate the distribution of  $(\mathbb{M}_1, \dots, \mathbb{M}_N) = (D^2 \mathcal{M}_p(\mathbf{z}^*))^{-1}(\mathbb{U}_1, \dots, \mathbb{U}_N)$ . More precisely, let  $\mathbf{Y}_1^B, \dots, \mathbf{Y}_m^B$  be a bootstrap sample of i.i.d. (conditionally given the sample  $\mathbf{Y}_1, \dots, \mathbf{Y}_m$ ) with common law  $Q_m$ . We 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 (7) 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 usefulness of Remark 4.7. 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

$$\text{Lag}_1(\mathbf{z}^*) = [0, \frac{1}{2}]^2, \text{Lag}_2(\mathbf{z}^*) = [\frac{1}{2}, 1] \times [0, \frac{1}{2}], \text{Lag}_3(\mathbf{z}^*) = [\frac{1}{2}, 1]^2, \text{Lag}_4(\mathbf{z}^*) = [0, \frac{1}{2}] \times [\frac{1}{2}, 1].$$

We analyze the behavior of the first cell. In Figure 2 we plot, for sample sizes  $m = 100, 500, 1000, 5000, 10000$ , the values of  $\text{Lag}_1^R(\hat{\mathbf{z}}^m) + \frac{4R(\Psi^{-1})^B(0.05)}{a\sqrt{m}}\mathbb{B}_d$ . Here  $(\Psi^{-1})^B$  denotes the quantile of the bootstrap approximation.

**Remark 4.8.** *The same technique can be applied to the Voronoi cells, i.e.*

$$\text{Vor}_k(\mathbf{z}) := \{\mathbf{y} \in \mathbb{R}^d : |\mathbf{c}_k - \mathbf{y}|^2 < |\mathbf{c}_i - \mathbf{y}|^2, \text{ for all } i \neq k\}, \quad k = 1, \dots, N,$$

where  $\{\mathbf{c}_i\}_{i=1}^N \in \arg \min_{\{\mathbf{a}_i\}_{i=1}^N \subset \mathbb{R}^d} \int \min_{0 \leq i \leq N} |\mathbf{a}_i - \mathbf{y}|^2 dQ(\mathbf{y})$  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 (27) can be made, which yields the differentiability.

**Remark 4.9.** *Note that Theorem 4.1 and the following upper bound*

$$\ell_d(\text{Lag}_i(\mathbf{z}) \setminus \text{Lag}_i(\mathbf{z}')) \leq C\|\mathbf{z} - \mathbf{z}'\|_\infty,$$

proved in ([Bansil and Kitagawa, 2020, Lemma 5.5](#)), suggest that  $\sqrt{\frac{nm}{n+m}}(\hat{T}^{n,m} - T)$  is stochastically bounded in norm  $L^1(\ell_d)$ , where  $\ell_d$  denotes the  $d$ -dimensional Lebesgue measure. Here  $\hat{T}^{n,m}$  denotes (when exists) the optimal transport map from  $Q_m$  and  $P_n$  and  $T$  is defined in (17). In order to provide an insight into this claim, we can bound

$$\begin{aligned} \int \|\hat{T}^{n,m} - T\| d\ell_d &\leq N \sup_{i=1, \dots, N} \|\mathbf{x}_i\| (\ell_d(\text{Lag}_i(\mathbf{z}^*) \setminus \text{Lag}_i(\hat{\mathbf{z}}^{n,m})) + \ell_d(\text{Lag}_i(\hat{\mathbf{z}}^{n,m}) \setminus \text{Lag}_i(\mathbf{z}^*))) \\ &\leq 2CN\|\hat{\mathbf{z}}^{n,m} - \mathbf{z}^*\|_\infty \sup_{i=1, \dots, N} \|\mathbf{x}_i\|. \end{aligned}$$

As a consequence, Theorem 4.1 states that  $\sqrt{\frac{nm}{n+m}} \int \|\hat{T}^{n,m} - T\| d\ell_d$  is stochastically bounded. This bound holds for the estimator  $\hat{T}^{n,m}$ . The recent pre-print [Pooladian, Divol and Niles-Weed \(2023\)](#) provides a bound on the  $L^2(Q)$  distance between an estimator based on the Sinkhorn relaxation and  $T$ . Research on the possible weak limits of  $\sqrt{\frac{nm}{n+m}}(\hat{T}^{n,m} - T)$  will be the scope of further research.

## 5. Applications to Hotelling's location model

As mentioned in the introduction and driven by the application described in [Galichon \(2016\)](#), the equilibrium of the Hotelling's location model becomes

$$\sup_{\mathbf{z}} \sum_{i=1}^N z_i p_i + \int \inf_{i=1, \dots, N} \{|\mathbf{x}_i - \mathbf{y}|^2 - z_i\} dQ(\mathbf{y}), \quad (30)$$

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 (30). The set of population that prefers to consume from the fountain  $i$ —called demand set—is  $\text{Lag}_i(\mathbf{z})$ .

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  $(\mathbf{Y}_1, \dots, \mathbf{Y}_m)$  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

$\{\text{Lag}_i(\mathbf{z}^m)\}_{i=1}^N$ . Note that we obtain also the empirical solution of (30), which does not play an important role in this problem. We assume that the population  $Q$  satisfies the assumptions of Theorem 4.3, which asserts that, denoting  $\Psi^{-1}$  the quantile function of  $|(\mathbb{M}_1, \dots, \mathbb{M}_N)|_\infty$ , then

$$\mathbb{P}\left(|\mathbf{z}^* - \mathbf{z}^m|_\infty > \frac{\Psi^{-1}(\alpha)}{\sqrt{m}}\right) \longrightarrow 1 - \alpha, \quad \alpha \in (0, 1) \quad (31)$$

On the other hand, in view of Lemma 4.6, we have

$$\liminf_m \mathbb{P}\left(\text{Lag}_k^R(\mathbf{z}^*) \subset \text{Lag}_k^R(\hat{\mathbf{z}}^m) + \frac{4R\Psi^{-1}(\alpha)}{\delta_m\sqrt{m}}\mathbb{B}_d\right) \geq \alpha, \quad \alpha \in (0, 1) \quad (32)$$

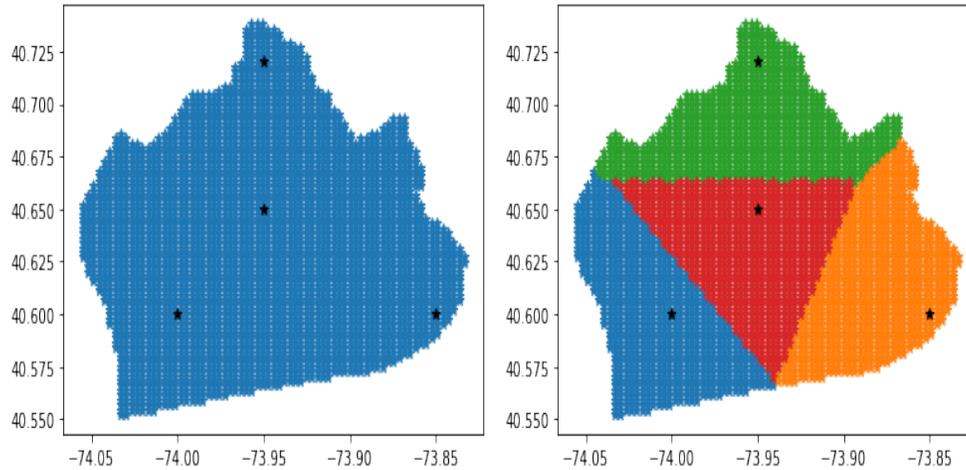
with

$$\delta_m = \sup_{\mathbf{y} \in \text{Lag}_k^R(\mathbf{z}^m) \cap \text{Lag}_k^R(\mathbf{z}^*)} \min\left(d_\infty(\{\mathbf{y}\}, \partial \text{Lag}_k^R(\mathbf{z}^*)), d_\infty(\{\mathbf{y}\}, \partial \text{Lag}_k^R(\mathbf{z}^m))\right).$$

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<sup>1</sup> 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  $(-74.0, 40.6)$ ,  $(-73.85, 40.6)$ ,  $(-73.95, 40.72)$  and  $(-73.95, 40.65)$  with same amount of stock. The data is displayed in Figure 3.

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  $z_i^* - z_i^m$ ,  $i = 1, 2, 3, 4$ , of the prices. We obtain the following results

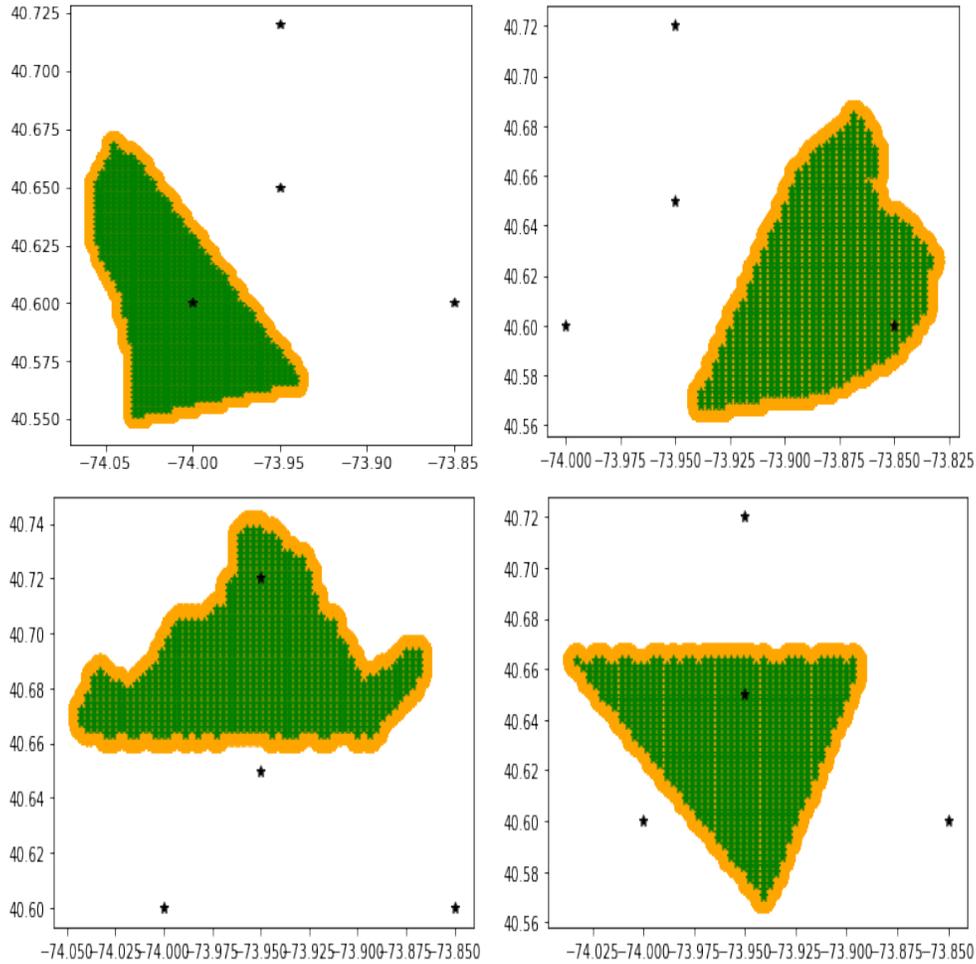
<sup>1</sup>According to the United States Census Bureau <http://www.census.gov/quickfacts/>



**Figure 3.** In the left original data of the sample ( $m = 3,129$ ) of Brooklyn’s population distribution. On the Right, the empirical Laguerre cells  $\text{Lag}_i^R(\mathbf{z}^*)$  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.

$\mathbf{z}^* = \mathbf{z}^m + [0, 5.217 \cdot 10^{-04}]^4$ ,  $z_1^* = z_1^m + 10^{-04}[-4.302, 4.242]$ ,  $z_2^* = z_2^m + 10^{-04}[-5.260, 5.486]$ ,  $z_3^* = z_3^m + 10^{-04}[-5.260, 5.486]$ ,  $z_4^* = z_4^m + 10^{-04}[-1.814, 1.416]$ . Here  $z_i^*$  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  $i = 4$ , at  $(-73.95, 40.65)$ . The same notation is shared by  $z_i^m$ .

Lemma 4.6 enables to obtain, in Figure 4, 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  $\delta_n$  is approximated by taking the maximum distance between points in the empirical cell.



**Figure 4.** Estimated upper confidence intervals for the set  $\text{Lag}_i^R(\mathbf{z}^*)$ . 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  $i = 4$ , (lower right) at  $(-73.95, 40.65)$ . 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.

## Acknowledgements

The authors would like to thank Luis-Alberto Rodríguez for showing us the paper [Cárcamo, Cuevas and Rodríguez \(2020\)](#), which is key for the proof of Theorem 2.4, and the anonymous reviewers, for encouraging us to extend the results of the paper and providing us useful references.

## Funding

The research of Eustasio del Barrio and Alberto González Sanz 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.

## References

- ACOSTA, G. and DURÁN, R. G. (2004). An Optimal Poincaré Inequality in L1 for Convex Domains. *Proc. Amer. Math. Soc.* **132** 195–202.
- BANSIL, M. and KITAGAWA, J. (2020). Quantitative Stability in the Geometry of Semi-discrete Optimal Transport. *International Mathematics Research Notices*.
- BING, X., BUNEA, F. and NILES-WEED, J. (2022). The Sketched Wasserstein Distance for mixture distributions. <https://doi.org/10.48550/ARXIV.2206.12768>
- BONNANS, J. F. and SHAPIRO, A. (2000). Perturbation Analysis of Optimization Problems. In *Springer Series in Operations Research*.
- CAFFARELLI, L. A. (1996). Boundary Regularity of Maps with Convex Potentials–II. *Annals of Mathematics* **144** 453–496.
- CORDERO-ERAUSQUIN, D. and FIGALLI, A. (2019). Regularity of monotone transport maps between unbounded domains. *Discrete Contin. Dyn. Syst.* **39** 7101–7112.
- CÁRCAMO, J., CUEVAS, A. and RODRÍGUEZ, L.-A. (2020). Directional differentiability for supremum-type functionals: Statistical applications. *Bernoulli* **26** 2143–2175. <https://doi.org/10.3150/19-BEJ1188>
- DE LARA, L., GONZÁLEZ SANZ, A., ASHER, N. and LOUBES, J.-M. (2021). Transport-based Counterfactual Models. *Pre-Print*.
- DEL BARRIO, E., GONZÁLEZ SANZ, A. and HALLIN, M. (2022). Nonparametric Multiple-Output Center-Outward Quantile Regression. <https://doi.org/10.48550/ARXIV.2204.11756>
- DEL BARRIO, E. and LOUBES, J.-M. (2019). Central limit theorems for empirical transportation cost in general dimension. *Ann. Probab.* **47** 926–951. <https://doi.org/10.1214/18-AOP1275>
- DEL BARRIO, E., GONZÁLEZ SANZ, A. and HALLIN, M. (2020). A note on the regularity of optimal-transport-based center-outward distribution and quantile functions. *J. Multivariate Anal.* **180** 104671. <https://doi.org/10.1016/j.jmva.2020.104671>
- DEL BARRIO, E., SANZ, A. G. and LOUBES, J.-M. (2021). Central Limit Theorems for General Transportation Costs. *To appear in Ann. Henri Poincaré*.
- DEL BARRIO, E., SANZ, A. G., LOUBES, J.-M. and NILES-WEED, J. (2022). An improved central limit theorem and fast convergence rates for entropic transportation costs. *ArXiv* **abs/2204.09105**.
- FANG, Z. and SANTOS, A. (2018). Inference on Directionally Differentiable Functions. *The Review of Economic Studies* **86** 377–412. <https://doi.org/10.1093/restud/rdy049>
- GALICHON, A. (2016). *Optimal Transport Methods in Economics*. Princeton University Press. <https://doi.org/10.23943/princeton/9780691172767.001.0001>
- GALLOUËT, T. and MÉRIGOT, Q. (2018). A Lagrangian Scheme à la Brenier for the Incompressible Euler Equations. *Found. Comput. Math.* **18** 835–865.

- GANGBO, W. and MCCANN, R. J. (1996). The geometry of optimal transportation. *Acta Math.* **177** 113–161. <https://doi.org/10.1007/BF02392620>
- GAYRAUD, N. T., RAKOTOMAMONJY, A. and CLERC, M. (2017). Optimal transport applied to transfer learning for P300 detection. In *BCI 2017-7th Graz Brain-Computer Interface Conference* 6.
- GINÉ, E. and NICKL, R. (2015). Mathematical Foundations of Infinite-Dimensional Statistical Models. In *Cambridge Series in Statistical and Probabilistic Mathematics, [40]*. Cambridge University Press, New York.
- GOLDFELD, Z., KATO, K., NIETERT, S. and RIOUX, G. (2022a). Limit distribution theory for smooth  $p$ -Wasserstein distances. <https://doi.org/10.48550/ARXIV.2203.00159>
- GOLDFELD, Z., KATO, K., RIOUX, G. and SADHU, R. (2022b). Statistical inference with regularized optimal transport. *arXiv preprint*.
- GOLDFELD, Z., KATO, K., RIOUX, G. and SADHU, R. (2022c). Limit Theorems for Entropic Optimal Transport Maps and the Sinkhorn Divergence. <https://doi.org/10.48550/ARXIV.2207.08683>
- GONZÁLEZ-DELGADO, J., GONZÁLEZ SANZ, A., CORTÉS, J. and NEUVIAL, P. (2021). Two-sample goodness-of-fit tests on the flat torus based on Wasserstein distance and their relevance to structural biology.
- GONZÁLEZ SANZ, A., LOUBES, J.-M. and NILES-WEED, J. (2022). Weak limits of entropy regularized Optimal Transport; potentials, plans and divergences. <https://doi.org/10.48550/ARXIV.2207.07427>
- GORDALIZA, P., DEL BARRIO, E., FABRICE, G. and LOUBES, J.-M. (2019). Obtaining Fairness using Optimal Transport Theory. In *Proceedings of the 36th International Conference on Machine Learning* (K. CHAUDHURI and R. SALAKHUTDINOV, eds.). *Proc. Mach. Learn. Res.* **97** 2357–2365. Proc. Mach. Learn. Res.
- GRAF, S. and LUSCHGY, H. (2000). *Foundations of Quantization for Probability Distributions*. Springer-Verlag, Berlin, Heidelberg.
- HALLIN, M., DEL BARRIO, E., CUESTA-ALBERTOS, J. and MATRÁN, C. (2021). Distribution and quantile functions, ranks and signs in dimension  $d$ : A measure transportation approach. *Ann. Statist.* **49** 1139–1165. <https://doi.org/10.1214/20-AOS1996>
- HARTMANN, V. and SCHUHMACHER, D. (2020). Semi-discrete optimal transport: a solution procedure for the unsquared Euclidean distance case. *Math. Meth. Oper. Res.* **92** 133–163. <https://doi.org/10.1561/22000000073>
- HUNDRIESER, S., STAUDT, T. and MUNK, A. (2022). Empirical Optimal Transport between Different Measures Adapts to Lower Complexity. *arXiv preprint*.
- HUNDRIESER, S., KLATT, M., STAUDT, T. and MUNK, A. (2022). A Unifying Approach to Distributional Limits for Empirical Optimal Transport. *arXiv preprint*.
- JIANG, R., PACCHIANO, A., STEPLETON, T., JIANG, H. and CHIAPPA, S. (2020). Wasserstein fair classification. In *Uncertainty in Artificial Intelligence* 862–872. Proc. Mach. Learn. Res.
- KITAGAWA, J., MÉRIGOT, Q. and THIBERT, B. (2019). Convergence of a Newton algorithm for semi-discrete optimal transport. *J. Eur. Math. Soc.* **21** 2603–2651. <https://doi.org/10.4171/JEMS/889>
- LINIAL, N., MANSOUR, Y. and RIVEST, R. L. (1991). Results on learnability and the Vapnik-Chervonenkis dimension. *Information and Computation* **90** 33–49. [https://doi.org/10.1016/0890-5401\(91\)90058-A](https://doi.org/10.1016/0890-5401(91)90058-A)
- LÉVY, B., MOHAYAEI, R. and VON HAUSEGGER, S. (2020). A fast semi-discrete optimal transport algorithm for a unique reconstruction of the early Universe. *Pre-Print*.
- MANOLE, T., BALAKRISHNAN, S. and WASSERMAN, L. A. (2022). Minimax confidence intervals for the Sliced Wasserstein distance. *Electron. J. Stat.*
- MASSART, P. and PICARD, J. (2007). Concentration inequalities and model selection. In *École d’été de probabilités de Saint-Flour*.
- MENA, G. and NILES-WEED, J. (2019). Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. In *Advances in Neural Information Processing Systems* **32**. Curran Associates, Inc.
- MEYRON, J. (2019). Initialization Procedures for Discrete and Semi-Discrete Optimal Transport. *Computer-Aided Design* **115** 13–22. <https://doi.org/10.1016/j.cad.2019.05.037>
- MILMAN, E. (2007). On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.* **177** 1–43.
- POLLARD, D. (1982). A Central Limit Theorem for  $k$ -Means Clustering. *Ann. Probab.* **10** 919–926. <https://doi.org/10.1214/aop/1176993713>
- POOLADIAN, A.-A., DIVOL, V. and NILES-WEED, J. (2023). Minimax estimation of discontinuous optimal transport maps: The semi-discrete case. <https://doi.org/10.48550/ARXIV.2301.11302>

- RATHMAIR, M. (2019). On how Poincaré inequalities imply weighted ones. *Monatsh Math* **118** 753–763.
- ROCKAFELLAR, R. T. (1970). *Convex analysis*. Princeton Mathematical Series. Princeton University Press, Princeton, N. J.
- ROCKAFELLAR, R. T. and WETS, R. J. B. (1998). *Variational Analysis*. Springer Verlag, Heidelberg, Berlin, New York.
- RÖMISCH, W. (2014). *Delta Method, Infinite Dimensional* In *Wiley StatsRef: Statistics Reference Online*. American Cancer Society. <https://doi.org/10.1002/9781118445112.stat01949>
- SADHU, R., GOLDFELD, Z. and KATO, K. (2021). Limit Distribution Theory for the Smooth 1-Wasserstein Distance with Applications. <https://doi.org/10.48550/ARXIV.2107.13494>
- SEGERS, J. (2022). Graphical and uniform consistency of estimated optimal transport plans. <https://doi.org/10.48550/ARXIV.2208.02508>
- SHEN, J., QU, Y., ZHANG, W. and YU, Y. (2018). Wasserstein Distance Guided Representation Learning for Domain Adaptation. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- SOMMERFELD, M. and MUNK, A. (2018). Inference for empirical Wasserstein distances on finite spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 219–238.
- TAMELING, C., SOMMERFELD, M. and MUNK, A. (2019). Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications. *Ann. Appl. Probab.* **29** 2744–2781. <https://doi.org/10.1214/19-AAP1463>
- VAART, A. W. V. D. and WELLNER, J. A. (1996). *Weak convergence and empirical processes*. Springer, New York, NY.
- VERDINELLI, I. and WASSERMAN, L. (2019). Hybrid Wasserstein distance and fast distribution clustering. *Electron. J. Stat.* **13** 5088–5119. <https://doi.org/10.1214/19-EJS1639>
- VILLANI, C. (2003). *Topics in optimal transportation*. Graduate studies in mathematics **58**. American Mathematical Soc.
- VILLANI, C. (2008). *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften **338**. Springer, Berlin. OCLC: ocn244421231.
- VITALE, R. A. (1985). Lp metrics for compact, convex sets. *J. Approx. Theory* **45** 280–287. [https://doi.org/10.1016/0021-9045\(85\)90051-6](https://doi.org/10.1016/0021-9045(85)90051-6)
- WEED, J. and BACH, F. (2019). Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli* **25** 2620–2648. <https://doi.org/10.3150/18-BEJ1065>

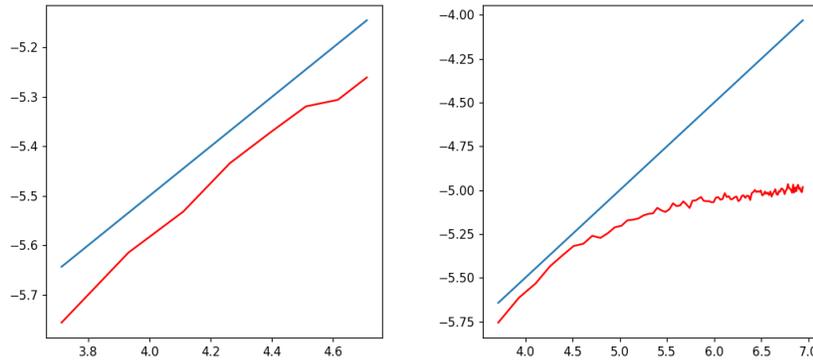
## Supplementary Material

### Supplementary Material of Central Limit Theorems for semi-discrete Wasserstein Distances

Supplementary Material of CLT for semi-discrete Wasserstein Distances

## 6. Simulations

First, we illustrate the precision of the upper bound in Theorem 2.6 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 5. 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.



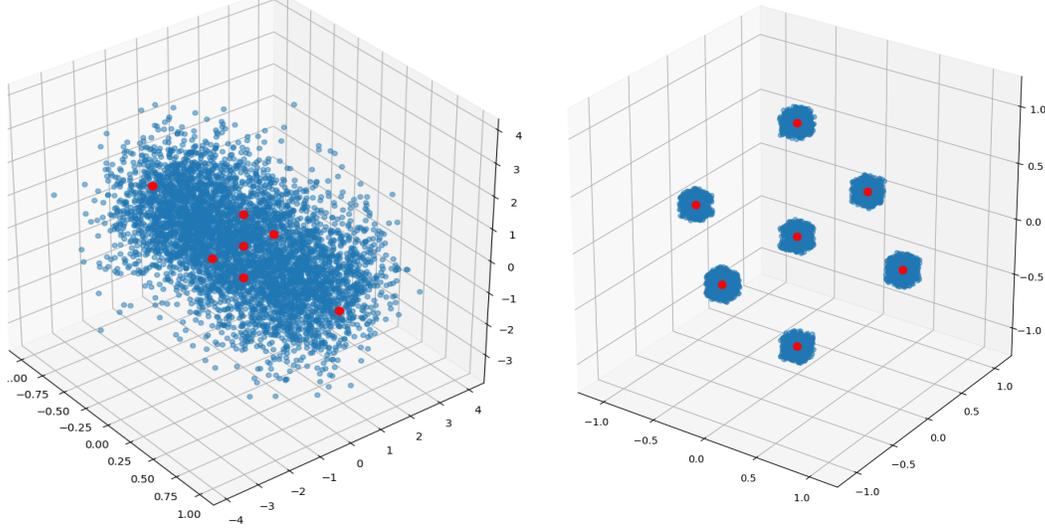
**Figure 5.** Plot, in double logarithmic scale, of  $E|\mathcal{W}_1(P^N, U_m) - \mathcal{W}_1(P^N, U)|$  (y axis) with respect to  $N$  (x axis).

The next part of section is devoted to illustrate empirically Theorems 2.4 and 3.2. 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 2.4 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, the dual problem has a unique solution. In consequence, the mapping is fully Hadamard differentiable (Corollary 2.4 in Cárcamo, Cuevas and Rodríguez (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

$$\mathbb{X} = \{\mathbf{x}_i\}_{i=1}^7 = \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (-1, 0, 0), (0, -1, 0), (0, 0, -1), (0, 0, 0)\}.$$

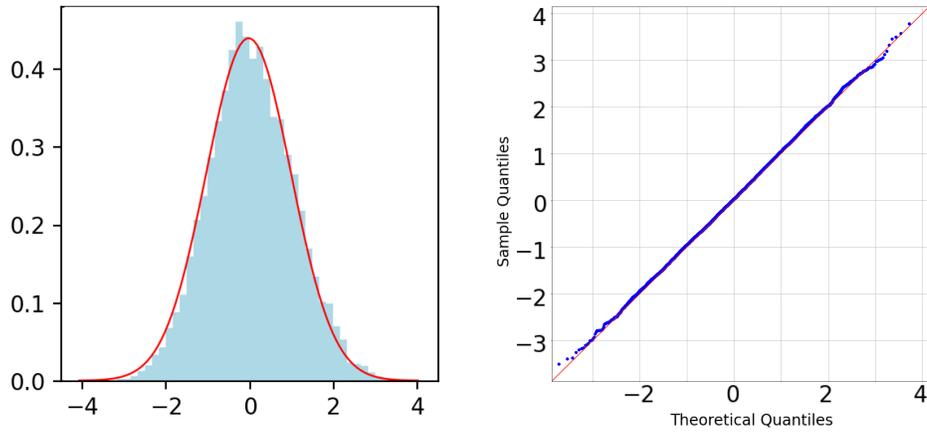
The continuous probability  $Q \in \mathcal{P}(\mathbb{R}^3)$  is the direct product  $\mathcal{U}(-1, 1) \times \mathcal{N}(0, 1) \times \mathcal{N}(0, 1)$ . Note that its



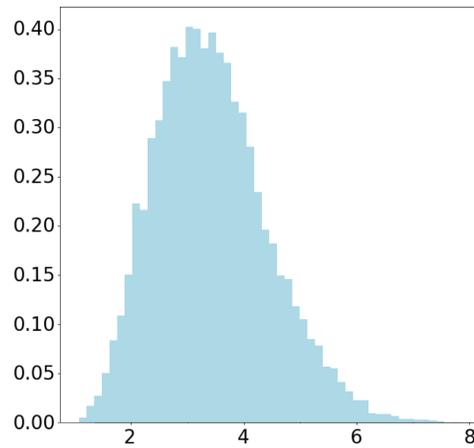
**Figure 6.** 3D visualization of the data set. In blue the continuous distribution  $Q$  and in red the discrete one  $P$ . Left:  $Q$  has a density with connected support and satisfies the assumptions of Theorem 3.2. Right:  $Q$  has a density but not a connected support.

support is connected with Lebesgue negligible boundary –we can visualize the data in Figure 6– and satisfies the assumptions of Theorem 3.2. 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 7 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 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 6. 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  $\sqrt{m} \frac{(\mathcal{W}_2^2(P, Q_m) - \mathcal{W}_2^2(P, Q))}{\sigma_2(Q_m, \mathbf{x}^m)}$  with the original sample. The results are shown in Figure 8, 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 8 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.



**Figure 7.** Illustration of Theorem 3.2 using bootstrap procedures. Histograms (left) and Q–Q plot (right) of the bootstrap estimation of  $\sqrt{m} \frac{W_2^2(P, Q_m) - W_2^2(P, Q)}{\sigma_2(Q_m, \mathbf{x}^m)}$ .



**Figure 8.** Illustration of Theorem 2.4, by using Monte Carlo's method, for  $Q$  with disconnected support.

## 7. Proofs

**Proof of Theorem 2.4.** 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

$$\sqrt{\frac{nm}{n+m}} \left( \sup_{|\mathbf{z}| \leq K} g_c(P_n, Q_m, \mathbf{z}) - \sup_{|\mathbf{z}| \leq K} g_c(P, Q, \mathbf{z}) \right)_{n,m}.$$

The weak limit depends on the set of restricted optimal points.

$$\text{Opt}_c^K(P, Q) := \left\{ \mathbf{z} : \sup_{|\mathbf{s}| \leq K} g_c(P, Q, \mathbf{s}) = g_c(P, Q, \mathbf{z}), \quad z_1 = 0 \right\}.$$

**Lemma 7.1.** *Set  $K > 0$ , under the assumptions of Theorem 2.4, we have the limit*

$$\sqrt{\frac{nm}{n+m}} \left( \sup_{|\mathbf{z}| \leq K} g_c(P_n, Q_m, \mathbf{z}) - \sup_{|\mathbf{z}| \leq K} g_c(P, Q, \mathbf{z}) \right) \xrightarrow{w} \sup_{\mathbf{z} \in \text{Opt}_c^K(P, Q)} \left( \sqrt{\lambda} \sum_{i=1}^N z_i \mathbb{U}_i + (\sqrt{1-\lambda}) \mathbb{G}_Q^c(\mathbf{z}) \right),$$

with  $(\mathbb{U}_1, \dots, \mathbb{U}_N)$  and  $\mathbb{G}_Q^c$  as in Theorem 2.4.

**Proof of Lemma 7.1.** For each  $K > 0$  we define the restricted set

$$\mathcal{F}_c^K = \left\{ \mathbf{y} \mapsto \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\}, \quad \mathbf{x}_i \in \mathbb{X}, \text{ and } z_1 = 0, |\mathbf{z}| \leq K \right\},$$

Lemma 8.1 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

$$C : \ell^\infty(\mathcal{F}_c^K) \longrightarrow \ell^\infty(\bar{\mathbb{B}}_K(\mathbf{0}))$$

$$f \mapsto \left( \mathbf{z} \mapsto f \left( \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\} \right) \right)$$

is actually continuous, hence for any  $f, g \in \ell^\infty(\mathcal{F}_c^K)$ , we have

$$\sup_{\mathbf{z} \in \bar{\mathbb{B}}_K(\mathbf{0})} \left| f \left( \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\} \right) - g \left( \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\} \right) \right| = \sup_{\phi \in \ell^\infty(\mathcal{F}_c^K)} |f(\phi) - g(\phi)|.$$

Moreover, the multivariate CLT implies  $\sqrt{n}(\mathbf{p}_n - \mathbf{p}) \xrightarrow{w} (\mathbb{U}_1, \dots, \mathbb{U}_N) \sim N(\mathbf{0}, \Sigma(\mathbf{p}))$ , where  $\Sigma(\mathbf{p})$  is defined in (11). Since the sequences  $\sqrt{n}(\mathbf{p}_n - \mathbf{p})$  and  $\sqrt{m}(Q_m - Q)$  are independent we derive the following result.

**Lemma 7.2.** *Under the assumptions of Theorem 2.4, we have the limit*

$$\sqrt{\frac{nm}{n+m}} (g_c(P_n, Q_m, \cdot) - g_c(P, Q, \cdot)) \xrightarrow{w} \sqrt{\lambda} \langle \mathbf{X}, \cdot \rangle + \sqrt{1-\lambda} C(\mathbb{G}_Q) \text{ in } \ell^\infty(\bar{\mathbb{B}}_K(\mathbf{0})), \quad (33)$$

with  $(\mathbb{U}_1, \dots, \mathbb{U}_N) = \mathbf{X}$ .

Let  $(\mathcal{B}, d)$  be a compact metric space, Corollary 2.3 in [Cárcamo, Cuevas and Rodríguez \(2020\)](#), 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  $C(\mathcal{B})$  (the space of continuous functions from  $\mathcal{B}$  to  $\mathbb{R}$ ) with respect to  $F$  in a direction  $G \in C(\mathcal{B})$ . Recall that a function  $f : \Theta \rightarrow \mathbb{R}$ , defined in a Banach space,  $\Theta$ , is said to be *Hadamard directionally differentiable* at  $\theta \in \Theta$  tangentially to  $\Theta_0 \subset \Theta$  if there exists a function  $f'_\theta : \Theta_0 \rightarrow \mathbb{R}$  such that

$$\frac{f(\theta + t_n h_n) - f(\theta)}{t_n} \xrightarrow{n \rightarrow \infty} f'_\theta(h), \text{ for all sequences } t_n \searrow 0 \text{ and } h_n \rightarrow h, \text{ for all } h \in \Theta_0.$$

If  $F \in C(\mathcal{B})$  is not identically 0, the precise formula for the derivative, provided by Corollary 2.3 in [Cárcamo, Cuevas and Rodríguez \(2020\)](#), is

$$\delta'_F(G) = \sup_{\{\mathbf{z}: F(\mathbf{z})=\delta(F)\}} G(\mathbf{z}), \text{ for } G \in C(\mathcal{B}). \quad (34)$$

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  $\text{Opt}_c^K(P, Q)$ . The following result rewrites (34) in our setting.

**Lemma 7.3.** *Set  $K > 0$ , under the assumptions of Theorem 2.4, the map  $\delta$  is Hadamard directionally differentiable at  $g_c(P, Q, \cdot)$ , tangentially to the set  $C(\bar{\mathbb{B}}_K(\mathbf{0}))$  with derivative, for  $G \in C(\bar{\mathbb{B}}_K(\mathbf{0}))$ ,*

$$\delta'_{g_c(P, Q, \cdot)}(G) = \sup_{\mathbf{z} \in \text{Opt}_c^K(P, Q)} G(\mathbf{z}).$$

The last step is the application of the delta-method. Let  $\Theta$  be a Banach space,  $\theta \in \Theta$  and  $\{Z_n\}_{n \in \mathbb{N}}$  be a sequence of random variables such that  $Z_n : \Omega_n \rightarrow \Theta$  and  $r_n(Z_n - \theta) \xrightarrow{w} Z$  for some sequence  $r_n \rightarrow +\infty$  and some random element  $Z$  that takes values in  $\Theta_0 \subset \Theta$ . If  $f : \Theta \rightarrow \mathbb{R}$  is Hadamard differentiable at  $\theta$  tangentially to  $\Theta_0 \subset \Theta$ , with derivative  $f'_\theta(\cdot) : \Theta_0 \rightarrow \mathbb{R}$ , 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 (33) belongs to  $C(\bar{\mathbb{B}}_K(\mathbf{0}))$ . 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

$$\sup_{f, g \in \mathcal{F}_c^K, \rho_Q(f, g) < \delta_n} |\mathbb{G}_Q(f) - \mathbb{G}_Q(g)| \xrightarrow{a.s.} 0. \quad (35)$$

We want now to analyze 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  $\mathbf{z}^f \in \mathbb{B}_K(\mathbf{0})$  such that  $f(\mathbf{y}) = \inf_{i=1, \dots, m} \{c(\mathbf{x}_i, \mathbf{y}) - z_i^f\}$ . Lemma 2.3 states that

$$\{f, g \in \mathcal{F}_c^K : |\mathbf{z}^f - \mathbf{z}^g| < \delta\} \subset \{f, g \in \mathcal{F}_c^K : \|f - g\|_\infty < \delta\} \subset \{f, g \in \mathcal{F}_c^K : \rho_Q(f, g) < \delta\}. \quad (36)$$

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 (36) and (35), 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 7.2 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 the Banach space  $\ell^\infty(\mathbb{B}_K(\mathbf{0}))$  having a version in  $C(\mathbb{B}_K(\mathbf{0}))$ . Applying the so-called delta-method to the function  $\delta$  and Lemma 7.3 we derive the limit

$$\sqrt{\frac{nm}{n+m}} \left( \sup_{|\mathbf{z}| \leq K} g_c(P_n, Q_m, \mathbf{z}) - \sup_{|\mathbf{z}| \leq K} g_c(P, Q, \mathbf{z}) \right) \xrightarrow{w} \sup_{\mathbf{z} \in \text{Opt}_c^K(P, Q)} Z(\mathbf{z}).$$

Note, that the process  $\mathbf{z} \mapsto \sqrt{1-\lambda} \mathbb{G}_Q(\inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\})$  is Gaussian in  $\mathbb{R}^N$  with covariance function  $\Xi_Q^c$ . 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.  $\square$

Unfortunately, the optimal solutions need not be universally bounded. In order to go from the bounded to the unbounded case, we observe that Lemma 2.1 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 2.1 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 (14), 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

$$\sqrt{\frac{nm}{n+m}} \left( \mathcal{T}_c(P_n, Q_m) - \sup_{|\mathbf{z}| \leq K^*+1} g_c(P_n, Q_m, \mathbf{z}) \right) + \sqrt{\frac{nm}{n+m}} \left( \sup_{|\mathbf{z}| \leq K^*+1} g_c(P_n, Q_m, \mathbf{z}) - \mathcal{T}_c(P, Q) \right), \quad (37)$$

and Lemma 7.1 implies the weak convergence of the second term to

$$\sup_{\mathbf{z} \in \text{Opt}_c^{K^*+1}(P, Q)} \left( \sqrt{\lambda} \sum_{i=1}^N z_i \mathbb{U}_i + (\sqrt{1-\lambda}) \mathbb{G}_Q^c(\mathbf{z}) \right) = \sup_{\mathbf{z} \in \text{Opt}_c^0(P, Q)} \left( \sqrt{\lambda} \sum_{i=1}^N z_i \mathbb{U}_i + (\sqrt{1-\lambda}) \mathbb{G}_Q^c(\mathbf{z}) \right),$$

where the equality is a direct consequence of Lemma 2.1. It only remains to prove that the first term of (37) 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 (37).

- The second one is  $K_{n,m} \geq K^* + 1$ , which implies the bound

$$0 \leq \mathcal{T}_c(P_n, Q_m) - \sup_{|\mathbf{z}| \leq K^*+1} g_c(P_n, Q_m, \mathbf{z}) \leq \sup_{|\mathbf{z}| \leq K_{n,m}} g_c(P_n, Q_m, \mathbf{z}) - \sup_{|\mathbf{z}| \leq K^*+1} g_c(P_n, Q_m, \mathbf{z}). \quad (38)$$

Note that the right side of the inequality (38) can be rewritten as

$$\begin{aligned} \sup_{|\mathbf{z}| \leq K_{n,m}} g_c(P_n, Q_m, \mathbf{z}) - \sup_{|\mathbf{z}| \leq K^*+1} g_c(P_n, Q_m, \mathbf{z}) \\ = \sup_{|\mathbf{z}| \leq K_{n,m}} \inf_{|\mathbf{z}'| \leq K^*+1} g_c(P_n, Q_m, \mathbf{z}) - g_c(P_n, Q_m, \mathbf{z}') \end{aligned}$$

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

$$\begin{aligned} |g_c(P_n, Q_m, \mathbf{z}) - g_c(P_n, Q_m, \mathbf{z}')| \\ \leq \sum_{i=1}^N |z_i - z'_i| p_i^n + \int | \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\} - \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z'_i\} | dQ_m(\mathbf{y}), \end{aligned}$$

we can conclude from (10) 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

$$0 \leq \mathcal{T}_c(P_n, Q_m) - \sup_{|\mathbf{z}| \leq K^*+1} g_c(P_n, Q_m, \mathbf{z}) \leq |K_{n,m} - K^* - 1| \mathbb{1}_{(K_{n,m} \geq K^*+1)}. \quad (39)$$

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 2.4.

**Remark 7.4.** When dealing with the case where the asymptotics depend only on the empirical distribution  $P_n$ , note that Assumption (13), which depends only on  $Q$

$$\int c(\mathbf{x}_i, \mathbf{y}) dQ(\mathbf{y}) < \infty, \text{ for all } i = 1, \dots, m,$$

is enough to prove the CLT. Actually, the multidimensional CLT yields that

$$\sqrt{n} (g_c(P_n, Q, \cdot) - g_c(P, Q, \cdot)) \xrightarrow{w} \langle \mathbf{X}, \cdot \rangle \text{ in } \ell^\infty(\mathbb{B}_K(\mathbf{0})),$$

with  $(\mathbb{U}_1, \dots, \mathbb{U}_N) = \mathbf{X}$ . Therefore, all of the arguments above can be now repeated verbatim. □

**Proof of Theorem 2.6.** 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

$$\mathcal{W}_1(P, Q) = \sup_{\mathbf{z} \in \mathbb{R}^N} g_1(P, Q, \mathbf{z}) = \sup_{|\mathbf{z}| \leq \text{diam}(\mathbb{X})} g_1(P, Q, \mathbf{z}), \quad (40)$$

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 (40), as

$$E \left| \sup_{|\mathbf{z}| \leq \text{diam}(\mathbb{X})} g_1(P, Q, \mathbf{z}) - \sup_{|\mathbf{z}| \leq \text{diam}(\mathbb{X})} g_1(P, Q_m, \mathbf{z}) \right|,$$

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

$$\mathcal{F}_1 = \left\{ \mathbf{y} \mapsto \inf_{i=1, \dots, N} \{|\mathbf{x}_i - \mathbf{y}| - z_i\}, \mathbf{x}_i \in \mathbb{X}, \text{ and } |\mathbf{z}| \leq \text{diam}(\mathbb{X}) \right\}.$$

We set  $D = \text{diam}(\mathbb{X})$  in order to simplify the following formulas. Denote by  $N(\epsilon, \mathcal{F}_1, \|\cdot\|_{L^2(Q_m)})$  the covering number with respect to the metric  $L^2(Q_m)$ . Lemma 4.14 in Massart and Picard (2007) and Lemma 2.3 imply that

$$\sqrt{\log \left( 2N(\epsilon, \mathcal{F}_1, \|\cdot\|_{L^2(Q_m)}) \right)} \leq \sqrt{N \log \left( \frac{2D}{\epsilon} + 1 \right) + \log(2)}.$$

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

$$\int_0^{2D+a_{N,m}} \sqrt{\log\left(2N(\epsilon, \mathcal{F}_1, \|\cdot\|_{L^2(Q_m)})\right)} d\epsilon \leq \sqrt{N} \int_0^{2D+a_{N,m}} \sqrt{\log\left(\frac{2D}{\epsilon} + 1\right)} d\epsilon + 2D + a_{N,m} \log(2). \quad (41)$$

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 (41). Therefore, using Jensen's inequality, we obtain

$$\begin{aligned} & \int_0^{2D+a_{N,m}} \sqrt{\log\left(\frac{2D}{\epsilon} + 1\right)} d\epsilon \\ & \leq (2D + a_{N,m}) \sqrt{\frac{2D \log(2D + a_{N,m}) + (4D + a_{N,m}) \log\left(\frac{4D+a_{N,m}}{2D+a_{N,m}}\right)}{2D + a_{N,m}}} \\ & \leq \sqrt{2D(2D + a_{N,m}) \log(2D + a_{N,m}) + (4D + a_{N,m})(2D + a_{N,m}) \log\left(\frac{4D + a_{N,m}}{2D + a_{N,m}}\right)}. \end{aligned}$$

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

$$\sqrt{m} E \left| \sup_{f \in \mathcal{F}_c^D} \int f(\mathbf{y})(dQ_m(\mathbf{y}) - dQ(\mathbf{y})) \right| \leq 8\sqrt{2}\sqrt{N} E \{(4D + a_{N,m}) (\log(2) + \sqrt{2D + 1})\}.$$

Using triangle inequality, we have

$$E a_{N,m} = E \left( 2\sqrt{\int \sup_{i=1,\dots,N} |\mathbf{y} - \mathbf{x}_i|^2 dQ_m(\mathbf{y})} \right) \leq 2\sqrt{\int |\mathbf{y}|^2 dQ(\mathbf{y})} + 2D,$$

which proves the result.

For the generalization to further potential costs, we observe that equation (2.5) in [del Barrio, Sanz and Loubes \(2021\)](#) yields

$$|z_i^*| \leq |\mathbf{x}_0 - \mathbf{x}_i| 4 \operatorname{diam}(\mathcal{Y})^{p-1}$$

and therefore, repeating the previous argumentation, we obtain the result.  $\square$

**Proof of Theorem 4.1.**

We note that the population potential is described through (20), 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

$$\nabla_{\mathbf{z}} \mathcal{M}_{\mathbf{p}}^{n,m}(\mathbf{z}) = (-Q_m(\operatorname{Lag}_1(\mathbf{z})) + p_1^n, \dots, -Q_m(\operatorname{Lag}_N(\mathbf{z})) + p_N^n), \quad (42)$$

where  $\operatorname{Lag}_k$ , for  $k = 1, \dots, N$ , are defined in (21). 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 7.5.** *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*

$$\begin{aligned} \nabla \mathcal{M}_{\mathbf{p}}(\mathbf{z}) &= \nabla_{\mathbf{z}} \mathcal{M}_{\mathbf{p}}(\mathbf{z})|_{\langle \mathbf{1} \rangle^\perp}, \\ D^2 \mathcal{M}_{\mathbf{p}}(\mathbf{z}) &= D_{\mathbf{z}^*}^2 \mathcal{M}_{\mathbf{p}}(\mathbf{z})|_{\langle \mathbf{1} \rangle^\perp}. \end{aligned}$$

Moreover, if  $\mathbf{z}^* \in \langle \mathbf{1} \rangle^\perp \cap \operatorname{Opt}_c(P, Q)$ , there exists a positive constant  $C$  such that

$$\langle D^2 \mathcal{M}_{\mathbf{p}}(\mathbf{z}^*) \mathbf{v}, \mathbf{v} \rangle \leq -C \inf_{i=1, \dots, N} |p_i|^3 |\mathbf{v}|^2, \text{ for all } \mathbf{v} \in \langle \mathbf{1} \rangle^\perp. \quad (43)$$

**Proof.** Note that it only remains to prove that (43) holds. But this is a direct consequence of (23). In fact, since  $\mathbf{z}^*$  is the unique  $\mathbf{z} \in \operatorname{Opt}(P, Q)$ , then  $Q(\operatorname{Lag}_k(\mathbf{z}^*)) = p_k$  for  $k = 1, \dots, N$  and we can conclude.  $\square$

We want now to prove the tightness of  $\sqrt{\frac{nm}{n+m}}|\hat{\mathbf{z}}^{n,m} - \mathbf{z}^*|$ . Note that Lemma 7.5 implies the well-separability property

$$\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*) - \mathcal{M}_{\mathbf{p}}(\hat{\mathbf{z}}^{n,m}) \geq c|\hat{\mathbf{z}}^{n,m} - \mathbf{z}^*|^2, \quad \text{for some } c > 0. \quad (44)$$

The relations

$$\begin{aligned} & |\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*) - \mathcal{M}_{\mathbf{p}}(\hat{\mathbf{z}}^{n,m}) - \mathcal{M}_{\mathbf{p}}^{n,m}(\mathbf{z}^*) + \mathcal{M}_{\mathbf{p}}^{n,m}(\hat{\mathbf{z}}^{n,m})| \\ &= \left| \sum_{k=2}^N (p_k - p_k^n)(z_k - \hat{z}_k^n) + \int \left( \inf_{i=2, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\} - \inf_{i=2, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - \hat{z}_i^n\} \right) (dQ - dQ_m)(\mathbf{y}) \right| \\ &\leq |\mathbf{p}^n - \mathbf{p}|_2 |\mathbf{z}^* - \hat{\mathbf{z}}^{n,m}|_2 + \left\| \inf_{i=2, \dots, N} \{c(\mathbf{x}_i, \cdot) - z_i\} - \inf_{i=2, \dots, N} \{c(\mathbf{x}_i, \cdot) - \hat{z}_i^n\} \right\|_\infty \|Q - Q_m\|_{\overline{\operatorname{span}(\mathcal{F}_c)}}, \end{aligned}$$

where  $\overline{\text{span}(\mathcal{F}_c)}$  is the topological closure in  $C(\mathcal{Y})$  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

$$|\mathcal{M}_{\mathbf{p}}(\mathbf{z}^*) - \mathcal{M}_{\mathbf{p}}(\hat{\mathbf{z}}^{n,m}) - \mathcal{M}_{\mathbf{p}}^{n,m}(\mathbf{z}^*) - \mathcal{M}_{\mathbf{p}}^{n,m}(\hat{\mathbf{z}}^{n,m})| \leq C \sqrt{\frac{n+m}{nm}} |\mathbf{z}^* - \hat{\mathbf{z}}^{n,m}|, \quad \text{for some } C > 0. \quad (45)$$

Therefore, (44) and (45) give

$$\begin{aligned} C \sqrt{\frac{n+m}{nm}} |\mathbf{z}^* - \hat{\mathbf{z}}^{n,m}| &\geq \mathcal{M}_{\mathbf{p}}(\mathbf{z}^*) - \mathcal{M}_{\mathbf{p}}(\hat{\mathbf{z}}^{n,m}) - \mathcal{M}_{\mathbf{p}}^{n,m}(\mathbf{z}^*) + \mathcal{M}_{\mathbf{p}}^{n,m}(\hat{\mathbf{z}}^{n,m}) \\ &\geq \mathcal{M}_{\mathbf{p}}(\mathbf{z}^*) - \mathcal{M}_{\mathbf{p}}(\hat{\mathbf{z}}^{n,m}) \geq c |\hat{\mathbf{z}}^{n,m} - \mathbf{z}^*|^2, \end{aligned}$$

from where we deduce the tightness of  $\sqrt{\frac{nm}{n+m}} |\hat{\mathbf{z}}^{n,m} - \mathbf{z}^*|$ .

Lemma 7.5 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

$$\nabla_{\mathbf{z}} \mathcal{M}_{\mathbf{p}}(\hat{\mathbf{z}}^{n,m}) - \nabla_{\mathbf{z}} \mathcal{M}_{\mathbf{p}}^{n,m}(\hat{\mathbf{z}}^{n,m}) = D^2 \mathcal{M}_{\mathbf{p}}(\mathbf{z}^*)(\hat{\mathbf{z}}^{n,m} - \mathbf{z}^*) + O_P(|\hat{\mathbf{z}}^{n,m} - \mathbf{z}^*|^2). \quad (46)$$

The left hand side of (46) can be written as

$$\begin{aligned} (Q_m(\text{Lag}_1(\hat{\mathbf{z}}^{n,m})) - Q(\text{Lag}_1(\hat{\mathbf{z}}^{n,m})) + p_1 - p_1^n, \dots, \\ Q_m(\text{Lag}_N(\hat{\mathbf{z}}^{n,m})) - Q(\text{Lag}_N(\hat{\mathbf{z}}^{n,m})) + p_N - p_N^n), \end{aligned}$$

The following result provides its weak limit.

**Lemma 7.6.** *Suppose that the assumptions of Lemma 7.5 hold. If  $m = m(n)$  is such that  $n \rightarrow \infty$  and  $\frac{m}{n+m} \rightarrow \lambda \in (0, 1)$ , we have*

$$\sqrt{\frac{nm}{n+m}} \left( \nabla_{\mathbf{z}} \mathcal{M}_{\mathbf{p}}(\hat{\mathbf{z}}^{n,m}) - \nabla_{\mathbf{z}} \mathcal{M}_{\mathbf{p}}^{n,m}(\hat{\mathbf{z}}^{n,m}) \right) \xrightarrow{w} (\mathbb{U}_1, \dots, \mathbb{U}_N).$$

**Proof.** From the relation

$$\begin{aligned} \nabla_{\mathbf{z}} \mathcal{M}_{\mathbf{p}}(\hat{\mathbf{z}}^{n,m}) - \nabla_{\mathbf{z}} \mathcal{M}_{\mathbf{p}}^{n,m}(\hat{\mathbf{z}}^{n,m}) \\ = (Q_m(\text{Lag}_1(\hat{\mathbf{z}}^{n,m})) - Q(\text{Lag}_1(\hat{\mathbf{z}}^{n,m})) + p_1 - p_1^n, \\ \dots, Q_m(\text{Lag}_N(\hat{\mathbf{z}}^{n,m})) - Q(\text{Lag}_N(\hat{\mathbf{z}}^{n,m})) + p_N - p_N^n) \end{aligned}$$

and the multivariate central limit theorem, we claim the convergence

$$\sqrt{m}(Q_m(\text{Lag}_1(\hat{\mathbf{z}}^{n,m})) - Q(\text{Lag}_1(\hat{\mathbf{z}}^{n,m})), \dots, Q_m(\text{Lag}_N(\hat{\mathbf{z}}^{n,m})) - Q(\text{Lag}_N(\hat{\mathbf{z}}^{n,m}))) \xrightarrow{w} (\mathbb{U}_1, \dots, \mathbb{U}_N).$$

To prove it we set  $k$  and observe that

$$\text{Lag}_k(\mathbf{z}) = \bigcap_{i=1, \dots, N} \{\mathbf{y} \in \mathbb{R}^d : c(\mathbf{x}_k, \mathbf{y}) - c(\mathbf{x}_i, \mathbf{y}) < z_k - z_i\}, \quad (47)$$

which means that the class of all possible cells  $\{\text{Lag}_k(\mathbf{z}) : \mathbf{z} \in \mathbb{R}^N\} \subset 2^{\mathcal{Y}}$  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

$$\{\{\mathbf{y} \in \mathbb{R}^d : c(\mathbf{x}_k, \mathbf{y}) - c(\mathbf{x}_i, \mathbf{y}) - s < 0\} : s \in \mathbb{R}\}$$

is  $D$ , the one of  $\{\text{Lag}_k(\mathbf{z}) : \mathbf{z} \in \mathbb{R}^N\}$  is  $2ND \log(3N)$  (see Lemma 3.2.3. in [Linial, Mansour and Rivest \(1991\)](#)). 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

$$\mathcal{N}(\epsilon, \{\mathbb{1}_{\text{Lag}_k(\mathbf{z})} : \mathbf{z} \in \mathbb{R}^N\}, L_2(Q)) \leq C \frac{1}{\epsilon^{8N \log(3N)}},$$

which means that the class  $\{\mathbb{1}_{\text{Lag}_k(\mathbf{z})} : \mathbf{z} \in \mathbb{R}^N\}$  is  $Q$ -Donsker. This means that, for every  $\epsilon > 0$

$$\lim_{\delta \rightarrow 0} \limsup_{m \rightarrow \infty} \mathbb{P} \left( \sup_{Q(\text{Lag}_k(\mathbf{u}) \Delta \text{Lag}_k(\mathbf{v})) < \delta} \left| \sqrt{m} \int \mathbb{1}_{\text{Lag}_k(\mathbf{u})} - \mathbb{1}_{\text{Lag}_k(\mathbf{v})} d(Q_m - Q) \right| > \epsilon \right) = 0, \quad (48)$$

where

$$\begin{aligned} Q(\text{Lag}_k(\mathbf{u}) \Delta \text{Lag}_k(\mathbf{v})) &= Q((\text{Lag}_k(\mathbf{u}) \setminus \text{Lag}_k(\mathbf{v})) \cup (\text{Lag}_k(\mathbf{v}) \setminus \text{Lag}_k(\mathbf{u}))) \\ &= \int (\mathbb{1}_{\text{Lag}_k(\mathbf{u})} - \mathbb{1}_{\text{Lag}_k(\mathbf{v})})^2 dQ. \end{aligned}$$

First we bound

$$Q(\text{Lag}_k(\hat{\mathbf{z}}^{n,m}) \Delta \text{Lag}_k(\mathbf{z}^*)) \leq Q(\text{Lag}_k(\hat{\mathbf{z}}^{n,m}) \setminus \text{Lag}_k(\mathbf{z}^*)) + Q(\text{Lag}_k(\mathbf{z}^*) \setminus \text{Lag}_k(\hat{\mathbf{z}}^{n,m})) \quad (49)$$

The we show that  $Q(\text{Lag}_k(\hat{\mathbf{z}}^{n,m}) \setminus \text{Lag}_k(\mathbf{z}^*)) \xrightarrow{P} 0$ , the same holds for  $Q(\text{Lag}_k(\mathbf{z}^*) \setminus \text{Lag}_k(\hat{\mathbf{z}}^{n,m}))$ , yielding the limit in probability of (49) towards 0. Using (47), we have

$$\text{Lag}_k(\hat{\mathbf{z}}^{n,m}) \setminus \text{Lag}_k(\mathbf{z}^*) = \bigcup_{i=1, \dots, N} \text{Lag}_k(\hat{\mathbf{z}}^{n,m}) \setminus \{\mathbf{y} \in \mathbb{R}^d : c(\mathbf{x}_k, \mathbf{y}) - c(\mathbf{x}_i, \mathbf{y}) < z_k^* - z_i^*\}.$$

which, in view of the union bound and (47), gives

$$\begin{aligned} Q(\text{Lag}_k(\hat{\mathbf{z}}^{n,m}) \setminus \text{Lag}_k(\mathbf{z}^*)) & \\ & \leq \sum_{i=1}^N Q(\text{Lag}_k(\hat{\mathbf{z}}^{n,m}) \setminus \{\mathbf{y} \in \mathbb{R}^d : c(\mathbf{x}_k, \mathbf{y}) - c(\mathbf{x}_i, \mathbf{y}) < z_k^* - z_i^*\}) \\ & \leq \sum_{i=1}^N Q(\{\mathbf{y} \in \mathbb{R}^d : z_k^{n,m} - z_i^{n,m} < c(\mathbf{x}_k, \mathbf{y}) - c(\mathbf{x}_i, \mathbf{y}) < z_k^* - z_i^*\}). \end{aligned}$$

Since  $\sqrt{\frac{nm}{n+m}}|\hat{\mathbf{z}}^{n,m} - \mathbf{z}^*$  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  $z_k^{n,m} - z_i^{n,m} \rightarrow z_k^* - z_i^*$  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,  $Q(\text{Lag}_k(\hat{\mathbf{z}}^{n,m}) \setminus \text{Lag}_k(\mathbf{z}^*)) \xrightarrow{P} 0$  and, by symmetry,

$$Q(\text{Lag}_k(\hat{\mathbf{z}}^{n,m}) \Delta \text{Lag}_k(\mathbf{z}^*)) \xrightarrow{P} 0. \quad (50)$$

Moreover, by repeating the same argument, we obtain that if  $\mathbf{u} \rightarrow \mathbf{v}$ , then

$$Q(\text{Lag}_k(\mathbf{u}) \Delta \text{Lag}_k(\mathbf{v})) \rightarrow 0 \quad (51)$$

Thanks to (50), (51) and (48), we are under the hypotheses of Lemma 3.3.5 in Vaart and Wellner (1996); as a consequence we have

$$\begin{aligned} \sqrt{m}|Q_m(\text{Lag}_k(\hat{\mathbf{z}}^{n,m})) - Q(\text{Lag}_k(\hat{\mathbf{z}}^{n,m})) - (Q_m(\text{Lag}_k(\mathbf{z}^*)) - Q(\text{Lag}_k(\mathbf{z}^*)))| \\ = \sqrt{m} \left| \int \mathbb{1}_{\text{Lag}_k(\hat{\mathbf{z}}^{n,m})} - \mathbb{1}_{\text{Lag}_k(\mathbf{z}^*)} d(Q_m - Q) \right| \xrightarrow{P} 0. \end{aligned}$$

This implies

$$\sqrt{m}(Q_m(\text{Lag}_1(\hat{\mathbf{z}}^{n,m})) - Q(\text{Lag}_1(\hat{\mathbf{z}}^{n,m}))) = \sqrt{m}(Q_m(\text{Lag}_k(\mathbf{z}^*)) - Q(\text{Lag}_k(\mathbf{z}^*))) + o_P(1).$$

As  $k$  was chosen arbitrarily, we obtain also

$$\begin{aligned} \sqrt{m}(Q_m(\text{Lag}_1(\hat{\mathbf{z}}^{n,m})) - Q(\text{Lag}_1(\hat{\mathbf{z}}^{n,m})), \dots, Q_m(\text{Lag}_N(\hat{\mathbf{z}}^{n,m})) - Q(\text{Lag}_N(\hat{\mathbf{z}}^{n,m}))) \\ \sqrt{m}(Q_m(\text{Lag}_1(\hat{\mathbf{z}}^*)) - Q(\text{Lag}_1(\hat{\mathbf{z}}^*)), \dots, Q_m(\text{Lag}_N(\hat{\mathbf{z}}^*)) - Q(\text{Lag}_N(\hat{\mathbf{z}}^*))) + o_P(1). \end{aligned}$$

The multi-variate central limit theorem yields the limit of

$$\sqrt{m}(Q_m(\text{Lag}_1(\hat{\mathbf{z}}^*)) - Q(\text{Lag}_1(\hat{\mathbf{z}}^*)), \dots, Q_m(\text{Lag}_N(\hat{\mathbf{z}}^*)) - Q(\text{Lag}_N(\hat{\mathbf{z}}^*)))$$

as a centered Gaussian r.v.  $(X_1, \dots, X_N)$  with covariance

$$E(X_i X_j) = Q(\text{Lag}_i(\hat{\mathbf{z}}^*) \cap \text{Lag}_j(\hat{\mathbf{z}}^*)) - Q(\text{Lag}_i(\hat{\mathbf{z}}^*))Q(\text{Lag}_j(\hat{\mathbf{z}}^*)) = \mathbb{1}_{i=j} p_i^2 - p_i p_j.$$

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.  $\square$

Lemma 7.6, (46) and the continuous mapping theorem conclude the proof.  $\square$

### Proof of Theorem 4.3.

The first step is to show that (28) holds. Set  $k \in \{1, \dots, N\}$ . Since

$$\text{Lag}_k^R(\mathbf{z}) = R\mathbb{B}_d \cap \bigcap_{i=1, \dots, N} \{\mathbf{y} \in \mathbb{R}^d : 2\langle \mathbf{x}_k - \mathbf{x}_i, \mathbf{y} \rangle \geq |\mathbf{x}_k|^2 - |\mathbf{x}_i|^2 - z_k + z_i\} \cap R\mathbb{B}_d,$$

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 4.5, we have

$$h_{\text{Lag}_k^R(\mathbf{z})}(\mathbf{v}) = \inf \left\{ R|\mathbf{v}_k| + \sum_{i \neq k} h_i(\mathbf{v}_i) : \sum_{i=1}^N \mathbf{v}_i = \mathbf{v} \right\}.$$

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

$$h_{\text{Lag}_k^R(\mathbf{z})}(\mathbf{v}) = \inf \left\{ R|\mathbf{v} - \sum_{i \neq k} t_i(\mathbf{x}_i - \mathbf{x}_k)| + \sum_{i \neq k} t_i(-|\mathbf{x}_k|^2 + |\mathbf{x}_i|^2 + z_k - z_i) : t_i \geq 0 \right\}$$

and (27) 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 4.5, (28) holds.

The next step is to show the Hadamard differentiability of  $h_{\text{Lag}_k^R(\mathbf{z}) + \{-\mathbf{y}^0\}}(\mathbf{v})$  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  $2N$ -Lipschitz w.r.t.  $\mathbf{z}$ —it is immediate to check out that  $|F(\mathbf{z}, \mathbf{t}) - F(\mathbf{z}', \mathbf{t})| \leq 2N \|\mathbf{z} - \mathbf{z}'\|_\infty$ — and by the relation

$$|\inf_{\mathbf{t}} F(\mathbf{z}, \mathbf{t}) - \inf_{\mathbf{t}} F(\mathbf{z}', \mathbf{t})| \leq \sup_{\mathbf{t}} |F(\mathbf{z}, \mathbf{t}) - F(\mathbf{z}', \mathbf{t})| \leq 2N \|\mathbf{z} - \mathbf{z}'\|_\infty,$$

the same holds for  $h_{\text{Lag}_k^R(\mathbf{z}) + \{-\mathbf{y}^0\}}(\mathbf{v})$ . Then we only need to prove the Gateaux differentiability. Sufficient conditions are given by Proposition 4.13 in [Bonnans and Shapiro \(2000\)](#), which are as follows:

1. The function  $F(\cdot, \mathbf{t})$  is Gateaux differentiable in  $\mathbb{R}^N$  with derivative  $D_{\mathbf{z}}F(\mathbf{z}, \mathbf{t})$ ,
2. the function  $F$  is continuous in  $\mathbb{R}^N \times \mathbb{R}^{N-1}$ , and
3. 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

$$\text{Lev}_\alpha F(\mathbf{z}, \cdot) = \{\mathbf{t} \in \mathbb{R}^{N-1} : t_i \geq 0, F(\mathbf{z}, \mathbf{t}) \leq \alpha\}$$

is non empty and contained in  $C$ .

Since the objective function is linear in  $\mathbf{z}$  and continuous in  $\mathbf{t}$ , (1) and (2) 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  $\mathbf{z} \rightarrow \psi_j(\mathbf{z})$ , 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,

$$\frac{a}{2} \sum_{j \neq k} t_j \leq \sum_{j \neq k} t_j \psi_j(\mathbf{z}) \leq h_{\text{Lag}_k^R(\mathbf{z}) + \{-\mathbf{y}^0\}}(\mathbf{v}) \leq 2R,$$

for all  $\mathbf{z} \in \mathcal{U}^*$ . The set of  $\mathbf{t} \in \mathbb{R}^{N-1}$  such that  $t_j \geq 0$  and  $\sum_{j \neq k} t_j \leq \frac{4R}{a}$  plays the role of  $C$  in (3), which is thus proven. The Hadamard derivative in a direction  $\mathbf{d}$  of  $h_{\text{Lag}_k^R(\mathbf{z}) + \{-\mathbf{y}^0\}}(\mathbf{v})$  in  $\mathbf{z}^*$  is thus given by

$$\inf_{\mathbf{t} \in \text{Sol}(\mathbf{z}^*, \mathbf{v}, \mathbf{y}^0)} \left\{ \sum_{j \neq k} t_j (d_k - d_j) \right\}.$$

Set  $\mathbf{d} \in \mathbb{R}^N$ . Now we claim that

$$\begin{aligned} & \frac{\left( \int \left| h_{\text{Lag}_k^R(\mathbf{z}^* + t_n \mathbf{d}_n)}(\mathbf{v}) - h_{\text{Lag}_k^R(\mathbf{z}^*)}(\mathbf{v}) \right|^p d\mathcal{H}^{d-1}(\mathbf{v}) \right)^{\frac{1}{p}}}{t_n} \\ & \rightarrow \left( \int \left| \inf_{\mathbf{t} \in \text{Sol}(\mathbf{z}^*, \mathbf{v}, \mathbf{y}^0)} \left\{ \sum_{j \neq k} t_j (d_k - d_j) \right\} \right|^p \right)^{\frac{1}{p}}, \quad (52) \end{aligned}$$

for any  $t_n \searrow 0$  and  $\mathbf{d}_n \rightarrow \mathbf{d}$ . Lemma 4.6 states the existence of some  $M > 0$  such that that

$$\limsup_n t_n d_\infty(\text{Lag}_k^R(\mathbf{z}^* + t_n \mathbf{h}_n), \text{Lag}_k^R(\mathbf{z}^*)) \leq M.$$

The quantity  $(t_n d_\infty(\text{Lag}_k^R(\mathbf{z}^* + t_n \mathbf{h}_n)))^p$  is constant w.r.t.  $\mathbf{v}$  and dominates

$$\frac{\left| h_{\text{Lag}_k^R(\mathbf{z}^* + t_n \mathbf{d}_n)}(\mathbf{v}) - h_{\text{Lag}_k^R(\mathbf{z}^*)}(\mathbf{v}) \right|^p}{t_n},$$

for all  $\mathbf{v} \in \mathbb{S}^{d-1}$ . In view of the dominated convergence theorem and the point-wise Hadamard differentiability (previous step), (52) holds. Finally, following verbatim the proof of Theorem 1 in Römisch (2014), we conclude.  $\square$

## 8. Proofs of Lemmas

### Proof of Lemma 2.1.

First, strong duality (3) yields that

$$\begin{aligned} \mathcal{T}_c(P, Q) &= \sup_{(f, g) \in \Phi_c(P, Q)} \int f(\mathbf{x}) dP(\mathbf{x}) + \int g(\mathbf{y}) dQ(\mathbf{y}) \\ &= \sup_{(f, g) \in \Phi_c(P, Q)} \sum_{i=1}^N f(\mathbf{x}_i) p_i + \int g(\mathbf{y}) dQ(\mathbf{y}). \end{aligned}$$

Set  $(z_1, \dots, z_N) = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_N))$ , 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  $i = 1, \dots, N$ . Therefore,  $g(\mathbf{y}) \leq \inf_{i=1, \dots, N} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\}$  and  $\mathcal{T}_c(P, Q) = \sup_{\mathbf{z} \in \mathbb{R}^N} g_c(P, Q, \mathbf{z})$ .

Let  $\mathbf{z}^* = (z_1^*, \dots, z_N^*) \in \mathbb{R}^N$  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

$$\begin{aligned} \mathcal{T}_c(P, Q) &\leq \sum_{i=1}^N z_i^* p_i + \int \{c(\mathbf{y}, \mathbf{x}_u) - z_u^*\} dQ(\mathbf{y}) \\ &\leq (1 - p_l) z_u^* + p_l z_l^* + \int \{c(\mathbf{y}, \mathbf{x}_u) - z_u^*\} dQ(\mathbf{y}) \\ &\leq -p_l z_u^* + p_l z_l^* + \int c(\mathbf{y}, \mathbf{x}_u) dQ(\mathbf{y}) \\ &\leq p_l (z_l^* - z_u^*) + \int c(\mathbf{y}, \mathbf{x}_u) dQ(\mathbf{y}), \end{aligned}$$

which implies  $p_l (z_u^* - z_l^*) \leq \int c(\mathbf{y}, \mathbf{x}_u) dQ(\mathbf{y}) - \mathcal{T}_c(P, Q)$  and

$$\sup_{i, j=1, \dots, N} |z_i^* - z_j^*| \leq \frac{1}{\inf_i p_i} \left( \sup_{i=1, \dots, N} \int c(\mathbf{y}, \mathbf{x}_i) dQ(\mathbf{y}) - \mathcal{T}_c(P, Q) \right).$$

Since adding additive constant does not change  $g_c(P, Q, \mathbf{z})$ , we conclude.  $\square$

**Lemma 8.1.** *Under the assumptions of Theorem 2.4, the class  $\mathcal{F}_c^K$  is  $Q$ -Donsker.*

### Proof of Lemma 8.1.

We use bracketing numbers, see Definition 2.1.6 in Vaart and Wellner (1996). Lemma 2.3 implies that

$$N_{[]} (2\epsilon, \mathcal{F}_c^K, \|\cdot\|_{L^2(Q)}) \leq N(\epsilon, \mathbb{B}_K(\mathbf{0}), |\cdot|).$$

Therefore, Lemma 4.14 in [Massart and Picard \(2007\)](#) implies that

$$\int_0^\infty \sqrt{\log \left( N_{[]}(\epsilon, \mathcal{F}_c^K, \|\cdot\|_{L^2(Q)}) \right)} d\epsilon < \infty. \quad (53)$$

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

$$\int F(\mathbf{y})^2 dQ(\mathbf{y}) \leq 2K + 2 \int \sup_{i=1, \dots, m} c(\mathbf{x}_i, \mathbf{y})^2 dQ(\mathbf{y}) < \infty.$$

Using Theorem 3.7.38 in [Giné and Nickl \(2015\)](#) we obtain the result.  $\square$