



**HAL**  
open science

# A Central Limit Theorem 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 Theorem for Semidiscrete Wasserstein Distances. 2021. hal-03232450v1

**HAL Id: hal-03232450**

**<https://hal.science/hal-03232450v1>**

Preprint submitted on 21 May 2021 (v1), last revised 16 Feb 2023 (v3)

**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.

# A Central Limit Theorem for Semidiscrete Wasserstein Distances.

Eustasio del Barrio<sup>(1)\*</sup>, Alberto González-Sanz<sup>(2)</sup>, and Jean-Michel Loubes<sup>(3)†</sup>

<sup>(1)(2)</sup>IMUVA, Universidad de Valladolid, Spain

<sup>(2)(3)</sup>IMT, Université de Toulouse France

<sup>(1)</sup>tasio@eio.uva.es    <sup>(2)</sup>alberto.gonzalez sanz@math.univ-toulouse.fr

<sup>(3)</sup>loubes@math.univ-toulouse.fr

## Abstract

We address the problem of proving a Central Limit Theorem for the empirical optimal transport cost,  $\sqrt{n}\{\mathcal{T}_c(P_n, Q) - \mathcal{W}_c(P, Q)\}$ , in the semi discrete case, i.e when the distribution  $P$  is finitely supported. We show that the asymptotic distribution is the supremum 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  $p \geq 1$ . Finally, the semidiscrete framework provides a control on the second derivative of the dual formulation, which yields the first central limit theorem for the optimal transport potentials.

## 1 Introduction

A large number of problems in statistics or computer science require the comparison between histograms or, more generally, measures. Optimal transport has proven to be an important tool to compare probability measures since it enables to define a metric over the set of distributions which convey 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

---

\*Research 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.

†Research 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.

variety of fields, moreover, in statistics and in particular in Machine learning, OT based methods have been developed to tackle problems in fairness as in [Jiang et al., 2020, Gordaliza et al., 2019, del Barrio et al., 2019], in domain adaptation ([Shen et al., 2018]), or transfer learning ([Gayraud et al., 2017]). Hence there is a growing need for theoretical results to support such applications and provide theoretical guarantees on the asymptotic distribution.

In this framework, the optimal transport between distributions has to be estimated from a random sample of the empirical distributions. Hence some controls have to be added on the rate of convergence of the the empirical version of the optimal transport cost to its true value. Moreover determining the asymptotic behavior of the limit distribution enables to control precisely the deviations of the estimation errors and certify the accuracy of the estimation.

This work focuses on the semi-discrete optimal transport, i.e. when one of both probabilities is supported on a discrete set. Such a problem has been studied in many contexts, including on resource allocation problem, points versus demand distribution, positions of sites such that the mean allocation cost is minimal ([Hartmann and Schuhmacher, 2020]), 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]), generation of blue noise distributions with applications for instance to low-level hardware implementation in printers( [de Goes et al., 2012]), in astronomy ([Lévy et al., 2020]). But in a more statistical point of view it can also be used to implement Goodness-of-fit-tests, in detecting 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] and to the new transport based generalization of the distribution function, proposed by [del Barrio et al., 2020], when the probability is discrete.

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

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

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

We assume the observation of a sample  $X_1, \dots, X_n$  i.i.d. with law  $P$ . The empirical measure  $P_n = \frac{1}{n} \sum_{k=1}^n \delta_{X_k}$  defines a random  $\mathcal{T}_c(P_n, Q)$ . Supposing that  $P, Q \in \mathcal{P}(\mathbb{R}^d)$  and both are absolutely continuous with respect to the Lebesgue measure, [Fournier and Guillin, 2015] proved that  $E\mathcal{W}_1(P_n, Q)$  converges to 0 with an order approximately of  $\frac{1}{n^{1/d}}$ . By using the triangular inequality, valid in Wasserstein distances, we can not expect better order of convergence for the difference  $\{\mathcal{W}_1(P_n, Q) - \mathcal{W}_1(P, Q)\}$ . But when it comes to the difference  $\sqrt{n}\{\mathcal{T}_c(P_n, Q) - E\mathcal{T}_c(P_n, Q)\}$ , the results of [del Barrio et al., 2021, del Barrio and Loubes, 2019] prove, using the Efron-Stein's inequality, that it follows a Gaussian asymptotic behavior. In the case of  $P, Q$  being supported in a finite (resp. countable) set, in [Sommerfeld and Munk, 2018] (resp. [Tameling et al., 2019]) proved that  $\sqrt{n}\{\mathcal{W}_c(P_n, Q) - \mathcal{W}_c(P, Q)\}$  has a weak limit  $X$ . Such a limit is the supremum of a Gaussian process. They start by the identification of the space of distribution supported in a finite set  $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$  with  $\mathbb{R}^m$ , and then give a proof based on the directional Hadamard differentiability of the functional  $(\mathbf{p}, \mathbf{q}) \mapsto \mathcal{T}_c(\sum_{i=1}^m p_i \delta_{\mathbf{x}_i}, \sum_{i=1}^m q_i \delta_{\mathbf{x}_i})$ .

In this work we are concerned with the asymptotic behaviour of  $\sqrt{n}\{\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q)\}$  when  $P$  is finitely supported but not  $Q$ . When the quadratic cost is involved and  $Q$  is absolutely continuous with respect to the Lebesgue measure, with convex support, [del Barrio and Loubes, 2019] proves that the limit is in fact Gaussian. Their approach is also based in some differentiability properties of the optimal transport problem. But one of their main arguments is that the optimal transport potential is unique which is no longer true in general costs and neither in general Polish spaces. We need to use weaker notions of derivative and follow similar arguments of those of [Sommerfeld and Munk, 2018] and [Tameling et al., 2019] to derive our main result, Theorem 2.4. As pointed out before, the result of [Sommerfeld and Munk, 2018] establishes that, if  $P$  and  $Q$  are both supported in a finite set, then

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

and  $(X_1, \dots, X_n)$  is a centered Gaussian vector and  $\text{Opt}_c(P, Q)$  is the set of solutions of the dual problem (3), both described in section 2. Lately [Tameling et al., 2019] extended the same result for probabilities supported

in countable spaces. Our work covers and generalizes the previously mentioned result of [Sommerfeld and Munk, 2018] through Theorem 2.4. We use similar tools but we make a slightly different extension. First we define in (5) a functional  $\mathcal{M}$  continuously differentiable with respect to  $\mathbf{p}$  in the positive hyperoctant, and where the optimization is made in the whole space  $\mathbb{R}^m$ , see Lemma 2.1. Hence it is a unconstrained optimization problem, consequently we can apply Danskin's theorem (Theorem 4.13 in [Bonnans and Shapiro, 2000]) instead of its Lagrangian counterpart, i.e. Theorem 4.24 in [Bonnans and Shapiro, 2000]). Such a slightly modification of the functional allows to realize that the central limit theorem holds assuming only finite support on one of both probabilities, which yields the central limit theorem for the semidiscrete case. Summarizing, the main difference between the approach of [Sommerfeld and Munk, 2018] and the one we make in section 2 is the lack of assumptions on the probability  $Q$ . Section 3 goes further and analyses the cases where the transport potential is unique up to additive constants. In such cases, under some assumption of regularity on the cost and on  $Q$ , the limit is not a supremum anymore, but simply a centered Gaussian random variable. The last section studies the semidiscrete O.T. in manifolds and gives a CLT not for the transport cost but for the solutions of the dual problem (3). We underline that if both probabilities are continuous and the space is not one dimensional, we cannot expect such type of central limit for the potentials, since, as commented before, the expected value of the estimation of the transport cost converges with rate  $O(n^{-\frac{1}{d}})$  and no longer  $o(n^{-\frac{1}{2}})$ . When the two samples are discrete, even if such a rate is  $o(n^{-\frac{1}{2}})$ , the lack of uniqueness of the dual problem does not allow to prove such type of problems. In consequence, the semidiscrete is the unique case where such results, for the potentials of the O.T. problem in general dimension, can be expected.

## 2 Central Limit Theorem

The goal is to prove a Central Limit theorem for the optimal transport cost in general Polish spaces  $\mathcal{X}, \mathcal{Y}$  for the semi-discrete problem. For that reason during the rest of the paper we will assume that  $P$  is supported in a finite set. Let  $\mathbb{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_m\} \subset \mathcal{X}$  be such that  $\mathbf{x}_i \neq \mathbf{x}_j$ , for  $i \neq j$ , and denote by  $\mathcal{P}(\mathbb{X})$  the set of probabilities defined in such set, i.e. any  $P \in \mathcal{P}(\mathbb{X})$  can be written as

$$P := \sum_{k=1}^m p_k \delta_{\mathbf{x}_k}, \text{ where } p_i > 0, \text{ for all } i = 1, \dots, m, \text{ and } \sum_{k=1}^m p_k = 1. \quad (4)$$

In consequence  $P$  is characterized by the vector  $\mathbf{p} = (p_1, \dots, p_m) \in \mathbb{R}^m$ . The following result shows that for the previous defined  $P$ , the optimal transport problem is equivalent to a optimization problem on the first coordinate of the

following functional over  $\mathbb{R}^m \times \mathbb{R}^m$  defined as

$$\mathcal{M} : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$$

$$(\mathbf{z}, \mathbf{p}) \mapsto \mathcal{M}(\mathbf{z}, \mathbf{p}) := \frac{1}{\|\mathbf{p}\|_1} \sum_{k=1}^m |p_k| z_k + \int \min_{i=1, \dots, m} \{c(\mathbf{y}, \mathbf{x}_i) - z_i\} dQ(\mathbf{y}), \quad (5)$$

where  $\|\mathbf{p}\|_1 = \sum_{k=1}^m |p_k|$ . The following result, which is a straightforward adaptation of Proposition 4.2 in [del Barrio and Loubes, 2019], shows that for the previous defined  $P$ , the optimal transport problem is equivalent to a optimization problem on the first coordinate of  $\mathcal{M}$ .

**Lemma 2.1.** *Let  $P \in \mathcal{P}(\mathbb{X})$ ,  $Q \in \mathcal{P}(\mathcal{Y})$  and  $c$  be a cost satisfying (2), then*

$$\mathcal{T}_c(P, Q) = \sup_{\mathbf{z} \in \mathbb{R}^m} \mathcal{M}(\mathbf{z}, \mathbf{p}). \quad (6)$$

**Remark 2.2.** *Note that if  $\varphi$  denotes an optimal transport potential from  $P$  to  $Q$  for the cost  $c$ , then*

$$\mathcal{T}_c(P, Q) = \mathcal{M}((\varphi(\mathbf{x}_1), \dots, \varphi(\mathbf{x}_m)), \mathbf{p}),$$

which makes the link between the optimal transport potentials and optimal values of (6) as  $\mathbf{z} = (\varphi(\mathbf{x}_1), \dots, \varphi(\mathbf{x}_m))$ .

Let  $X_1, \dots, X_n$  be a sequence of i.i.d. random variables with law  $P$ , since  $X_k \in \mathbb{X}$  for all  $k = 1, \dots, n$ , the empirical measure  $P_n := \frac{1}{n} \sum_{k=1}^n \delta_{X_k}$  belongs also to  $\mathcal{P}(\mathbb{X})$ . In consequence it can be written as  $P_n := \sum_{k=1}^m p_k^n \delta_{\mathbf{x}_k}$ , where  $p_1^n, \dots, p_m^n$  are real random variables such that  $p_i^n \geq 0$ , for all  $i = 1, \dots, m$ , and  $\sum_{k=1}^m p_k^n = 1$ . We want to study the weak limit of the sequence  $\{\sqrt{n}(\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q))\}_{n \in \mathbb{N}}$ . Consider a centered Gaussian vector,  $(X_1, \dots, X_m)$  with variance matrix where

$$\Sigma(\mathbf{p}) := \begin{bmatrix} p_1(1-p_1) & -p_1 p_2 & \cdots & -p_1 p_m \\ -p_2 p_1 & p_2(1-p_2) & \cdots & -p_2 p_m \\ \vdots & \vdots & \ddots & \vdots \\ -p_m p_1 & \cdots & p_m p_{m-1} & p_m(1-p_m) \end{bmatrix}. \quad (7)$$

and define the class of optimal values

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

The following result shows that the class of optimal values  $\text{Opt}_c(P, Q)$  is non-empty and the main theorem makes sense.

**Lemma 2.3.** *Let  $P \in \mathcal{P}(\mathbb{X})$ ,  $Q \in \mathcal{P}(\mathcal{Y})$  and  $c$  be such that (2) holds, then  $\text{Opt}_c(P, Q) \neq \{\emptyset\}$ .*

The proof of Lemma 2.3 is postponed to the appendix and is based on Lemma 2.6, described below and whose proof can be found also in the appendix. The following theorem shows that  $\{\sqrt{n}(\mathcal{T}_c(P_n, Q) - \mathcal{T}_c(P, Q))\}_{n \in \mathbb{N}}$  has a weak limit.

**Theorem 2.4.** *Let  $P \in \mathcal{P}(\mathbb{X})$ ,  $Q \in \mathcal{P}(\mathcal{Y})$  and  $c$  be such that it satisfies (2) and*

$$\int c(\mathbf{y}, \mathbf{x}_i) dQ(\mathbf{y}) < \infty, \quad \text{for all } i = 1, \dots, m. \quad (9)$$

Then

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

where  $(X_1, \dots, X_n) \sim \mathcal{N}(\mathbf{0}, \Sigma(\mathbf{p}))$  and  $\Sigma(\mathbf{p})$  is defined in (7).

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 the metric  $d$ . Then the value  $\mathcal{T}_{d^p}(P, Q)$  follows also the asymptotic behaviour described in Theorem 2.4. Using the classic delta-method, the following theorem establishes the asymptotic behavior of the well known  $p$ -Wasserstein distance  $\mathcal{W}_p(P, Q) = \sqrt[p]{\mathcal{T}_{d^p}(P, Q)}$ .

**Corollary 2.5.** *Let  $P \in \mathcal{P}(\mathbb{X})$  and  $Q \in \mathcal{P}(\mathcal{Z})$  be such that*

$$\int d(\mathbf{y}, \mathbf{x}_0) dQ(\mathbf{y}) < \infty, \quad \text{for some } \mathbf{x}_0 \in \mathcal{X}. \quad (10)$$

Then, for any  $p \geq 1$ , we have

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

where  $(X_1, \dots, X_n) \sim \mathcal{N}(\mathbf{0}, \Sigma(\mathbf{p}))$  and  $\Sigma(\mathbf{p})$  is defined in (7).

*Proof of Theorem 2.4.*

We have divided the proof in a sequence of lemmas, finally we will use the Delta-Method for Hadamard differentiable maps. Recall that a function  $f : \Theta \rightarrow \mathbb{R}$ , defined in an open set  $\Theta \subset \mathbb{R}^m$ , is said to be *Hadamard directionally differentiable* at  $\theta \in \Theta$  if there exists a function  $f'_\theta : \mathbb{R}^m \rightarrow \mathbb{R}$  such that

$$\frac{f(\theta + t_n \mathbf{h}_n) - f(\theta)}{t_n} \xrightarrow[n \rightarrow \infty]{} f'_\theta(\mathbf{h}), \quad \text{for all sequences } t_n \searrow 0 \text{ and } \mathbf{h}_n \rightarrow \mathbf{h}.$$

Let  $\Theta \subset \mathbb{R}^m$  be an open set,  $\theta \in \Theta$  and  $\{X_n\}_{n \in \mathbb{N}}$  be a sequence of random variables such that  $X_n : \Omega_n \rightarrow \Theta$  and  $r_n(X_n - \theta) \xrightarrow{w} X$  for some sequence  $r_n \rightarrow +\infty$  and some random element  $X$  that takes values in  $\mathbb{R}^m$ . If  $f : \Theta \rightarrow \mathbb{R}$  is Hadamard differentiable at  $\theta$  with derivative  $f'_\theta(\cdot) : \mathbb{R}^m \rightarrow \mathbb{R}$ , then Theorem 1 in

[Römisch, 2014], so-called delta-method, states that  $r_n(f(X_n) - f(\theta)) \xrightarrow{w} f'_\theta(X)$ . Since  $p_i > 0$  for all  $i = 1, \dots, m$ , it is enough to consider the hyperoctant

$$\mathcal{U}_+ := \{\mathbf{p} \in \mathbb{R}^m : p_i > 0, i = 1, \dots, m\}.$$

The aim is to prove that the map

$$\begin{aligned} \Gamma : \mathcal{U}_+ &\rightarrow \mathbb{R} \\ \mathbf{p} &\mapsto \sup_{\mathbf{z} \in \mathbb{R}^m} \mathcal{M}(\mathbf{z}, \mathbf{p}) \end{aligned} \quad (11)$$

is Hadamard differentiable. With that in mind, we start by proving the following technical lemma.

**Lemma 2.6.** *Let  $Q \in \mathcal{P}(\mathcal{Y})$  be such that (9) holds for the cost  $c$  which satisfies (2). Then the functional  $\mathcal{M}$  satisfies:*

1. *it is continuous in  $\mathbb{R}^m \times \mathcal{U}_+$ ,*
2. *it is continuously differentiable in the second variable over the set  $\mathcal{U}_+$ , moreover*

$$\nabla_{\mathbf{p}} \mathcal{M}(\mathbf{z}, \mathbf{p}) = \left( \frac{z_i \sum_{j=1}^m p_j - \sum_{j=1}^m z_j p_j}{\left( \sum_{j=1}^m p_j \right)^2} \right)_{i=1, \dots, m}, \quad \text{for all } \mathbf{z} \in \mathbb{R}^m,$$

3. *for all  $\mathbf{p}_0 \in \mathcal{U}_+$  there exists  $\alpha \in \mathbb{R}$  and  $\epsilon > 0$  satisfying that the set*

$$\text{lev}_\alpha \mathcal{M}(\cdot, \mathbf{p}) := \{\mathbf{z} \in \mathbb{R}^m : \mathcal{M}(\mathbf{z}, \mathbf{p}) \geq \alpha, z_1 = 0\}$$

*is compact and non empty, for all  $\mathbf{p}_0 \in \mathcal{U}_+$  such that  $|\mathbf{p}_0 - \mathbf{p}| < \epsilon$ .*

Note that Lemma 2.6 yields that the functional  $\mathcal{M}$  satisfies the hypothesis of Danskin's theorem, see for instance Theorem 4.13 in [Bonnans and Shapiro, 2000]. In consequence  $\Gamma$  is directionally differentiable. By proving that  $\Gamma$  is locally Lipschitz in  $\mathcal{U}_+$ , we obtain the following result.

**Lemma 2.7.** *The map  $\Gamma$  defined in (11) is Hadamard differentiable, with derivative*

$$\Gamma'_{\mathbf{p}}(\mathbf{q}) = \sup_{\mathbf{z} \in \text{Opt}_c(P, Q)} \sum_{i=1}^m \frac{z_i \sum_{j=1}^m p_j - \sum_{j=1}^m z_j p_j}{\left( \sum_{j=1}^m p_j \right)^2} q_i, \quad \text{for all } \mathbf{q} \in \mathbb{R}^m. \quad (12)$$

Note that since  $P = \sum_{j=1}^m p_j \delta_{\mathbf{x}_j}$  is a probability,  $\sum_{j=1}^m p_j = 1$  and (12) can be simplified:

$$\Gamma'_{\mathbf{p}}(\mathbf{q}) = \sum_{i=1}^m \left( z_i - \sum_{j=1}^m z_j p_j \right) q_i, \quad \text{for all } \mathbf{q} \in \mathbb{R}^m. \quad (13)$$



Then recall that the multivariate central limit theorem yields that

$$\sqrt{n}(\mathbf{p}_n - \mathbf{p}) \xrightarrow{w} \mathbf{X} = (X_1, \dots, X_m) \sim N(\mathbf{0}, \Sigma(\mathbf{p})),$$

where  $\Sigma(\mathbf{p})$  is defined in (7). The aforementioned Delta-Method for Hadamard differentiable maps and (13) yields that

$$\sqrt{n}(\Gamma(\mathbf{p}_n) - \Gamma(\mathbf{p})) \xrightarrow{w} \sup_{\mathbf{z} \in \text{Opt}_c(P, Q)} \sum_{i=1}^m \left( z_i - \sum_{j=1}^m z_j p_j \right) X_i.$$

Since  $\mathbf{z} \in \text{Opt}_c(P, Q)$  implies that  $\mathbf{z} - \mathbf{1} \sum_{j=1}^m z_j p_j \in \text{Opt}_c(P, Q)$ , then we have that

$$\sqrt{n}(\Gamma(\mathbf{p}_n) - \Gamma(\mathbf{p})) \xrightarrow{w} \sup_{\mathbf{z} \in \text{Opt}_c(P, Q)} \sum_{i=1}^m z_i X_i,$$

which concludes the proof.  $\square$

**Remark 2.8.** *Note that, while the idea of proving a central limit theorem for the optimal transport cost through the Hadamard differentiability properties of a suitable functional is not new at all (the same was done by [Sommerfeld and Munk, 2018] and [Tameling et al., 2019]), the functional used in other references only holds for finitely (or countably) supported  $Q \in \mathcal{P}(\mathbb{Y})$ , with  $\mathbb{Y} = \{y_1, \dots, y_l\}$ . In [Sommerfeld and Munk, 2018] the dual problem deals with the functional. More precisely*

$$\mathcal{G}(\mathbf{u}, \mathbf{v}, \mathbf{p}, \mathbf{q}) := \sum_{k=1}^m p_k u_k + \sum_{k=1}^l q_k v_k.$$

The optimization problem is formulated under the following constraints

$$\sup_{\mathbf{u}, \mathbf{v}} \mathcal{G}(\mathbf{u}, \mathbf{v}, \mathbf{p}, \mathbf{q}), \quad \text{s.t. } u_k + v_j \leq c(x_k, y_j), \quad \text{for } k \in \{1, \dots, m\} \text{ and } j \in \{1, \dots, l\}.$$

With regards to [Tameling et al., 2019], the approach is based on a direct analysis of the primal formulation which is also a constrained optimization problem. The definition of  $\mathcal{M}$  and Lemma 6 allow to use Danskin's theorem instead of constrained versions such as Theorem 4.24 in [Bonnans and Shapiro, 2000]). Moreover the division by the term  $\|\mathbf{p}\|_1$  in the definition of  $\mathcal{M}$  implies the Hadamard differentiability not only for a fixed direction, but in a general sense. This is a consequence of the existence of finite solutions of  $\sup_{\mathbf{z} \in \mathbb{R}^m} \mathcal{M}(\mathbf{z}, \mathbf{p})$ , for every  $\mathbf{p} \in \mathcal{U}_+$ . Moreover, the point 3 in Lemma 2.6 is no longer true without such regularization by  $\|\mathbf{p}\|_1$ , and in consequence the argument does not hold.

### 3 Asymptotic Gaussian distribution optimal transport cost.

The aim of this section is to show that in the case where  $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^d$  and under some regularity assumptions, enumerated as (A1)-(A3), the limit distribution is

Gaussian. Let  $Q \in \mathcal{P}(\mathbb{R}^d)$  be a probability measure uniformly 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

$$K(r, \theta, \mathbf{z}, \mathbf{p}) := \{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x} - \mathbf{p}| |\mathbf{z}| \cos(\theta/2) \leq \langle \mathbf{z}, \mathbf{x} - \mathbf{p} \rangle \leq r |\mathbf{z}| \}, \quad (14)$$

with vertex at  $\mathbf{p}$  on which  $h$  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

$$\mathcal{T}_c(P, Q) := \inf_T \int c(\mathbf{x}, T(\mathbf{x})) dP(\mathbf{x}), \quad \text{and} \quad T_{\#}P = Q. \quad (15)$$

The notation  $T_{\#}P$  represents the *Push-Forward* measure, it is the measure such that for each measurable set  $A$  we have  $T_{\#}P(A) := P(T^{-1}(A))$ . The solution of (15) is called *optimal transport map* from  $P$  to  $Q$ . Moreover it is defined as the unique Borel function satisfying

$$T(\mathbf{x}) = \mathbf{x} - \nabla h^*(\nabla \varphi(\mathbf{x})), \quad \text{where } \varphi \text{ solves (3)}. \quad (16)$$

Here  $h^*$  denotes the convex conjugate of  $h$ , see [Rockafellar, 1970]. Such uniqueness enabled [del Barrio et al., 2021] to deduce the uniqueness, under additive constants, of the solutions of  $\varphi$ . They assumed (A1)-(A3) to show that if two  $c$ -concave functions 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 support of  $Q$  is connected and with Lebesgue negligible boundary  $\ell_d(\partial \text{supp}(Q)) = 0$ , the uniqueness, up to additive constants, of the solutions of (3) holds. The proof of the main theorem is 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.** *Under the hypothesis of Theorem 3.2, if  $\tilde{\mathbf{z}}, \mathbf{z} \in \text{Opt}_c(P, Q)$ , then  $\tilde{\mathbf{z}} = \mathbf{z} - L\mathbf{1}$  for some constant  $L \in \mathbb{R}$ .*

The following theorem states, under the previous assumptions, that the limit distribution of Theorem 2.4 is the centered Gaussian variable  $\sum_{i=1}^m z_i X_i$ , where  $\mathbf{z} \in \text{Opt}_c(P, Q)$ . Note that Lemma 3.1 implies that the random variable  $\sum_{i=1}^m z_i X_i$  follows the same distribution independently of the chosen  $\mathbf{z} \in \text{Opt}_c(P, Q)$ , which is  $\mathcal{N}(0, \sigma^2(P, \mathbf{z}))$ , with

$$\sigma^2(P, \mathbf{z}) = \text{Var}\left(\sum_{i=1}^m z_i X_i\right) \quad \text{and} \quad (X_1, \dots, X_m) \sim \mathcal{N}(\mathbf{0}, \Sigma(\mathbf{p})), \quad (17)$$

where  $\Sigma(\mathbf{p})$  is defined in (7). 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) and*

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

Then

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

with  $\sigma^2(P, \mathbf{z})$  defined in (17) for  $\mathbf{z} \in \text{Opt}_c(P, Q)$ .

Since, in particular, the potential costs  $c_p = |\cdot|^p$ , for  $p > 0$ , 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 (4) and  $Q \in \mathcal{P}(\mathbb{R}^d)$  be such that  $Q \ll \ell_d$ , has finite moments of order  $p$  and its support is connected with Lebesgue negligible boundary. Then, for every  $p > 1$ , we have that*

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

and

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

with  $\sigma^2(P, \mathbf{z})$  defined in (17) for  $\mathbf{z} \in \text{Opt}_{c_p}(P, Q)$ .

*Proof of Theorem 3.2.*

Note that Theorem 2.4 states that

$$\sqrt{n} (\mathcal{T}_p(P_n, Q) - \mathcal{T}_p(P, Q)) \xrightarrow{w} \sup_{z \in \text{Opt}_c(P, Q)} \sum_{i=1}^m z_i X_i,$$

where  $(X_1, \dots, X_n) \sim \mathcal{N}(\mathbf{0}, \Sigma(\mathbf{p}))$ ,  $\Sigma(\mathbf{p})$  is defined in (7) and  $\text{Opt}_c(P, Q)$  in (8). Lemma 3.1 shows that such class is in fact a singleton, up to additive constants.  $\square$

Note that Corollary 3.3 is a particular case of Corollary 2.5 in the cases where the optimal transport potential is unique—the hypothesis of Theorem 3.2 hold—which is the reason why the case  $p = 1$  can not be considered. With regards to the other potential costs,  $p > 1$ , it is straightforward to see that the hypothesis (A1)-(A3) hold, see for instance [del Barrio et al., 2021] or [Gangbo and McCann, 1996].

## 4 A central Limit theorem for the potentials.

The aim of the section is to deduce a CLT for the potentials  $(\psi, \varphi)$  of the transport problem. Recall that it refers to all pairs solving (3). In the semidiscrete case the potentials are pairs formed by  $\mathbf{z} = (z_1, \dots, z_m) \in \text{Opt}_c(P, Q)$  and  $\varphi(\mathbf{y}) := \inf_{i=1, \dots, m} \{c(\mathbf{y}, \mathbf{x}_i) - z_i\}$ . Note that if  $(\psi, \varphi)$  solves (3) then  $(\psi + C, \varphi - C)$ , for all constant  $C$ . That makes necessary the 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  $\mathbf{1} = (1, \dots, 1)$ ,

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

where  $\mathcal{M}(\mathbf{z}, \mathbf{p})$  is defined in (5). The idea behind the proof of the main result of the section, Theorem 4.2, is to use classical results in  $M$ -estimation. In consequence, the strictly positiveness of the Hessian of the previously defined  $\Gamma$  is required. Such study of the second derivative have already been addressed in [Kitagawa et al., 2019], where they used a completely different notation, more related to the angle of mathematical analysis than the statistical. Consequently, this section changes slightly the notations of the previous ones. In order to make the readers work easier, we will try to be as consistent as possible with the notation adopted in the foregoing sections and, at the same time, to adapt it to the one proposed in [Kitagawa et al., 2019]. First we will assume that  $\mathcal{Y}$  is an open domain of a  $d$ -dimensional Riemannian manifold endowed with the volume measure  $\mathcal{V}$  and metric  $d$ . We will follow the point of view and notations presented in [Kitagawa et al., 2019]. For further details we refer to this paper and references therein.

Following the approach of [Kitagawa et al., 2019], let's assume the following assumptions on the cost:

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

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

where  $D_{\mathbf{y}}c$  denotes the partial derivative of  $c$  w.r.t. the second variable. For every  $i \in \{1, \dots, m\}$  there exists  $\Omega_i \subset \mathbb{R}^d$  open and convex set, and a  $\mathcal{C}^{1,1}$  diffeomorphism  $\exp_i^c : \Omega_i \rightarrow \Omega$  such that the functions

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

Here quasi-convex, according to [Kitagawa et al., 2019], means that for every  $\lambda \in \mathbb{R}^d$  the sets  $f_{i,j}^{-1}([-\infty, \lambda])$  are convex. Assumptions on the cost are not enough at all, it becomes necessary that the probability is supported in a *c-convex set*  $\Omega$ , which means that  $(\exp_i^c)^{-1}(\Omega)$  is convex, for every  $i = 1, \dots, m$ .

Formally, let  $\mathcal{Y} \subset \mathcal{M}$  be a compact  $c$ -convex set,  $P \in \mathcal{P}(\mathbb{X})$  be as in (4) and suppose that

$$Q \in \mathcal{P}(\mathcal{Y}) \text{ satisfies } Q \ll \mathcal{V} \text{ with density } q \in \mathcal{C}^{0,\alpha}(\mathcal{Y}). \quad (\text{Hol})$$

Recall that  $f \in \mathcal{C}^{0,\alpha}(\mathcal{Y})$  if there exist  $C > 0$  such that  $\|f(x) - f(y)\| \leq Cd(x, y)^\alpha$ . The last required assumption in [Kitagawa et al., 2019] is that  $Q$  satisfies a *Poincaré Wirtinger inequality with constant  $C_{PW}$* : a probability measure  $Q$  supported in a compact set  $\mathcal{Y} \subset \mathcal{M}$  satisfies a Poincaré Wirtinger inequality with constant  $C_{PW}$  if for every  $f \in \mathcal{C}^1(\mathcal{Y})$  we have that for  $Y \sim Q$

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

In order to clarify the feasibility of such assumptions, we will provide some insights on them at the end of the section.

[Kitagawa et al., 2019] proved the following assertions.

1. Under assumptions (Reg) and (Twist) the function  $\mathcal{M}(\cdot, \mathbf{p})$  is concave and differentiable with derivative

$$\nabla_{\mathbf{z}}\mathcal{M}(\mathbf{z}, \mathbf{p}) = (-Q(A_2(\mathbf{z})) + p_2, \dots, -Q(A_m(\mathbf{z})) + p_m),$$

where

$$A_k(\mathbf{z}) := \{\mathbf{y} \in \mathbb{R}^d : c(\mathbf{x}_k, \mathbf{y}) - z_k < c(\mathbf{x}_i, \mathbf{y}) - z_i, \text{ for all } i \neq k\}. \quad (18)$$

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

$$\begin{aligned} \frac{\partial^2\mathcal{M}}{\partial z_i\partial z_j}(\mathbf{z}, \mathbf{p}) &= \int_{A_k(\mathbf{z}) \cap A_k(\mathbf{z})} \frac{1}{|\nabla_{\mathbf{y}}c(\mathbf{x}_i, \mathbf{y}) - \nabla_{\mathbf{y}}c(\mathbf{x}_j, \mathbf{y})|} dQ(\mathbf{y}), \text{ if } i \neq j, \\ \frac{\partial^2\mathcal{M}}{\partial^2 z_i}(\mathbf{z}, \mathbf{p}) &= -\sum_{j \neq i} \frac{\partial^2\mathcal{M}}{\partial z_i\partial z_j}(\mathbf{z}, \mathbf{p}). \end{aligned} \quad (19)$$

3. Under assumptions (Reg),(Twist) and (QC), and if  $Q$  satisfies (PW), defined below, then there exists a constant  $C$  such that

$$\langle D_{\mathbf{z}}^2\mathcal{M}(\mathbf{z}, \mathbf{p})\mathbf{v}, \mathbf{v} \rangle \leq -C\epsilon^3|\mathbf{v}|^2, \text{ for all } \mathbf{z} \in \mathcal{K}^\epsilon \text{ and } \mathbf{v} \in \langle \mathbf{1} \rangle^\perp, \quad (20)$$

where

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

These previous results imply immediately the following Lemma.

**Lemma 4.1.** *Let  $\mathcal{Y} \subset \mathcal{M}$  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 (Hol) on  $Q$ , we have that the function  $\mathcal{M}_{\mathbf{p}}$  is strictly concave and twice continuously differentiable, with*

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

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

$$\langle D^2 \mathcal{M}_{\mathbf{p}}(\bar{\mathbf{z}}) \mathbf{v}, \mathbf{v} \rangle \leq -C \inf_{i=1, \dots, m} |p_i|^3 |\mathbf{v}|^2, \text{ for all } \mathbf{v} \in \langle \mathbf{1} \rangle^\perp. \quad (21)$$

*Proof.* Note that it only remains to prove that (21) holds. But it is a direct consequence of (20). Actually, since  $\bar{\mathbf{z}}$  is the unique  $\mathbf{z} \in \text{Opt}(P, Q)$  then  $Q(A_k(\bar{\mathbf{z}})) = p_k$  for  $k = 1, \dots, m$  and we can conclude.  $\square$

Then we can formulate the main theorem of the section which yields a central limit theorem for the empirical estimation of the potentials. We follow classical arguments of  $M$ -estimation by writing the function  $\mathcal{M}_{\mathbf{p}}(\mathbf{z})$  as  $E(g(X, \mathbf{z}))$  with  $X \sim P$  and  $g : \mathbb{X} \times \langle \mathbf{1} \rangle^\perp \rightarrow \mathbb{R}$  defined by

$$g(\mathbf{x}_k, \mathbf{z}) := z_k + \int \inf_{i=1, \dots, m} \{c(\mathbf{y}, \mathbf{x}_i) - z_i\} dQ(\mathbf{y}), \quad (22)$$

for each  $\mathbf{z} = (z_1, \dots, z_m)$ . The weak limit is a centered multivariate Gaussian distribution with covariate matrix, depending on the optimal  $\tilde{\mathbf{z}}$ , defined by the bilinear map

$$\Sigma(\tilde{\mathbf{z}}) = (D^2 \mathcal{M}_{\mathbf{p}}(\tilde{\mathbf{z}}))^{-1} A (D^2 \mathcal{M}_{\mathbf{p}}(\tilde{\mathbf{z}}))^{-1}, \text{ where } A = \sum_{i=1}^m p_i \nabla_{\tilde{\mathbf{z}}} g(\mathbf{x}_i, \mathbf{z}) \nabla_{\tilde{\mathbf{z}}} g(\mathbf{x}_i, \mathbf{z})^t. \quad (23)$$

**Theorem 4.2.** *Let  $\mathcal{Y} \subset \mathcal{M}$  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 (Hol) on  $Q$ , we have that*

$$\sqrt{n}(\hat{\mathbf{z}}_n - \tilde{\mathbf{z}}) \xrightarrow{w} N(\mathbf{0}, \Sigma(\tilde{\mathbf{z}})), \quad (24)$$

where  $\tilde{\mathbf{z}} \in \langle \mathbf{1} \rangle^\perp \cap \text{Opt}_c(P, Q)$  (resp.  $\hat{\mathbf{z}}_n \in \langle \mathbf{1} \rangle^\perp \cap \text{Opt}_c(P_n, Q)$ ), and  $\Sigma(\tilde{\mathbf{z}})$  is defined in (23).

*Proof.* Now let  $g : \mathbb{X} \times \langle \mathbf{1} \rangle^\perp \rightarrow \mathbb{R}$  be defined in (22). It satisfies that if  $X \sim P$  then  $E(g(X, \mathbf{z})) = \mathcal{M}_{\mathbf{p}}(\mathbf{z})$ . Lemma 4.1 implies in particular that:

- (i) the function  $\mathbf{z} \mapsto g(\mathbf{x}_k, \mathbf{z})$  is concave for every  $\mathbf{x}_k$ .
- (ii) There exists a unique

$$\tilde{\mathbf{z}} \in \arg \sup_{\mathbf{z} \in \langle \mathbf{1} \rangle^\perp} E(g(X, \mathbf{z})).$$

(iii) The empirical potential is defined as

$$\hat{\mathbf{z}}_n \in \arg \sup_{\mathbf{z} \in (\mathbf{1})^\perp} P_n(g(X, \mathbf{z})).$$

(iv) The function  $\mathbf{z} \mapsto E(g(X, \mathbf{z}))$  is twice differentiable in  $\tilde{\mathbf{z}}$  with strictly negative definite Hessian matrix  $D^2\mathcal{M}_{\mathbf{p}}(\tilde{\mathbf{z}})$ .

(v) For every  $\mathbf{z} \in \mathbb{R}^{m-1}$  we have that

$$|\nabla_{\mathbf{z}}g(\mathbf{x}_k, \mathbf{z})| \leq |(Q(A_2(\mathbf{z})) + 1, \dots, Q(A_m(\mathbf{z})) + 1)| \leq \sqrt{2(m-1)}.$$

Then all the assumptions of Corollary 2.2 in [Huang, 2007] are satisfied by the function  $g$ . In consequence we have the limit

$$\sqrt{n}(\hat{\mathbf{z}}_n - \tilde{\mathbf{z}}) \xrightarrow{w} N\left(\mathbf{0}, (D^2\mathcal{M}_{\mathbf{p}}(\tilde{\mathbf{z}}))^{-1} A (D^2\mathcal{M}_{\mathbf{p}}(\tilde{\mathbf{z}}))^{-1}\right), \quad (25)$$

where  $A = E(\nabla_{\mathbf{z}}g(X, \mathbf{z})\nabla_{\mathbf{z}}g(X, \mathbf{z})^t)$ . Note that computing  $A$  we obtain the expression (23).  $\square$

For  $\tilde{\mathbf{z}}$  defined as in Theorem 4.2, set

$$\varphi(\mathbf{y}) := \inf_{i=1, \dots, m} \{c(\mathbf{y}, \mathbf{x}_i) - \tilde{z}_i\} \quad (26)$$

and note that it is an optimal transport map from  $Q$  to  $P$ , set also the value  $i(y) \in \{1, \dots, m\}$  where the infimum of (26) is attained. As previously we can define their empirical counterparts

$$\varphi_n(\mathbf{y}) := \inf_{i=1, \dots, m} \{c(\mathbf{y}, \mathbf{x}_i) - \hat{z}_i\}, \quad (27)$$

which is an optimal transport map from  $Q$  to  $P_n$ , and  $i_n(y)$  the index where the infimum of (27) is attained. Then we have that

$$\sqrt{n}(\hat{z}_{i_n(y)} - \tilde{z}_{i_n(y)}) \leq \sqrt{n}(\varphi_n(\mathbf{y}) - \varphi(\mathbf{y})) \leq \sqrt{n}(\hat{z}_{i(y)} - \tilde{z}_{i(y)}). \quad (28)$$

We can take supremums over  $\mathbf{y}$  in both sides of (28) and derive that

$$\sqrt{n} \sup_{i=1, \dots, m} (\hat{z}_i - \tilde{z}_i) = \sqrt{n} \sup_{\mathbf{y} \in \mathcal{Y}} (\varphi_n(\mathbf{y}) - \varphi(\mathbf{y})).$$

By symmetry we have that

$$\sqrt{n} \sup_{i=1, \dots, m} |\hat{z}_i - \tilde{z}_i| = \sqrt{n} \sup_{\mathbf{y} \in \mathcal{Y}} |\varphi_n(\mathbf{y}) - \varphi(\mathbf{y})|,$$

which implies the following corollary.

**Corollary 4.3.** *Under the hypothesis of Theorem 4.2, for  $\varphi$  and  $\varphi_n$  defined in (4) and (??), we have that*

$$\sqrt{n} \sup_{\mathbf{y} \in \mathcal{Y}} |\varphi_n(\mathbf{y}) - \varphi(\mathbf{y})| \xrightarrow{w} \sup_{i=1, \dots, m} |N_i|,$$

where  $(N_1, \dots, N_m) \sim N(\mathbf{0}, \Sigma(\tilde{\mathbf{z}}))$ .

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

1. Note that if we consider  $\mathcal{M} = \mathbb{R}^d$  and the quadratic cost, then (Reg), (Twist) and (QC) are obviously satisfied, by taking the function  $\exp_j$  as the identity. Actually the map  $\mathbf{y} \mapsto |\mathbf{x}_j - \mathbf{y}|^2$  is  $\mathcal{C}^\infty(\mathbb{R}^d)$  and  $\mathbf{y} - \mathbf{x}_j$  is its derivative w.r.t.  $\mathbf{y}$ . Finally note that the function

$$\mathbb{R}^d \ni \mathbf{p} \mapsto |\mathbf{x}_i - \mathbf{p}|^2 - |\mathbf{x}_j - \mathbf{p}|^2 = |\mathbf{x}_i|^2 - |\mathbf{x}_j|^2 + \langle \mathbf{x}_j - \mathbf{x}_i, \mathbf{p} \rangle$$

is linear in  $\mathbf{p}$  and consequently quasi-convex.

2. 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  $\Omega$ . In [Rathmair, 2019], Lemma 1 claims that (PW) is equivalent to the bound of  $\inf_{t \in \mathbb{R}} E(|f(Y) - t|)$ , for every  $f \in \mathcal{C}^1(\mathcal{Y})$ . Let  $Y \sim Q$  be such that there exists a  $\mathcal{C}^1(\mathcal{Y})$  map  $T$  satisfying the relation  $T(U) = Y$ , where  $U$  follows a uniform distribution on a compact convex set  $A$ . Since  $f \circ T \in \mathcal{C}^1(A)$ , by the powerful result of [Acosta and Durán, 2004], there exists  $C_A > 0$  such that

$$\begin{aligned} \inf_{t \in \mathbb{R}} E(|f(Y) - t|) &= \inf_{t \in \mathbb{R}} E(|f(T(U)) - t|) \leq C_A E(|\nabla f(T(U))| \cdot \|T'(U)\|_2) \\ &\leq C_A \sup_{\mathbf{u} \in A} \|T'(\mathbf{u})\|_2 E(|\nabla f(T(U))|), \end{aligned}$$

where  $\|T'(U)\|_2$  denotes the matrix norm. We deduce that in such cases (PW) holds. Note that the existence of such map relies on the well known existence of continuously differentiable optimal transport maps, which is treated by Caffarelli's theory. We refer to the most recent work [Cordero-Erausquin and Figalli, 2019] and references therein. However, as pointed out in [Kitagawa et al., 2019], more general probabilities can satisfy that assumption such as radial functions on  $\mathbb{R}^d$  with density

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

Moreover the spherical uniform  $\mathbb{U}_d$ , used in [Hallin et al., 2021] to generalize the distribution function to general dimension, where we first choose the radius uniformly and then, independently, we choose a point in the sphere  $\mathbf{S}_{d-1}$ , also satisfies (PW). This can be proved by using previous reasoning with the function  $T(\mathbf{x}) = \mathbf{x}|\mathbf{x}|^{d-1}$ , which is continuously differentiable. But note that such probability measure does not satisfy the regularity condition of continuous density over a convex set, in consequence some additional work should be done which is left as a future work. In the same way as the regularity of the transport can be derived in the continuous case by a careful treatment of the Monge-Ampère equation, see [del Barrio et al., 2020], we conjecture that Theorem 4.2 could hold also in that discrete framework.



## 5 Appendix

*Proof of Lemma 2.3.*

Note that Lemma 2.6 yields that there exist  $\alpha > 0$  such that

$$\text{lev}_\alpha \mathcal{M}(\cdot, \mathbf{p}) := \{\mathbf{z} \in \mathbb{R}^m : \mathcal{M}(\mathbf{z}, \mathbf{p}) \geq \alpha, z_1 = 0\}$$

is non empty and compact. Lemma 2.6 also claims that  $\mathcal{M}$  is continuous, then

$$\sup_{\mathbf{z} \in \mathbb{R}^m} \mathcal{M}(\mathbf{z}, \mathbf{p}) = \sup_{\mathbf{z} \in \text{lev}_\alpha \mathcal{M}(\cdot, \mathbf{p})} \mathcal{M}(\mathbf{z}, \mathbf{p})$$

whose supremum is attained in some  $\mathbf{z} \in \text{lev}_\alpha \mathcal{M}(\cdot, \mathbf{p})$  by Weierstrass extreme value theorem.  $\square$

*Proof of Lemma 2.6.*

To prove the first point we realize that  $\mathcal{M} = F_1 + F_2$ , where

$$F_1((\mathbf{z}, \mathbf{p})) := \int \inf_{i=1, \dots, m} \{c(\mathbf{y} - \mathbf{x}_i) - z_i\} dQ(\mathbf{y}) \quad \text{and} \quad F_2((\mathbf{z}, \mathbf{p})) := \frac{1}{\|\mathbf{p}\|_1} \sum_{k=1}^m |p_k| z_k$$

are both continuous functions in  $\mathbb{R}^m \times \mathcal{U}_+$ . An easy computation shows the point 2. of Lemma 2.6. For the part 3, we set  $\mathbf{p}_0 \in \mathcal{U}_+$  and a sequence  $\{\mathbf{p}_n\}_{n \in \mathbb{N}}$  converging to  $\mathbf{p}_0$  then the measure  $P_n := \frac{1}{\sum_{j=1}^m p_j^n} \sum_{j=1}^m p_j^n \delta_{\mathbf{x}_j}$  converges weakly to  $P_0 = \frac{1}{\sum_{j=1}^m p_j^0} \sum_{j=1}^m p_j^0 \delta_{\mathbf{x}_j}$ . In consequence  $\mathcal{T}_c(P_n, Q) \rightarrow \mathcal{T}_c(P, Q) = \lambda > 0$ , when  $n \rightarrow \infty$ . This means that there exists some  $\epsilon > 0$  such that  $\sup_{\mathbf{z} \in \mathbb{R}^m} \mathcal{M}(\mathbf{z}, \mathbf{p}) \geq \alpha := \lambda/2$ , for all  $\mathbf{p}$  such that  $|\mathbf{p}_0 - \mathbf{p}| < \epsilon$ . Since  $\mathcal{U}_+$  is open, we can assume that  $\mathbb{B}_{2\epsilon}(\mathbf{p}_0) = \{\mathbf{q} \in \mathbb{R}^m : |\mathbf{p}_0 - \mathbf{q}| < 2\epsilon\} \subset \mathcal{U}_+$ . This means that

$$1 - \epsilon > \frac{p_j}{\|\mathbf{p}\|_1} > \epsilon, \quad \text{for all } \mathbf{p} \in \mathbb{B}_\epsilon(\mathbf{p}_0) \quad (29)$$

Therefore the set  $\text{lev}_\alpha \mathcal{M}(\cdot, \mathbf{p}) := \{\mathbf{z} \in \mathbb{R}^m : \mathcal{M}(\mathbf{z}, \mathbf{p}) \geq \alpha, z_1 = 0\}$  is not empty for every  $\mathbf{p} \in \mathbb{B}_\epsilon(\mathbf{p}_0)$ . The only point remaining concerns the compactness of

$\text{lev}_\alpha \mathcal{M}(\cdot, \mathbf{p})$ . Note that

$$\begin{aligned}
& \frac{1}{\|\mathbf{p}\|_1} \sum_{k=1}^m |p_k| z_k + \int \inf_{i=1, \dots, m} \{c(\mathbf{y} - \mathbf{x}_i) - z_i\} dQ(\mathbf{y}) \\
& \leq \frac{1}{\|\mathbf{p}\|_1} \sum_{k=1}^m p_k z_k + \inf_{i=1, \dots, m} \left\{ \int c(\mathbf{y} - \mathbf{x}_i) dQ(\mathbf{y}) - z_i \right\} \\
& \leq \frac{1}{\|\mathbf{p}\|_1} \sum_{k=1}^m p_k z_k + \sup_{i=1, \dots, m} \int c(\mathbf{y} - \mathbf{x}_i) dQ(\mathbf{y}) - \sup_{i=1, \dots, m} z_i \\
& = \frac{1}{\|\mathbf{p}\|_1} \sum_{k=1}^m p_k \left( z_k - \sup_{i=1, \dots, m} z_i \right) + \sup_{i=1, \dots, m} \int c(\mathbf{y} - \mathbf{x}_i) dQ(\mathbf{y}) \\
& = \frac{1}{\|\mathbf{p}\|_1} \sum_{k=2}^m p_k \left( z_k - \sup_{i=1, \dots, m} z_i \right) + \sup_{i=1, \dots, m} \int c(\mathbf{y} - \mathbf{x}_i) dQ(\mathbf{y}) + \frac{1}{\|\mathbf{p}\|_1} p_1 \left( - \sup_{i=1, \dots, m} z_i \right) \\
& \leq K - \frac{p_1}{\|\mathbf{p}\|_1} \left( \sup_{i=1, \dots, m} z_i \right),
\end{aligned}$$

where  $K = \sup_{i=1, \dots, m} \int c(\mathbf{y} - \mathbf{x}_i) dQ(\mathbf{y}) \in \mathbb{R}$ . From (29) we deduce that  $\sup_k z_k$  has to be bounded  $\sup_k z_k \leq M$ . The task now becomes to show that  $\inf_k z_k$  has to be bounded. Since  $z_1 = 0$  then

$$\begin{aligned}
\mathcal{M}(\mathbf{z}, \mathbf{p}) & \leq \inf_j \frac{p_j}{\|\mathbf{p}\|_1} \inf_k z_k + \sup_j \frac{p_j}{\|\mathbf{p}\|_1} \sup_k z_k + \sup_i \int c(\mathbf{y} - \mathbf{x}_i) dQ(\mathbf{y}) - z_1 \\
& \leq M + K + \inf_j \frac{p_j}{\|\mathbf{p}\|_1} \inf_k z_k.
\end{aligned}$$

Finally (29) implies that  $\inf_k z_k \geq M'$ . In consequence the set  $\text{lev}_\alpha \mathcal{M}(\cdot, \mathbf{p})$  is bounded, it is clear that also closed, hence compact.  $\square$

*Proof of Lemma 2.7.*

Danskin's theorem (Theorem 4.13 in [Bonnans and Shapiro, 2000]) considers an optimization problem

$$G(u) = \max_{x \in X} g(x, u) \quad \text{s.t. } x \in \Phi,$$

where  $u \in U$  and it is a Banach space with the norm  $|\cdot|$ ,  $X$  is a topological Hausdorff space,  $\Phi \subset X$  is nonempty and closed and  $g$  is continuous. If the function  $g(x, \cdot)$  is differentiable,  $g$  and its derivative w.r.t.  $u$  in the direction  $u$   $D_u g(x, u)$  are continuous in  $X \times U$  and for all  $u_0 \in U$  there exists  $\alpha \in \mathbb{R}$  and  $\epsilon > 0$  satisfying that the set

$$\text{lev}_\alpha \mathcal{M}(\cdot, \mathbf{p}) := \{x \in \Phi : f(x, u) \geq \alpha\}$$

is compact and non empty, for all  $u$  such that  $|u - u_0| < \epsilon$ . Then the function  $G$  is Fréchet directionally differentiable in  $u_0$ , which means that, for every  $h \in U$ , the limit

$$G'_{u_0}(h) = \lim_{t \downarrow 0} \frac{G(u_0 + th) - G(u_0)}{t}$$

exists and satisfies that

$$G(u_0 + h) = G(u_0) + G'_{u_0}(h) + o(|h|).$$

Note that Lemma 2.6 yields that the functional  $\mathcal{M}$  satisfies the assumptions of Danskin's theorem in  $\mathbb{R}^d \times \mathcal{U}_+$ , but  $\mathcal{U}_+$  is not a Banach space, consequently some additional work has to be done. Set  $\mathbf{p} \in \mathcal{U}_+$  and note that there exists some  $\epsilon > 0$  such that  $\overline{\mathbb{B}_\epsilon(\mathbf{p})} \subset \mathcal{U}_+$ . The function  $\mathcal{M}$  restricted to such set  $\mathcal{M}|_{\mathbb{R}^m \times \overline{\mathbb{B}_\epsilon(\mathbf{p})}}$  is a continuous function in  $\mathbb{R}^m \times \overline{\mathbb{B}_\epsilon(\mathbf{p})}$ , then we can apply Theorem 1 in [Whitney, 1934] to conclude that there exists a function  $\tilde{\mathcal{M}} : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$  such that

- $\tilde{\mathcal{M}} = \mathcal{M}$  in  $\mathbb{R}^m \times \overline{\mathbb{B}_\epsilon(\mathbf{p})}$ ,
- $\tilde{\mathcal{M}}$  is analytic in  $\mathbb{R}^m \times \mathbb{R}^m \setminus \left( \mathbb{R}^m \times \overline{\mathbb{B}_\epsilon(\mathbf{p})} \right)$ .

Note that the equality

$$\Gamma(\mathbf{p}) = \sup_{\mathbf{z} \in \mathbb{R}^m} \mathcal{M}(\mathbf{z}, \mathbf{p}) = \sup_{\mathbf{z} \in \mathbb{R}^m} \tilde{\mathcal{M}}(\mathbf{z}, \mathbf{p})$$

implies that we can apply Danskin's theorem to  $\tilde{\mathcal{M}}$  and in consequence  $\Gamma$  is Fréchet directionally differentiable in  $\mathbf{p}$ . We claim that  $\Gamma$  is locally Lipschitz. To prove it recall that Lemma 2.3 yields that there exist  $\mathbf{z}$  such that  $\mathcal{M}(\mathbf{z}, \mathbf{p}) = \mathcal{T}_c(P, Q)$  and  $z_1 = 0$ . Lemma 2.6 claims also that there exists  $\alpha \in \mathbb{R}$  and  $\epsilon > 0$  such  $\text{lev}_\alpha \mathcal{M}(\cdot, \mathbf{p})$  is non empty and compact for all  $\mathbf{p}' \in \mathbb{B}_\epsilon(\mathbf{p})$ . Then for all  $\mathbf{p}' \in \mathbb{B}_\epsilon(\mathbf{p})$  and all  $\mathbf{z}'$  solving  $\mathcal{M}(\mathbf{z}', \mathbf{p}') = \mathcal{T}_c(P', Q)$  and  $z'_1 = 0$ , we have that  $|z'_1| \leq M$ . Moreover such  $\epsilon$  can be chosen small enough in order to have  $\mathbb{B}_{2\epsilon}(\mathbf{p}) \subset \mathcal{U}_+$ . In consequence we have that

$$\begin{aligned} \Gamma(\mathbf{p}) - \Gamma(\mathbf{p}') &\leq \mathcal{M}(\mathbf{z}, \mathbf{p}) - \mathcal{M}(\mathbf{z}, \mathbf{p}') = \sum_{j=1}^m \left( \frac{p_j}{|\mathbf{p}|_1} - \frac{p'_j}{|\mathbf{p}'|_1} \right) z_j \\ &\leq M \sum_{j=1}^m \left| \frac{p_j}{|\mathbf{p}|_1} - \frac{p'_j}{|\mathbf{p}'|_1} \right| = M |F(\mathbf{p}) - F(\mathbf{p}')|_1, \end{aligned}$$

where  $F(\mathbf{p}) := \left( \frac{p_1}{|\mathbf{p}|_1}, \dots, \frac{p_m}{|\mathbf{p}|_1} \right)$ . We remark that the reverse inequality holds exactly with the same arguments. Then it suffices to prove that  $F$  is Lipschitz in  $\mathbb{B}_\epsilon(\mathbf{p})$ . It can be seen by computing for any  $\mathbf{p}' \in \mathbb{B}_\epsilon(\mathbf{p})$  the derivative

$$F'(\mathbf{p}') = \frac{1}{\left( \sum_{i=1}^m p'_i \right)^2} \begin{bmatrix} -\sum_{i \neq 1} p'_i & -1 & \cdots & -1 \\ -1 & -\sum_{i \neq 2} p'_i & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & \cdots & -1 & -\sum_{i \neq m} p'_i \end{bmatrix},$$

and checking that the matrix norm  $\|F'(\mathbf{p}')\|_1 \leq K := \frac{m-1+|\mathbf{p}|_1+\epsilon}{(|\mathbf{p}|_1-\epsilon)^2}$ . Note that this inequality holds since  $\epsilon$  is small enough to make  $|\mathbf{p}|_1 - \epsilon > 0$ . Therefore we

have that

$$|\Gamma(\mathbf{p}) - \Gamma(\mathbf{p}')| \leq MK|\mathbf{p} - \mathbf{p}'|_1 \leq \sqrt{m}MK|\mathbf{p} - \mathbf{p}'|, \quad \text{for all } \mathbf{p}' \in \mathbb{B}_\epsilon(\mathbf{p}).$$

Then  $\Gamma$  is locally Lipschitz and directionally differentiable in  $\mathcal{U}_+$ . By Proposition 2.49 in [Bonnans and Shapiro, 2000], we conclude that it is differentiable at  $\mathbf{p}$  in the Hadamard sense.  $\square$

*Proof of Lemma 3.1.*

Set  $\tilde{\mathbf{z}}, \mathbf{z} \in \text{Opt}(P, Q)$ . Lemma 2.1 implies that the  $c$ -concave functions  $\psi(\mathbf{y}) := \inf_{i=1, \dots, m} \{c(\mathbf{x}_i, \mathbf{y}) - z_i\}$  and  $\tilde{\psi}(\mathbf{y}) := \inf_{i=1, \dots, m} \{c(\mathbf{x}_i, \mathbf{y}) - \tilde{z}_i\}$  are both solutions of the dual optimal transport problem from  $Q$  to  $P$ . Moreover, from optimal transport theory, the map  $\psi^c$  (resp.  $\tilde{\psi}^c$ ) satisfies that  $\varphi(\mathbf{x}_j) = z_j$  (resp.  $\tilde{\varphi}(\mathbf{x}_j) = \tilde{z}_j$ ). Hence Theorem 2.6 in [del Barrio et al., 2021] implies that  $\psi = \tilde{\psi} + L$  holds in the interior of the support of  $Q$ . Since

$$\tilde{z}_i := \tilde{\psi}^c(\mathbf{x}_i) = \psi^c(\mathbf{x}_i) - L = z_i - L,$$

we have that  $\tilde{\mathbf{z}} = \mathbf{z} - L$  for some constant  $L \in \mathbb{R}$ .  $\square$

## References

- [Acosta and Durán, 2004] Acosta, G. and Durán, R. G. (2004). An optimal poincaré inequality in  $\ell_1$  for convex domains. *Proceedings of the American Mathematical Society*, 132(1):195–202.
- [Bonnans and Shapiro, 2000] Bonnans, J. F. and Shapiro, A. (2000). Perturbation analysis of optimization problems. In *Springer Series in Operations Research*.
- [Cordero-Erausquin and Figalli, 2019] Cordero-Erausquin, D. and Figalli, A. (2019). Regularity of monotone transport maps between unbounded domains.
- [de Goes et al., 2012] de Goes, F., Breeden, K., Ostromoukhov, V., and Desbrun, M. (2012). Blue noise through optimal transport. *ACM Trans. Graph.*, 31(6).
- [del Barrio et al., 2020] 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. *Journal of Multivariate Analysis*, 180:104671.
- [del Barrio et al., 2021] del Barrio, E., González-Sanz, A., and Loubes, J.-M. (2021). Central limit theorems for general transportation costs.
- [del Barrio et al., 2019] del Barrio, E., Gordaliza, P., and Loubes, J.-M. (2019). A central limit theorem for  $L_p$  transportation cost on the real line with application to fairness assessment in machine learning. *Information and Inference: A Journal of the IMA*, 8(4):817–849.

- [del Barrio and Loubes, 2019] del Barrio, E. and Loubes, J.-M. (2019). Central limit theorems for empirical transportation cost in general dimension. *The Annals of Probability*, 47(2):926 – 951.
- [Fournier and Guillin, 2015] Fournier, N. and Guillin, A. (2015). On the rate of convergence in wasserstein distance of the empirical measure. *Probab. Theory Relat. Fields*, 162:707–738.
- [Gallouët and Mérigot, 2018] Gallouët, T. and Mérigot, Q. (2018). A lagrangian scheme à la brenier for the incompressible euler equations. *Found Comput Math*, 18:835–865.
- [Gangbo and McCann, 1996] Gangbo, W. and McCann, R. J. (1996). The geometry of optimal transportation. *Acta Math.*, 177(2):113–161.
- [Gayraud et al., 2017] 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*, page 6.
- [Gordaliza et al., 2019] Gordaliza, P., del Barrio, E., Fabrice, G., and Loubes, J.-M. (2019). Obtaining fairness using optimal transport theory. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 2357–2365. PMLR.
- [Hallin et al., 2021] 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. *The Annals of Statistics*, 49(2):1139 – 1165.
- [Hartmann and Schuhmacher, 2020] 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.
- [Huang, 2007] Huang, J. (2007). Central limit theorems for m-estimates.
- [Jiang et al., 2020] Jiang, R., Pacchiano, A., Stepleton, T., Jiang, H., and Chiappa, S. (2020). Wasserstein fair classification. In *Uncertainty in Artificial Intelligence*, pages 862–872. PMLR.
- [Kitagawa et al., 2019] 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.
- [Lévy et al., 2020] Lévy, B., Mohayaei, R., and von Haeubler, S. (2020). A fast semi-discrete optimal transport algorithm for a unique reconstruction of the early universe.
- [Meyron, 2019] Meyron, J. (2019). Initialization procedures for discrete and semi-discrete optimal transport. *Computer-Aided Design*, 115:13–22.

- [Rathmair, 2019] Rathmair, M. (2019). On how poincaré inequalities imply weighted ones. *Monatsh Math*, 118:753–763.
- [Rockafellar, 1970] Rockafellar, R. T. (1970). *Convex analysis*. Number 28. Princeton university press.
- [Römisch, 2014] Römisch, W. (2014). *Delta Method, Infinite Dimensional*. American Cancer Society.
- [Shen et al., 2018] Shen, J., Qu, Y., Zhang, W., and Yu, Y. (2018). Wasserstein distance guided representation learning for domain adaptation.
- [Sommerfeld and Munk, 2018] Sommerfeld, M. and Munk, A. (2018). Inference for empirical wasserstein distances on finite spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):219–238.
- [Taming et al., 2019] Taming, C., Sommerfeld, M., and Munk, A. (2019). Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications. *Ann. Appl. Probab.*, 29(5):2744–2781.
- [Verdinelli and Wasserman, 2019] Verdinelli, I. and Wasserman, L. (2019). Hybrid Wasserstein distance and fast distribution clustering. *Electronic Journal of Statistics*, 13(2):5088 – 5119.
- [Villani, 2003] Villani, C. (2003). *Topics in optimal transportation*. Number 58 in Graduate studies in mathematics. American Mathematical Soc.
- [Villani, 2008] Villani, C. (2008). *Optimal Transport: Old and New*. Number 338 in Grundlehren der mathematischen Wissenschaften. Springer, Berlin. OCLC: ocn244421231.
- [Whitney, 1934] Whitney, H. (1934). Analytic extensions of differentiable functions defined in closed sets. *Transactions of the American Mathematical Society*, 36(1):63–89.