



**HAL**  
open science

# Convergence of asymptotic costs for random Euclidean matching problems

Michael Goldman, Dario Trevisan

► **To cite this version:**

Michael Goldman, Dario Trevisan. Convergence of asymptotic costs for random Euclidean matching problems. Probability and Mathematical Physics, In press. hal-02932968v2

**HAL Id: hal-02932968**

**<https://hal.science/hal-02932968v2>**

Submitted on 4 Dec 2020

**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.

# CONVERGENCE OF ASYMPTOTIC COSTS FOR RANDOM EUCLIDEAN MATCHING PROBLEMS

MICHAEL GOLDMAN AND DARIO TREVISAN

ABSTRACT. We investigate the average minimum cost of a bipartite matching between two samples of  $n$  independent random points uniformly distributed on a unit cube in  $d \geq 3$  dimensions, where the matching cost between two points is given by any power  $p \geq 1$  of their Euclidean distance. As  $n$  grows, we prove convergence, after a suitable renormalization, towards a finite and positive constant. We also consider the analogous problem of optimal transport between  $n$  points and the uniform measure. The proofs combine sub-additivity inequalities with a PDE ansatz similar to the one proposed in the context of the matching problem in two dimensions and later extended to obtain upper bounds in higher dimensions.

## 1. INTRODUCTION

The aim of this paper is to extend the results of [12, 9, 4, 11] on the existence of the thermodynamic limit for some random optimal matching problems. Because of their relations to computer science, statistical physics and quantization of measures, optimal matching problems have been the subject of intense research both from the mathematical and physical communities. We refer for instance to [29, 27, 10] for more details in particular regarding the vast literature.

Probably the simplest and most studied variant of these problems is the bipartite (or Euclidean bipartite) matching on the unit cube in  $d$  dimensions. Given  $p \geq 1$  and two independent families of i.i.d. random variables  $(X_i)_{i \geq 1}$  and  $(Y_i)_{i \geq 1}$  with common law the uniform (Lebesgue) measure on  $[0, 1]^d$ , the problem is to understand the behavior for large  $n$  of

$$\mathbb{E} \left[ \frac{1}{n} \min_{\pi} \sum_{i=1}^n |X_i - Y_{\pi(i)}|^p \right],$$

where the minimum is taken over all permutations  $\pi$  of  $\{1, \dots, n\}$ . It is by now well-known, see [1, 4, 6, 21] that for<sup>1</sup>  $n \gg 1$  (see [13] for some non-asymptotic bounds)

$$\mathbb{E} \left[ \frac{1}{n} \min_{\pi} \sum_{i=1}^n |X_i - Y_{\pi(i)}|^p \right] \sim \begin{cases} n^{-\frac{p}{2}} & \text{for } d = 1 \\ \left( \frac{\log n}{n} \right)^{\frac{p}{2}} & \text{for } d = 2 \\ n^{-\frac{p}{d}} & \text{for } d \geq 3. \end{cases} \quad (1.1)$$

Let us point out that while the case  $d \geq 3$ ,  $p \geq d/2$  is not explicitly covered in the literature, the proof of [21] clearly extends to any  $p \neq 2$  (see also [7]). Our main result is the following:

---

2010 *Mathematics Subject Classification.* 60D05, 90C05, 39B62, 60F25, 35J05.

*Key words and phrases.* Matching problem, optimal transport, geometric probability.

<sup>1</sup>The notation  $A \ll 1$ , which we only use in assumptions, means that there exists an  $\varepsilon > 0$  only depending on the dimension  $d$  and on  $p \geq 1$ , such that if  $A \leq \varepsilon$  then the conclusion holds. Similarly, the notation  $A \lesssim B$ , which we use in output statements, means that there exists a global constant  $C > 0$  depending on the dimension  $d$  and on  $p \geq 1$  such that  $A \leq CB$ . We write  $A \sim B$  if both  $A \lesssim B$  and  $B \lesssim A$ .

**Theorem 1.1.** *For every  $d \geq 3$  and  $p \geq 1$ , there exists a constant  $f_\infty^{bi} = f_\infty^{bi}(p, d) > 0$  such that*

$$\lim_{n \rightarrow \infty} n^{\frac{p}{d}} \mathbb{E} \left[ \frac{1}{n} \min_{\pi} \sum_{i=1}^n |X_i - Y_{\pi(i)}|^p \right] = f_\infty^{bi}. \quad (1.2)$$

This extends earlier results of [12, 9, 4, 11] where the same conclusion was obtained under the more restrictive condition  $p < d/2$ . See also [26] for bounds on  $f_\infty^{bi}(1, d)$  as  $d$  becomes large. As in the previously quoted papers, our proof is based on a sub-additivity argument and makes use of the classical observation that, thanks to the Birkhoff-Von Neumann Theorem, the bi-partite matching problem is actually an optimal transport problem. Indeed, if  $\mu = \sum_{i=1}^n \delta_{X_i}$  and  $\lambda = \sum_{i=1}^n \delta_{Y_i}$  are the associated empirical measures, then

$$\min_{\pi} \sum_{i=1}^n |X_i - Y_{\pi(i)}|^p = W_p^p(\mu, \lambda),$$

where  $W_p$  denotes the Wasserstein distance of order  $p$  (see [28, 25]). However, the papers [9, 4, 11] rely then upon combinatorial arguments, and in fact their results apply to a larger class of random optimization problems, while [12] strongly uses the dual formulation of optimal transport, which in the case  $p = 1$  is quite specific, since it becomes a maximization over the set of 1-Lipschitz functions.

The optimal transport point of view allows us to treat the defect in sub-additivity as a defect in local distribution of mass rather than a defect in local distribution of points. More precisely, even if  $\mu([0, 1]^d) = \lambda([0, 1]^d)$  it is in general not true that for a given partition of  $[0, 1]^d$  in sub-cubes  $Q_i$ ,  $\mu(Q_i) = \lambda(Q_i)$ . Therefore, in order to use sub-additivity, one needs to relax the definition of the matching problem to take into account this disparity. In [9, 4, 11] this is done by requiring that as many points as possible are matched. Here we allow instead for varying weights. That is, for  $\mu$  and  $\lambda$  containing potentially a different number of points, we consider the problem

$$\mathbb{E} \left[ W_p^p \left( \frac{\mu}{\mu([0, 1]^d)}, \frac{\lambda}{\lambda([0, 1]^d)} \right) \right].$$

The main sub-additivity argument for this quantity is contained in Lemma 3.1. In order to estimate the error in sub-additivity, we then use in Lemma 3.4 a PDE ansatz similar to the one proposed in the context of the matching problem in [10] and then used in [3] (see also [21, 17, 5]) to show that when  $d = 2$

$$\lim_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E} \left[ \frac{1}{n} W_2^2(\mu, \lambda) \right] = \frac{1}{2\pi}.$$

Notice however that our use of this linearization ansatz is quite different from [10, 3]. Indeed, for us the main contribution to the transportation cost is given by the transportation at the smaller scales and linearization is only used to estimate the higher order error term. On the other hand, in [10, 3] (see also [2]), the main contribution to the cost is given by the linearized i.e.  $H^{-1}$ , cost. This is somewhat in line with the prediction by [10] that for  $d \geq 3$ , the first order contribution to the Wasserstein cost is *not* given by the linearized problem while higher order corrections are. In any case, we give in Proposition 5.3 an alternative argument to estimate the error term without relying on the PDE ansatz. There we use instead an elementary comparison argument with one-dimensional transport plans. We included both proofs since we believe that they could prove useful in other contexts.

Let us make a few more comments on the proof of Theorem 1.1. As in [12, 9, 4], the proof of (1.2) is actually first done on a variant of the problem where the number of points follows a Poisson distribution instead of being deterministic. This is due to the fact that the restriction of a Poisson point process is again a Poisson point process. For this variant of the problem, rather than working on a fixed cube  $[0, 1]^d$  with an increasing density of

points, we prefer to make a blow-up at a scale  $L = n^{1/d}$  and consider in Theorem 5.1, a fixed Poisson point process of intensity 1 on  $\mathbb{R}^d$  but restricted to cubes  $Q_L = [0, L]^d$  with  $L \gg 1$  (hence the terminology thermodynamic limit). We believe that the sub-additivity argument is slightly clearer in these variables (a similar rescaling is actually implicitly used in [9, 4]). This setting is somewhat reminiscent of [18], where super-additivity is used to construct an optimal coupling between the Poisson point process and the Lebesgue measure on  $\mathbb{R}^d$ . In order to pass from the Poisson version of the problem to the deterministic one, we prove a general *de-Poissonization* result in Proposition 6.1 which can hopefully be useful in other contexts.

Besides the bipartite matching we also treat in Theorem 4.1 and Theorem 6.2 the case of the matching to the reference measure. We actually treat this problem first since the proof is a bit simpler. Indeed, while the general scheme of the proof is identical to the bipartite case, the PDE ansatz used in Lemma 3.4 works well for “regular” measures and a more delicate argument is required for the bipartite matching. Notice that by Jensen and triangle inequalities, (1.1) also holds for the matching to the reference measure.

We point out that in [12, 4, 11], it is more generally proven that if the points  $X_i$  and  $Y_i$  have a common law  $\mu$  supported in  $[0, 1]^d$  instead of the Lebesgue measure (for measures with unbounded support a condition on the moments is required), then for  $1 \leq p < d/2$  and  $d \geq 3$ ,

$$\limsup_{n \rightarrow \infty} n^{\frac{p}{d}} \mathbb{E} \left[ \frac{1}{n} \min_{\pi} \sum_{i=1}^n |X_i - Y_{\pi(i)}|^p \right] \leq f_{\infty}^{\text{bi}} \int_{[0,1]^d} \left( \frac{d\mu}{dx} \right)^{1-\frac{p}{d}}.$$

However, when  $p > d/2$  and without additional assumptions on  $\mu$ , the asymptotic rate may be different and thus this inequality may fail, see e.g. [13]. Positive results for specific densities can be obtained nonetheless. For instance, it is proven in [22] that for the standard Gaussian measure  $\mu$  on  $\mathbb{R}^d$ ,  $d \geq 3$ , the asymptotic bound

$$n^{\frac{p}{d}} \mathbb{E} \left[ \frac{1}{n} \min_{\pi} \sum_{i=1}^n |X_i - Y_{\pi(i)}|^p \right] \sim 1$$

holds true also for  $d/2 \leq p < d$ .

Finally, we notice that usual results on concentration of measure allow us to improve from convergence of the expectations to strong convergence. However, we are able to cover only the case  $1 \leq p < d$ , see Remark 6.5.

The plan of the paper is the following. In Section 2, we fix some notation and recall basic moment and concentration bounds for Poisson random variables. In Section 3, we state and prove our two main lemmas namely the sub-additivity estimate Lemma 3.1 and the error estimate Lemma 3.4. In Section 4, we then prove the existence of the thermodynamic limit for the matching problem of a Poisson point process to the reference measure. The analog result for the bipartite matching between two Poisson point processes is obtained in Section 5. Finally, in Section 6 we pass from the Poissonized problem to the deterministic one and discuss stronger convergence results.

## 2. NOTATION AND PRELIMINARY RESULTS

We use the notation  $|A|$  for the Lebesgue measure of a Borel set  $A \subseteq \mathbb{R}^d$ , and  $\int_A f$  for the Lebesgue integral of a function  $f$  on  $A$ . For  $L > 0$ , we let  $Q_L = [0, L]^d$ . We denote by  $|p|$  the Euclidean norm of a vector  $p \in \mathbb{R}^N$ . For a function  $\phi$ , we use the notation  $\nabla \phi$  for the gradient,  $\nabla \cdot \phi$  for the divergence,  $\Delta \phi$  for the Laplacian and  $\nabla^2 \phi$  for the Hessian.

**2.1. Optimal transport.** In this section we introduce some notation for the Wasserstein distance and recall few simple properties that will be used throughout. Proofs can be found in any of the monographs [28, 25, 24] with expositions of theory of optimal transport, from different perspectives.

Given  $p \geq 1$ , a Borel subset  $\Omega \subseteq \mathbb{R}^d$  and two positive Borel measures  $\mu, \lambda$  with  $\mu(\Omega) = \lambda(\Omega) \in (0, \infty)$  and finite  $p$ -th moments, the Wasserstein distance of order  $p \geq 1$  between  $\mu$  and  $\lambda$  is defined as the quantity

$$W_p(\mu, \lambda) = \left( \min_{\pi \in \mathcal{C}(\mu, \lambda)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p d\pi(x, y) \right)^{\frac{1}{p}},$$

where  $\mathcal{C}(\mu, \lambda)$  is the set of couplings between  $\mu$  and  $\lambda$ . Moreover, if  $\mu(\Omega) = \lambda(\Omega) = 0$ , we define  $W_p(\mu, \lambda) = 0$ , while if  $\mu(\Omega) \neq \lambda(\Omega)$ , we let  $W_p(\mu, \lambda) = \infty$ .

Let us recall that since  $W_p$  is a distance, we have the triangle inequality

$$W_p(\mu, \nu) \leq W_p(\mu, \lambda) + W_p(\nu, \lambda). \quad (2.1)$$

We will also use the classical sub-additivity inequality

$$W_p^p \left( \sum_i \mu_i, \sum_i \lambda_i \right) \leq \sum_i W_p^p(\mu_i, \lambda_i), \quad (2.2)$$

for a finite set of positive measures  $\mu_i, \lambda_i$ . This follows from the observation that if  $\pi_i \in \mathcal{C}(\mu_i, \lambda_i)$ , then  $\sum_i \pi_i \in \mathcal{C}(\sum_i \mu_i, \sum_i \lambda_i)$ .

**Remark 2.1.** In fact, our results deal with the *transportation cost*  $W_p^p(\mu, \lambda)$  rather than  $W_p(\mu, \lambda)$ . To keep notation simple, we write

$$W_\Omega^p(\mu, \lambda) = W_p^p(\mu \llcorner \Omega, \lambda \llcorner \Omega).$$

Moreover, if a measure is absolutely continuous with respect to Lebesgue measure, we only write its density. For example,

$$W_\Omega^p \left( \mu, \frac{\mu(\Omega)}{|\Omega|} \right),$$

denotes the transportation cost between  $\mu \llcorner \Omega$  to the uniform measure on  $\Omega$  with total mass  $\mu(\Omega)$ .

Occasionally we may write  $W_\Omega(\mu, \lambda)$  instead of  $(W_\Omega^p(\mu, \lambda))^{1/p}$ . This may lead to some ambiguity, but it should be clear from the context.

**2.2. Poisson point processes.** As in [12, 9, 4], we exploit invariance properties of Poisson point processes on  $\mathbb{R}^d$  with uniform intensity in order to obtain simpler sub-additivity estimates. We refer e.g. to [19] for a general introduction to Poisson point processes. Here we only recall that a Poisson point process on  $\mathbb{R}^d$  with intensity one can be defined as a random variable taking values on locally finite atomic measures

$$\mu = \sum_i \delta_{X_i}$$

such that, for every  $k \geq 1$ , for any disjoint Borel sets  $A_1, \dots, A_k \subseteq \mathbb{R}^d$ , the random variables  $\mu(A_1), \dots, \mu(A_k)$  are independent and  $\mu(A_i)$  has a Poisson distribution of parameter  $|A_i|$ , for every  $i = 1, \dots, k$ . In particular, if  $A \subseteq \mathbb{R}^d$  is Lebesgue negligible, then  $\mu(A) = 0$  almost surely.

Existence of a Poisson point process of intensity one is obtained via a superposition argument, noticing that on every bounded subset  $\Omega \subseteq \mathbb{R}^d$  the law of  $\mu \llcorner \Omega$  can be easily described: conditionally on  $\mu(\Omega) = n$ , the measure  $\mu \llcorner \Omega$  has the same law as the random measure

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

where  $(X_i)_{i=1}^n$  are independent random variables with uniform law on  $\Omega$ . Uniqueness in law can be also obtained, so that translation invariance of Lebesgue measure entails that the process is stationary, i.e., any deterministic translation of the random measure  $\mu$  leaves its law unchanged.

Let us finally recall the classical Cramér-Chernoff concentration bounds for Poisson random variables.

**Lemma 2.2.** *Let  $N$  be a Poisson random variable with parameter  $n \gg 1$ . Then, for every  $t > 0$*

$$\mathbb{P}[|N - n| \geq t] \leq 2 \exp\left(-\frac{t^2}{2(t+n)}\right). \quad (2.3)$$

As a consequence, for every  $q \geq 1$ ,

$$\mathbb{E}[|N - n|^q] \lesssim_q n^{\frac{q}{2}}. \quad (2.4)$$

*Proof.* For the concentration bound (2.3), see for instance [8]. By the layer-cake representation,

$$\begin{aligned} \mathbb{E}[|N - n|^q] &\lesssim \int_0^\infty t^{q-1} \exp\left(-\frac{t^2}{2(t+n)}\right) dt \\ &\lesssim \int_0^{\sqrt{n}} t^{q-1} dt + \int_{\sqrt{n}}^n t^{q-1} \exp\left(-c\frac{t^2}{n}\right) dt + \int_n^\infty t^{q-1} \exp(-ct) dt \\ &\lesssim_q n^{\frac{q}{2}} + n^{\frac{q}{2}} + n^{q-1} \exp(-cn) \lesssim_q n^{\frac{q}{2}}. \end{aligned}$$

□

### 3. THE MAIN LEMMAS

Our sub-additivity argument rests on a general but relatively simple lemma (which we only apply here for rectangles).

**Lemma 3.1.** *For every  $p \geq 1$ , there exists a constant  $C > 0$  depending only on  $p$  such that the following holds. For every Borel set  $\Omega \subset \mathbb{R}^d$ , every Borel partition  $(\Omega_i)_{i \in \mathbb{N}}$  of  $\Omega$ , every measures  $\mu, \lambda$  on  $\Omega$ , and every  $\varepsilon \in (0, 1)$ ,*

$$W_\Omega^p\left(\mu, \frac{\mu(\Omega)}{\lambda(\Omega)}\lambda\right) \leq (1 + \varepsilon) \sum_i W_{\Omega_i}^p\left(\mu, \frac{\mu(\Omega_i)}{\lambda(\Omega_i)}\lambda\right) + \frac{C}{\varepsilon^{p-1}} W_\Omega^p\left(\sum_i \frac{\mu(\Omega_i)}{\lambda(\Omega_i)}\chi_{\Omega_i}\lambda, \frac{\mu(\Omega)}{\lambda(\Omega)}\lambda\right). \quad (3.1)$$

*Proof.* We first use the triangle inequality (2.1) to get

$$W_\Omega^p\left(\mu, \frac{\mu(\Omega)}{\lambda(\Omega)}\lambda\right) \leq \left(W_\Omega\left(\mu, \sum_i \frac{\mu(\Omega_i)}{\lambda(\Omega_i)}\chi_{\Omega_i}\lambda\right) + W_\Omega\left(\sum_i \frac{\mu(\Omega_i)}{\lambda(\Omega_i)}\chi_{\Omega_i}\lambda, \frac{\mu(\Omega)}{\lambda(\Omega)}\lambda\right)\right)^p.$$

The proof is then concluded by combining the elementary inequality

$$(a + b)^p \leq (1 + \varepsilon)a^p + \frac{C}{\varepsilon^{p-1}}b^p \quad \forall a, b > 0 \text{ and } \varepsilon \in (0, 1), \quad (3.2)$$

with the sub-additivity of  $W_\Omega^p$  (2.2) in the form

$$W_\Omega^p\left(\mu, \sum_i \frac{\mu(\Omega_i)}{\lambda(\Omega_i)}\chi_{\Omega_i}\lambda\right) \leq \sum_i W_{\Omega_i}^p\left(\mu, \frac{\mu(\Omega_i)}{\lambda(\Omega_i)}\lambda\right).$$

□

**Remark 3.2.** Alternatively, we could have also stated (3.1) in the slightly more symmetric form:

$$\begin{aligned} W_\Omega^p\left(\frac{\mu}{\mu(\Omega)}, \frac{\lambda}{\lambda(\Omega)}\right) &\leq (1 + \varepsilon) \sum_i \frac{\mu(\Omega_i)}{\mu(\Omega)} W_{\Omega_i}^p\left(\frac{\mu}{\mu(\Omega_i)}, \frac{\lambda}{\lambda(\Omega_i)}\right) \\ &\quad + \frac{C}{\varepsilon^{p-1}} W_\Omega^p\left(\frac{1}{\mu(\Omega)} \sum_i \frac{\mu(\Omega_i)}{\lambda(\Omega_i)}\chi_{\Omega_i}\lambda, \frac{\lambda}{\lambda(\Omega)}\right). \end{aligned}$$

However, the sub-additivity argument turns out to be a little bit simpler using (3.1) instead.

Lemma 3.1 shows that in order to estimate the defect in sub-additivity, it is enough to bound the local defect of mass distribution. This will be done here through a PDE argument.

**Definition 3.3.** We say that a rectangle  $R = x + \prod_{i=1}^d [0, L_i]$  is of moderate aspect ratio if for every  $i, j$ ,  $L_i/L_j \leq 2$ . A partition  $\mathcal{R} = \{R_i\}$  of  $R$  is called admissible if for every  $i$ ,  $R_i$  is a rectangle of moderate aspect ratio and  $3^{-d}|R| \leq |R_i| \leq |R|$ . Notice that in particular  $\#\mathcal{R} \lesssim 1$  for every admissible partition.

**Lemma 3.4.** *Let  $R$  be a rectangle of moderate aspect ratio,  $\mu$  and  $\lambda$  be measures on  $R$  with equal mass, both absolutely continuous with respect to Lebesgue and such that  $\inf_R \lambda > 0$ . Then, for every  $p \geq 1$*

$$W_R^p(\mu, \lambda) \lesssim \frac{\text{diam}^p(R)}{(\inf_R \lambda)^{p-1}} \int_R |\mu - \lambda|^p. \quad (3.3)$$

*Proof.* Let  $\phi$  be a solution of the Poisson equation with Neumann boundary conditions

$$\Delta \phi = \mu - \lambda \quad \text{in } R \quad \text{and} \quad \nu \cdot \nabla \phi = 0 \quad \text{on } \partial R. \quad (3.4)$$

We first argue that

$$W_R^p(\mu, \lambda) \lesssim \frac{1}{(\inf_R \lambda)^{p-1}} \int_R |\nabla \phi|^p. \quad (3.5)$$

Let us point out that this estimate is well-known and has already been used in the context of the matching problem, see [3, 17] in the case  $p = 2$  and [21, Th. 2] for general  $p \geq 1$ . Still, we give a proof for the reader's convenience.

We first argue as in [17, Lem. 2.7], and use triangle inequality (2.1) and the monotonicity of  $W_R$  (2.2) to get

$$\begin{aligned} W_R(\mu, \lambda) &\leq W_R\left(\mu, \frac{1}{2}(\mu + \lambda)\right) + W_R\left(\frac{1}{2}(\mu + \lambda), \lambda\right) \\ &\leq W_R\left(\frac{1}{2}\mu, \frac{1}{2}\lambda\right) + W_R\left(\frac{1}{2}(\mu + \lambda), \lambda\right) \\ &= 2^{-\frac{1}{p}} (W_R(\mu, \lambda) + W_R(\mu + \lambda, 2\lambda)) \end{aligned}$$

and thus

$$W_R(\mu, \lambda) \lesssim W_R(\mu + \lambda, 2\lambda).$$

We now recall that by the Benamou-Brenier formula (see [25, Th. 5.28])

$$W_R^p(\mu + \lambda, 2\lambda) = \min_{\rho, j} \left\{ \int_0^1 \int_R \frac{1}{\rho^{p-1}} |j|^p : \partial_t \rho + \nabla \cdot j = 0, \rho_0 = \mu + \lambda, \rho_1 = 2\lambda \right\}.$$

Estimate (3.5) follows using

$$\rho_t = (1-t)\mu + t\lambda + \lambda \quad \text{and} \quad j = \nabla \phi$$

as competitor and noticing that for  $t \in [0, 1]$ ,  $\rho_t \geq \inf_R \lambda$ .

We now claim that

$$\int_R |\nabla \phi|^p \lesssim \text{diam}^p(R) \int_R |\mu - \lambda|^p, \quad (3.6)$$

which together with (3.5) would conclude the proof of (3.3). Estimate (3.6) is a direct consequence of Poincaré inequality and Calderón-Zygmund estimates for the Laplacian. However, since we did not find a precise reference for (global) Calderón-Zygmund estimates on rectangles with Neumann boundary conditions, we give here a short proof.

By scaling, we may assume that  $\text{diam}(R) = 1$ . Furthermore, using even reflections along  $\partial R$  we may replace Neumann boundary conditions by periodic ones in (3.4). By Poincaré

inequality [23, Prop. 12.29] (notice that thanks to the periodic boundary conditions we now have  $\int_R \nabla \phi = 0$ ),

$$\int_R |\nabla \phi|^p \lesssim \int_R |\nabla^2 \phi|^p.$$

By interior Calderón-Zygmund estimates (see for instance [15, Th. 7.3]), periodicity and the fact that  $R$  has moderate aspect ratio, we get

$$\int_R |\nabla^2 \phi|^p \lesssim \int_R |\mu - \lambda|^p + \left( \int_R |\nabla^2 \phi|^2 \right)^{\frac{p}{2}}.$$

By Bochner's formula and Hölder inequality,

$$\int_R |\nabla^2 \phi|^2 = \int_R |\mu - \lambda|^2 \lesssim \left( \int_R |\mu - \lambda|^p \right)^{\frac{2}{p}},$$

which concludes the proof of (3.6).  $\square$

**Remark 3.5.** For  $p = 2$ , combining the energy identity

$$\int_R |\nabla \phi|^2 = \int_R \phi(\lambda - \mu)$$

with Poincaré inequality, we see that (3.6) (and thus (3.3)) holds for any convex set  $R$ . Although the situation for  $p \geq 2$  is more subtle, see [14, Prop. 2], we believe that (3.6) holds for any rectangle, not necessarily of moderate aspect ratio.

#### 4. MATCHING TO THE REFERENCE MEASURE

In this section, we consider the optimal matching problem between  $\mu$  a Poisson point process on  $\mathbb{R}^d$  with intensity one and the Lebesgue measure. More precisely, for every  $L \geq 1$  we let

$$f^{\text{ref}}(L) = \mathbb{E} \left[ \frac{1}{|Q_L|} W_{Q_L}^p(\mu, \kappa) \right],$$

where  $Q_L = [0, L]^d$  and

$$\kappa = \frac{\mu(Q_L)}{|Q_L|}$$

is the generic constant for which this is well defined.

**Theorem 4.1.** *For every  $d \geq 3$  and  $p \geq 1$ , the limit*

$$f_\infty^{\text{ref}} = \lim_{L \rightarrow \infty} f^{\text{ref}}(L)$$

*exists and is strictly positive. Moreover, there exists  $C > 0$  depending on  $p$  and  $d$  such that for  $L \geq 1$ ,*

$$f_\infty^{\text{ref}} \leq f^{\text{ref}}(L) + \frac{C}{L^{\frac{d-2}{2}}}. \quad (4.1)$$

The proof follows the argument of [9] (see also [4]) and is mostly based on the following sub-additivity estimate.

**Proposition 4.2.** *For every  $d \geq 3$  and  $p \geq 1$ , there exists a constant  $C > 0$  such that for every  $L \geq 1$  and  $m \in \mathbb{N}$ ,*

$$f^{\text{ref}}(mL) \leq f^{\text{ref}}(L) + \frac{C}{L^{\frac{d-2}{2}}}. \quad (4.2)$$



*Proof.* We start by pointing out that since  $f^{\text{ref}}(L) \lesssim L^p$ , it is not restrictive to assume that  $L \gg 1$  in the proof of (4.2).

*Step 1.* [The dyadic case] For the sake of clarity, we start with the simpler case  $m = 2^k$  for some  $k \geq 1$ . We claim that

$$f^{\text{ref}}(2L) \leq f^{\text{ref}}(L) + \frac{C}{L^{\frac{d-2}{2}}}. \quad (4.3)$$

The desired estimate (4.2) would then follow iterating (4.3) and using that  $\sum_{k \geq 0} \frac{1}{2^{k(d-2)}} < \infty$  for  $d \geq 3$ . In order to prove (4.3), we divide the cube  $Q_{2L}$  in  $2^d$  sub-cubes  $Q_i = x_i + Q_L$  and let  $\kappa_i = \frac{\mu(Q_i)}{|Q_L|}$  (and  $\kappa = \frac{\mu(Q_{2L})}{|Q_{2L}|}$ ). Notice that we are considering a partition up to a Lebesgue negligible remainder, which gives no contribution almost surely. By (3.1), for every  $\varepsilon \in (0, 1)$ ,

$$W_{Q_{2L}}^p(\mu, \kappa) \leq (1 + \varepsilon) \sum_i W_{Q_i}^p(\mu, \kappa_i) + \frac{C}{\varepsilon^{p-1}} W_{Q_{2L}}^p \left( \sum_i \kappa_i \chi_{Q_i}, \kappa \right).$$

Dividing by  $|Q_{2L}|$ , taking expectations and using the fact that by translation invariance  $\mathbb{E}[\frac{1}{|Q_L|} W_{Q_i}^p(\mu, \kappa_i)] = f^{\text{ref}}(L)$ , we get

$$\begin{aligned} f^{\text{ref}}(2L) &\leq (1 + \varepsilon) \sum_i \frac{|Q_L|}{|Q_{2L}|} f^{\text{ref}}(L) + \frac{C}{\varepsilon^{p-1}} \mathbb{E} \left[ \frac{1}{|Q_{2L}|} W_{Q_{2L}}^p \left( \sum_i \kappa_i \chi_{Q_i}, \kappa \right) \right] \\ &= (1 + \varepsilon) f^{\text{ref}}(L) + \frac{C}{\varepsilon^{p-1}} \mathbb{E} \left[ \frac{1}{|Q_{2L}|} W_{Q_{2L}}^p \left( \sum_i \kappa_i \chi_{Q_i}, \kappa \right) \right]. \end{aligned}$$

We now estimate  $\mathbb{E}[\frac{1}{|Q_{2L}|} W_{Q_{2L}}^p(\sum_i \kappa_i \chi_{Q_i}, \kappa)]$ . Using  $\frac{1}{|Q_{2L}|} W_{Q_{2L}}^p(\sum_i \kappa_i \chi_{Q_i}, \kappa) \lesssim L^p \kappa$  together with

$$\mathbb{P} \left[ \kappa \leq \frac{1}{2} \right] \leq \exp(-cL^d),$$

which follows from the Cramér-Chernoff bounds (2.3), we may reduce ourselves to the event  $\{\kappa \geq \frac{1}{2}\}$ . Under this condition, by (3.3), we have

$$\begin{aligned} \frac{1}{|Q_{2L}|} W_{Q_{2L}}^p \left( \sum_i \kappa_i \chi_{Q_i}, \kappa \right) &\lesssim \frac{L^p}{L^d} \int_{Q_{2L}} \sum_i |\kappa_i - \kappa|^p \chi_{Q_i} \\ &\lesssim L^p \left( |\kappa - 1|^p + \sum_i |\kappa_i - 1|^p \right). \end{aligned}$$

Recalling that  $\mu(Q_i)$  are Poisson random variables of parameter  $|Q_i|$  and that  $\kappa_i = \frac{\mu(Q_i)}{|Q_i|}$ , we get from (2.4)

$$\mathbb{E}[|\kappa - 1|^p] \sim \mathbb{E}[|\kappa_i - 1|^p] \lesssim \frac{1}{L^{\frac{pd}{2}}}.$$

Thus

$$\mathbb{E} \left[ \frac{1}{|Q_{2L}|} W_{Q_{2L}}^p \left( \sum_i \kappa_i \chi_{Q_i}, \kappa \right) \right] \lesssim \frac{1}{L^{\frac{p}{2}(d-2)}} \quad (4.4)$$

and we conclude that for  $\varepsilon \in (0, 1)$ ,

$$f^{\text{ref}}(2L) \leq (1 + \varepsilon) f^{\text{ref}}(L) + \frac{C}{\varepsilon^{p-1}} \frac{1}{L^{\frac{p}{2}(d-2)}}. \quad (4.5)$$

Optimizing in  $\varepsilon$  by choosing  $\varepsilon = L^{-(d-2)/2}$ , and using that  $f^{\text{ref}}(L)$  is bounded (by (1.1) and (2.3), see for instance [16, Prop. 2.7] for details) we conclude the proof of (4.3). Let us point out that we used here boundedness of  $f^{\text{ref}}(L)$  for simplicity but that as shown

below it can also be obtained as a consequence of our proof.

*Step 2.*[The general case] We now consider the case when  $m \in \mathbb{N}$  is not necessarily dyadic. We will partition  $Q_{mL}$  into rectangles of almost dyadic size and thus need to deal with slightly more general configurations than dyadic cubes. Let us introduce some notation. Let  $R$  be a rectangle with moderate aspect ratio and  $\mathcal{R} = \{R_i\}$  be an admissible partition of  $R$  (recall Definition 3.3). Slightly abusing notation, we define

$$f^{\text{ref}}(R) = \mathbb{E} \left[ \frac{1}{|R|} W_R^p(\mu, \kappa) \right]. \quad (4.6)$$

*Step 2.1.* We claim that the following variant of (4.5) holds: for every rectangle  $R$  of moderate aspect ratio with  $|R| \gg 1$ , every admissible partition  $\mathcal{R}$  of  $R$  and every  $\varepsilon \in (0, 1)$ , we have

$$f^{\text{ref}}(R) \leq (1 + \varepsilon) \sum_i \frac{|R_i|}{|R|} f^{\text{ref}}(R_i) + \frac{C}{\varepsilon^{p-1}} \frac{1}{|R|^{\frac{p(d-2)}{2d}}}. \quad (4.7)$$

Defining  $\kappa_i = \frac{\mu(R_i)}{|R_i|}$  and using (3.1) as above, we get

$$f^{\text{ref}}(R) \leq (1 + \varepsilon) \sum_i \frac{|R_i|}{|R|} f^{\text{ref}}(R_i) + \frac{C}{\varepsilon^{p-1}} \mathbb{E} \left[ \frac{1}{|R|} W_R^p \left( \sum_i \kappa_i \chi_{R_i}, \kappa \right) \right].$$

The estimate

$$\mathbb{E} \left[ \frac{1}{|R|} W_R^p \left( \sum_i \kappa_i \chi_{R_i}, \kappa \right) \right] \lesssim \frac{1}{|R|^{\frac{p(d-2)}{2d}}} \quad (4.8)$$

is then obtained arguing exactly as for (4.4), using first the Crámer-Chernoff bound (2.3) to reduce to the event  $\{\kappa \geq \frac{1}{2}\}$  and then (3.3) (recalling that  $\text{diam}(R) \sim |R|^{\frac{1}{d}}$  since  $R$  has moderate aspect ratio) in combination with (2.4) and the fact that  $\#\mathcal{R} \lesssim 1$  since  $\mathcal{R}$  is an admissible partition.

*Step 2.2.* Starting from the cube  $Q_{mL}$ , let us construct a sequence of finer and finer partitions of  $Q_{mL}$  by rectangles of moderate aspect ratios and side-length given by integer multiples of  $L$ . We let  $\mathcal{R}_0 = \{Q_{mL}\}$  and define  $\mathcal{R}_k$  inductively as follows. Let  $R \in \mathcal{R}_k$ . Up to translation we may assume that  $R = \prod_{i=1}^d (0, m_i L)$  for some  $m_i \in \mathbb{N}$ . We then split each interval  $(0, m_i L)$  into  $(0, \lfloor \frac{m_i}{2} \rfloor L) \cup (\lfloor \frac{m_i}{2} \rfloor L, m_i L)$ . It is readily seen that this induces an admissible partition of  $R$ . Let us point out that when  $m_i = 1$  for some  $i$ , the corresponding interval  $(0, \lfloor \frac{m_i}{2} \rfloor L)$  is empty. This procedure stops after a finite number of steps  $K$  once  $\mathcal{R}_K = \{Q_L + z_i, z_i \in [0, m-1]^d\}$ . It is also readily seen that  $2^{K-1} < m \leq 2^K$  and that for every  $k \in [0, K]$  and every  $R \in \mathcal{R}_k$  we have  $|R| \sim (2^{K-k} L)^d$ . Let us prove by a downward induction that there exists  $\Lambda > 0$  such that for every  $k \in [0, K]$  and every  $R \in \mathcal{R}_k$ ,

$$f^{\text{ref}}(R) \leq f^{\text{ref}}(Q_L) + \Lambda(1 + f^{\text{ref}}(Q_L)) L^{-\frac{d-2}{2}} \sum_{j=K-k}^K 2^{-j \frac{d-2}{2}}. \quad (4.9)$$

This is clearly true for  $k = K$ . Assume that it holds true for  $k + 1$ . Let  $R \in \mathcal{R}_k$ . Applying (4.7) with  $\varepsilon = (2^{K-k}L)^{-(d-2)/2} \ll 1$ , we get

$$\begin{aligned} f^{\text{ref}}(R) &\leq (1 + \varepsilon) \sum_{R_i \in \mathcal{R}_{k+1}, R_i \subset R} \frac{|R_i|}{|R|} f^{\text{ref}}(R_i) + \frac{C}{\varepsilon} \frac{1}{|R|^{\frac{p(d-2)}{2d}}} \\ &\stackrel{(4.9)}{\leq} (1 + \varepsilon) \left( f^{\text{ref}}(Q_L) + \Lambda(1 + f^{\text{ref}}(Q_L))L^{-\frac{d-2}{2}} \sum_{j=K-k+1}^K 2^{-j\frac{d-2}{2}} \right) + C(2^{K-k}L)^{-\frac{d-2}{2}} \\ &\leq f^{\text{ref}}(Q_L) + \Lambda(1 + f^{\text{ref}}(Q_L))L^{-\frac{d-2}{2}} \\ &\quad \times \left[ \sum_{j=K-k+1}^K 2^{-j\frac{d-2}{2}} + 2^{-(K-k)\frac{d-2}{2}} \left( \frac{C}{\Lambda} + L^{-\frac{d-2}{2}} \sum_{j=K-k+1}^K 2^{-j\frac{d-2}{2}} \right) \right]. \end{aligned}$$

If  $L$  is large enough, then  $(\sum_{j=K-k+1}^K 2^{-j\frac{d-2}{2}})L^{-(d-2)/2} \leq \frac{1}{2}$ . Finally, choosing  $\Lambda \geq 2C$  yields (4.9).

Applying (4.9) to  $R = Q_{mL}$  and using that  $\sum_{j \geq 0} 2^{-j\frac{d-2}{2}} < \infty$ , we get

$$f^{\text{ref}}(mL) \leq f^{\text{ref}}(L) + C(1 + f^{\text{ref}}(L)) \frac{1}{L^{\frac{d-2}{2}}}.$$

Since  $f^{\text{ref}}(L) \lesssim L^p$ , writing that every  $L \gg 1$  may be written as  $L = mL'$  for some  $m \in \mathbb{N}$  and  $L' \in [1, 2]$ , we conclude that  $f^{\text{ref}}(L)$  is bounded and thus (4.2) follows.  $\square$

**Remark 4.3.** We point out that as a consequence of the proof of Proposition 4.2 we have for every rectangle  $R$  of moderate aspect ratio (recall definition (4.6))

$$f^{\text{ref}}(R) \lesssim 1. \tag{4.10}$$

We can now prove Theorem 4.1

*Proof of Theorem 4.1.* The existence of a limit  $f_\infty^{\text{ref}}$  is obtained from (4.2) arguing exactly as in [9] using the continuity of  $L \mapsto f^{\text{ref}}(L)$  (which can be obtained for instance by dominated convergence). The fact that  $f^{\text{ref}}(L) \gtrsim 1$  and thus  $f_\infty^{\text{ref}} > 0$  follows from (1.1) and (2.3). Finally, (4.1) follows from (4.2) by sending  $m \rightarrow \infty$  for fixed  $L$ .  $\square$

**Remark 4.4.** It may be conjectured from [10] that

$$|f^{\text{ref}}(L) - f_\infty^{\text{ref}}| \lesssim \frac{1}{L^{d-2}}.$$

Let us notice that if we could replace the term  $\varepsilon^{-(p-1)}$  in (4.5) by a constant then letting  $\varepsilon \rightarrow 0$  we would get the lower bound

$$f^{\text{ref}}(L) - f_\infty^{\text{ref}} \gtrsim -\frac{1}{L^{\frac{p(d-2)}{2}}},$$

which, at least for  $p = 2$ , is in line with the conjectured rate. See also Remark 6.4 for rates in the case of a deterministic number of points.

## 5. BI-PARTITE MATCHING

We now turn to the bi-partite matching. For  $\mu$  and  $\lambda$  two independent Poisson point processes of intensity one, we want to study the asymptotic behavior as  $L \rightarrow \infty$  of

$$f^{\text{bi}}(L) = \mathbb{E} \left[ \frac{1}{|Q_L|} W_{Q_L}^p(\mu, \kappa\lambda) \right],$$

with  $\kappa = \frac{\mu(Q_L)}{\lambda(Q_L)}$ . Analogously to Theorem 4.1, we have:

**Theorem 5.1.** *For every  $d \geq 3$  and  $p \geq 1$ , the limit*

$$f_\infty^{bi} = \lim_{L \rightarrow \infty} f^{bi}(L)$$

*exists and is strictly positive. Moreover, there exists  $C > 0$  depending on  $p$  and  $d$  such that for  $L \geq 1$ ,*

$$f_\infty^{bi} \leq f^{bi}(L) + \frac{C}{L^{\frac{d-2}{2p}}}.$$

The proof of Theorem 5.1 follows the same line of arguments as for Theorem 4.1. We only detail the estimate of the sub-additivity defect i.e. the counterpart of (4.8), since it is more delicate in the bipartite case.

**Proposition 5.2.** *Let  $R$  be a rectangle of moderate aspect ratio with  $|R| \gg 1$  and  $\mathcal{R} = \{R_i\}$  be an admissible partition of  $R$  (recall Definition 3.3). Defining  $\kappa_i = \frac{\mu(R_i)}{\lambda(R_i)}$  and  $\kappa = \frac{\mu(R)}{\lambda(R)}$ , we have for every  $d \geq 3$  and  $p \geq 1$ ,*

$$\mathbb{E} \left[ \frac{1}{|R|} W_R^p \left( \sum_i \kappa_i \chi_{R_i} \lambda, \kappa \lambda \right) \right] \lesssim \frac{1}{|R|^{\frac{d-2}{2d}}}. \quad (5.1)$$

*Proof.* As opposed to (4.8), since  $\lambda$  is atomic, we cannot directly use (3.3). The idea is to use as intermediate step the matching between  $\lambda$  and the reference measure. Let us point out that since  $f^{\text{ref}}(R_i) = \mathbb{E} \left[ \frac{1}{|R_i|} W_{R_i}^p \left( \lambda, \frac{\lambda(R_i)}{|R_i|} \right) \right]$  is of order one, we cannot apply naively the triangle inequality. The main observation is that since  $|\kappa_i - \kappa| \ll 1$  with overwhelming probability, the amount of mass which actually needs to be transported is very small.

Let  $1 \gg \theta > 0$  to be optimized later on. For every  $i$  let

$$\theta_i := (\kappa - \kappa_i) + \theta$$

and assume that

$$\theta \geq 2 \max_i |\kappa - \kappa_i| \quad (5.2)$$

so that  $\frac{3}{2}\theta \geq \theta_i \geq \frac{1}{2}\theta > 0$ . Notice that thanks to the Crámer-Chernoff bounds (2.3), (5.2) is satisfied with overwhelming probability as long as  $\theta \gg |R|^{-1/2}$ . Using the triangle inequality (2.1) we have

$$\begin{aligned} & W_R^p \left( \sum_i \kappa_i \chi_{R_i} \lambda, \kappa \lambda \right) \\ & \lesssim W_R^p \left( \sum_i \kappa_i \chi_{R_i} \lambda, \sum_i \chi_{R_i} \left[ (\kappa_i - \theta) \lambda + \theta \frac{\lambda(R_i)}{|R_i|} dx \right] \right) \\ & + W_R^p \left( \sum_i \chi_{R_i} \left[ (\kappa_i - \theta) \lambda + \theta \frac{\lambda(R_i)}{|R_i|} dx \right], \sum_i \chi_{R_i} \left[ (\kappa_i - \theta) \lambda + \theta_i \frac{\lambda(R_i)}{|R_i|} dx \right] \right) \\ & + W_R^p \left( \sum_i \chi_{R_i} \left[ (\kappa_i - \theta) \lambda + \theta_i \frac{\lambda(R_i)}{|R_i|} dx \right], \kappa \lambda \right). \end{aligned}$$

Notice that  $\sum_i \theta \lambda(R_i) = \sum_i \theta_i \lambda(R_i)$  by definition of  $\theta_i$  and the fact that  $\kappa \lambda(R) = \mu(R) = \sum_i \kappa_i \lambda(R)$  so that the second term is well defined.

We now estimate the three terms separately. The first and third are estimated in a similar way and we thus focus only on the first one. By sub-additivity (2.2) of  $W_R^p$ , we have

$$W_R^p \left( \sum_i \kappa_i \chi_{R_i} \lambda, \sum_i \chi_{R_i} \left[ (\kappa_i - \theta) \lambda + \theta \frac{\lambda(R_i)}{|R_i|} dx \right] \right) \leq \theta \sum_i W_{R_i}^p \left( \lambda, \frac{\lambda(R_i)}{|R_i|} \right).$$

We turn to the middle term. Again by sub-additivity of  $W_R^p$ ,

$$\begin{aligned} W_R^p \left( \sum_i \chi_{R_i} \left[ (\kappa_i - \theta)\lambda + \theta \frac{\lambda(R_i)}{|R_i|} dx \right], \sum_i \chi_{R_i} \left[ (\kappa_i - \theta)\lambda + \theta_i \frac{\lambda(R_i)}{|R_i|} dx \right] \right) \\ \leq W_R^p \left( \sum_i \chi_{R_i} \theta \frac{\lambda(R_i)}{|R_i|}, \sum_i \chi_{R_i} \theta_i \frac{\lambda(R_i)}{|R_i|} \right). \end{aligned}$$

Using (3.3) in the event  $\{\lambda(R_i)/|R_i| \sim 1\}$  (which has overwhelming probability) and recalling that we assumed  $\frac{3}{2}\theta \geq \theta_i \geq \frac{1}{2}\theta$ , we have

$$W_R^p \left( \sum_i \chi_{R_i} \theta \frac{\lambda(R_i)}{|R_i|}, \sum_i \chi_{R_i} \theta_i \frac{\lambda(R_i)}{|R_i|} \right) \lesssim \frac{\text{diam}^p(R)}{\theta^{p-1}} \sum_i |\kappa - \kappa_i|^p |R_i|.$$

Putting these two estimates together, dividing by  $|R|$  and taking expectations we find

$$\begin{aligned} \mathbb{E} \left[ \frac{1}{|R|} W_R^p \left( \sum_i \kappa_i \chi_{R_i} \lambda, \kappa \lambda \right) \right] &\lesssim \theta \sum_i f^{\text{ref}}(R_i) + \frac{\text{diam}^p(R)}{\theta^{p-1}} \sum_i \mathbb{E} [|\kappa - \kappa_i|^p] \\ &\stackrel{(4.10)\&(2.4)}{\lesssim} \theta + \frac{1}{\theta^{p-1}} \frac{1}{|R|^{\frac{d-2}{2d}}}. \end{aligned}$$

Optimizing in  $\theta$  by choosing  $\theta = |R|^{-\frac{d-2}{2d}} \gg |R|^{-1/2}$  (so that (5.2) is satisfied) this yields (5.1).  $\square$

Comparing (4.4) and (5.1), one may wonder if (5.1) is suboptimal and could be improved. Let us prove that it is not the case, at least if we consider the slightly more regular situation of the matching to a deterministic grid.

**Proposition 5.3.** *Let  $\lambda = \sum_{X \in \mathbf{Z}^d} \delta_X$  and for every  $L \gg 1$ ,  $(\mu_1, \dots, \mu_{2d})$  be  $2^d$  independent Poisson random variables of parameter  $L^d$ . Writing for  $L \gg 1$ ,  $Q_{2L} = \cup_{i=1}^{2^d} Q_i$  where  $Q_i$  are disjoint cubes of sidelength  $L$  and defining  $\kappa_i = \frac{\mu_i}{\lambda(Q_i)}$ ,  $\kappa = \frac{\sum_i \mu_i}{\lambda(Q_{2L})}$ , we have for  $d \geq 3$  and  $p \geq 1$ ,*

$$\mathbb{E} \left[ \frac{1}{|Q_{2L}|} W_{Q_{2L}}^p \left( \sum_i \kappa_i \chi_{Q_i} \lambda, \kappa \lambda \right) \right] \sim \frac{1}{L^{\frac{d-2}{2}}}.$$

*Proof. Step 1.* [The lower bound] We start by proving that

$$\mathbb{E} \left[ \frac{1}{|Q_{2L}|} W_{Q_{2L}}^p \left( \sum_i \kappa_i \chi_{Q_i} \lambda, \kappa \lambda \right) \right] \gtrsim \frac{1}{L^{\frac{d-2}{2}}}. \quad (5.3)$$

For this we notice that if  $\pi$  is any admissible coupling, then  $|x - y| \geq 1$  for every  $(x, y)$  in the support of  $\pi$  with  $x \neq y$  and thus<sup>2</sup>

$$W_{Q_{2L}}^p \left( \sum_i \kappa_i \chi_{Q_i} \lambda, \kappa \lambda \right) \geq W_{Q_{2L}}^1 \left( \sum_i \kappa_i \chi_{Q_i} \lambda, \kappa \lambda \right).$$

Let  $\Sigma = \cup_i \partial Q_i \cap Q_{2L}$  and  $\varepsilon_i = \text{sign}(\kappa_i - \kappa)$ . Using  $\xi(x) = d(x, \Sigma) \sum_i \varepsilon_i \chi_{Q_i}(x)$  as test function in the dual formulation of  $W_{Q_{2L}}^1$ , we obtain

$$W_{Q_{2L}}^1 \left( \sum_i \kappa_i \chi_{Q_i} \lambda, \kappa \lambda \right) \geq \sum_i \sum_{X \in \mathbf{Z}^d \cap Q_i} d(X, \Sigma) |\kappa_i - \kappa| \gtrsim L^{d+1} \sum_i |\kappa_i - \kappa|.$$

<sup>2</sup>recall that  $W_{Q_{2L}}^1$  denotes the 1-Wasserstein distance.

Taking expectations we find (5.3).

*Step 2.*[The upper bound] We turn to the corresponding upper bound,

$$\mathbb{E} \left[ \frac{1}{|Q_{2L}|} W_{Q_{2L}}^p \left( \sum_i \kappa_i \chi_{Q_i} \lambda, \kappa \lambda \right) \right] \lesssim \frac{1}{L^{\frac{d-2}{2}}}. \quad (5.4)$$

One could argue exactly as for (5.1) but we provide an alternative proof which uses a one-dimensional argument instead of (3.3). Notice that this argument could have also been used to give a different proof of (4.4).

*Step 2.1.*[The one-dimensional estimate] Let  $\lambda = \sum_{X \in \mathbf{Z}} \delta_X$  and for  $\kappa_1, \kappa_2 > 0$  and  $L \gg 1$ , define  $\kappa = \frac{\kappa_1 \lambda(0, L/2) + \kappa_2 \lambda(L/2, L)}{\lambda(0, L)}$ . We claim that if  $|\kappa_1 - \kappa_2| \ll L^{-1} \min(\kappa_1, \kappa_2)$ , then

$$W_{(0, L)}^p(\kappa_1 \chi_{(0, L/2)} \lambda + \kappa_2 \chi_{(L/2, L)} \lambda, \kappa \lambda) \lesssim L^2 |\kappa_1 - \kappa_2|. \quad (5.5)$$

Let us assume without loss of generality that  $\kappa_1 \geq \kappa_2$ . The optimal transport map is essentially symmetric around  $L/2$  and is given in  $(0, L/2)$  by sending a mass  $(k+1) \frac{\lambda(L/2, L)}{\lambda(0, L)} (\kappa_1 - \kappa_2)$  from position  $k$  to  $k+1$  (which is admissible since by hypothesis  $k|\kappa_1 - \kappa_2| \ll \min(\kappa_1, \kappa_2)$ ) so that

$$W_{(0, L)}^p(\mu, \lambda) \lesssim \sum_{k=0}^{L/2} k |\kappa_1 - \kappa_2| \sim L^2 |\kappa_1 - \kappa_2|.$$

This proves (5.5).

*Step 2.2.*[Proof of (5.4)] The proof is made recursively by layers. In the first step, we pass from  $2^d$  cubes to  $2^{d-1}$  rectangles of the form  $x + (0, L) \times (0, L/2)^{d-1}$ . For this we remark that for every cube  $Q_i$  there is exactly one cube  $Q_j$  such that  $Q_j = Q_i \pm \frac{L}{2} e_1$  (where  $e_i$  is the canonical basis of  $\mathbb{R}^d$ ). Let us focus on  $Q_1 = (0, L/2)^d$  and  $Q_2 = \frac{L}{2} e_1 + Q_1$ . Let  $R = Q_1 \cup Q_2 = (0, L) \times (0, L/2)^{d-1}$ . Define  $\hat{\kappa} = \frac{\kappa_1 \lambda(Q_1) + \kappa_2 \lambda(Q_2)}{\lambda(R)} = \frac{\mu_1 + \mu_2}{\lambda(R)}$ . We claim that

$$\mathbb{E} \left[ \frac{1}{|R|} W_R^2(\kappa_1 \chi_{Q_1} \lambda + \kappa_2 \chi_{Q_2} \lambda, \hat{\kappa} \lambda) \right] \lesssim \frac{1}{L^{\frac{d-2}{2}}}.$$

For this we notice that in  $R$ , the measures  $\kappa_1 \chi_{Q_1} \lambda + \kappa_2 \chi_{Q_2} \lambda$  and  $\hat{\kappa} \lambda$  are constant in the directions orthogonal to  $e_1$  and thus

$$\frac{1}{|R|} W_R^p(\kappa_1 \chi_{Q_1} \lambda + \kappa_2 \chi_{Q_2} \lambda, \hat{\kappa} \lambda) \lesssim \frac{1}{L} W_{(0, L)}^p(\kappa_1 \chi_{(0, L/2)} \lambda + \kappa_2 \chi_{(L/2, L)} \lambda, \hat{\kappa} \lambda).$$

Since  $\kappa_i = \frac{\mu_i}{\lambda(Q_i)}$ , by the Cramér-Chernoff bounds (2.3), we have  $|\kappa_i - 1| = O(L^{-d/2})$  with overwhelming probability. Hence, if  $d \geq 3$  we may apply (5.5) to get

$$\mathbb{E} \left[ \frac{1}{|R|} W_R^p(\kappa_1 \chi_{Q_1} \lambda + \kappa_2 \chi_{Q_2} \lambda, \hat{\kappa} \lambda) \right] \lesssim L \mathbb{E}[|\kappa_1 - \kappa_2|] \lesssim \frac{1}{L^{\frac{d-2}{2}}}$$

which proves the claim.

We finally iterate this argument  $2^d$  times. At every step  $k$  we have  $2^{d-k}$  rectangles of the form  $x + (0, L)^k \times (0, L/2)^{d-k}$  and each iteration has a cost of the same order (namely  $L^{-(d-2)/2}$ ). Using triangle inequality this concludes the proof of (5.4).  $\square$

**Remark 5.4.** Notice that in the proof of (5.5), we locally move a mass of order  $L^{-(d-2)/2}$  which corresponds to the optimal choice of  $\theta$  in the proof of (5.1).

## 6. DE-POISSONIZATION

In this section we discuss how to transfer limit results from matching problems with Poisson point process, i.e., with a random number of points, to those of with a deterministic number of points.

We use the following general result.

**Proposition 6.1.** *Let  $f : (0, \infty) \times \mathbb{N}^k \rightarrow [0, \infty)$ ,  $(L, n) \mapsto f(L|n)$ , satisfy the following assumptions:*

- (1) (*p*-homogeneity)  $f(L|n) = L^p f(1|n)$ , for every  $n \in \mathbb{N}^k$ ,  $L > 0$ ,
- (2) (boundedness)  $f(1|n) \lesssim 1$ , for every  $n \in \mathbb{N}^k$ ,
- (3) (monotonicity)  $f(1|n) \leq f(1|m)$ , for every  $m, n \in \mathbb{N}^k$  such that  $m_i \leq n_i$  for  $i = 1, \dots, k$ .

Define

$$f(L) = \mathbb{E}[f(L|N_L)]$$

where  $N_L = (N_{L,i})_{i=1}^k$  are i.i.d. Poisson random variables with parameter  $L^d$ . Then,

$$\liminf_{n \rightarrow \infty} n^{\frac{p}{d}} f(1|(n, \dots, n)) = \liminf_{L \rightarrow \infty} f(L), \quad \text{and} \quad \limsup_{n \rightarrow \infty} n^{\frac{p}{d}} f(1|(n, \dots, n)) = \limsup_{L \rightarrow \infty} f(L).$$

*Proof.* Let  $0 < \delta < L^d$  and introduce the event

$$A = \left\{ |N_{L,i} - L^d| < \delta \quad \text{for } i = 1, \dots, k \right\}.$$

By independence and Poisson tail bounds (2.3),

$$\mathbb{P}(A) \geq \left( 1 - 2 \exp\left(-\frac{1}{2} \frac{\delta^2}{L^d + \delta}\right) \right)^k. \quad (6.1)$$

We decompose

$$f(L) = \mathbb{E}[f(L|N_L)|A] \mathbb{P}(A) + \mathbb{E}[f(L|N_L)|A^c] \mathbb{P}(A^c).$$

If  $A$  holds, we use monotonicity of  $f(L|n)$  to argue that

$$L^p f(1|a) \leq \mathbb{E}[f(L|N_L)|A] \leq L^p f(1|b),$$

where  $a = L^d + \delta$ ,  $b = L^d - \delta$ . Otherwise, we use (1) and (2) to obtain

$$0 \leq \mathbb{E}[f(L|N_L)|A^c] \leq L^p.$$

Combining these inequalities, we have

$$L^p f(1|a) \mathbb{P}(A) \leq f(L) \leq L^p (f(1|b) + (1 - \mathbb{P}(A))). \quad (6.2)$$

For any  $n \geq 1$ , let  $L = L(n)$  be such such that  $L^d + L^{d/2} \sqrt{2\beta \log L} = n$ , for some fixed  $\beta > p$ . Then, we have from (6.2) with  $\delta = L^{d/2} \sqrt{2\beta \log L}$ ,

$$f(1|n) \leq \frac{f(L)}{L^p \mathbb{P}(A)}.$$

As  $n \rightarrow \infty$ , we have

$$\frac{n^{\frac{p}{d}}}{L^p} = (1 + L^{-\frac{d}{2}} \sqrt{2\beta \log L})^{\frac{p}{d}} = 1 + O\left(\sqrt{\frac{\log n}{n}}\right).$$

Moreover, by (6.1),

$$\mathbb{P}(A) \geq \left( 1 - 2 \exp\left(-\frac{1}{2} \frac{\delta^2}{L^d + \delta}\right) \right)^k = 1 - O(L^{-\beta}) = 1 - O\left(n^{-\frac{\beta}{d}}\right).$$

It follows that

$$\limsup_{n \rightarrow \infty} n^{\frac{p}{d}} f(1|(n, \dots, n)) \leq \limsup_{L \rightarrow \infty} f(L), \quad \text{and} \quad \liminf_{n \rightarrow \infty} n^{\frac{p}{d}} f(1|(n, \dots, n)) \leq \liminf_{L \rightarrow \infty} f(L).$$

To obtain the converse inequalities, we argue analogously choosing instead  $L = L(n)$  such that  $L^d - L^{d/2} \sqrt{2\beta \log L} = n$ .  $\square$

We now apply Proposition 6.1 to matching problems. Let us first consider the case of matching to the reference measure.

**Theorem 6.2.** *Let  $d \geq 3$ ,  $p \geq 1$ , and  $(X_i)_{i \geq 1}$  be i.i.d. uniform random variables on  $[0, 1]^d$ . Then,*

$$\lim_{n \rightarrow \infty} n^{\frac{p}{d}} \mathbb{E} \left[ W_{[0,1]^d}^p \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, 1 \right) \right] = f_{\infty}^{\text{ref}},$$

with  $f_{\infty}^{\text{ref}}$  as in Theorem 4.1.

*Proof.* Recalling the notation  $Q_1 = [0, 1]^d$ , we introduce the function

$$f(L|n) = L^p \mathbb{E} \left[ W_{Q_1}^p \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, 1 \right) \right], \quad \text{if } n \geq 1,$$

and  $f(L|0) = 0$ . It is clearly bounded and  $p$ -homogeneous in the sense of Proposition 6.1. To show monotonicity, let  $1 \leq m \leq n$  and use the identity

$$\frac{1}{n} \sum_{i=1}^n \delta_{X_i} = \binom{n}{m}^{-1} \sum_{\substack{I \subseteq \{1, \dots, n\} \\ |I|=m}} \frac{1}{m} \sum_{i \in I} \delta_{X_i}$$

in combination with the convexity of the transportation cost, to obtain

$$\begin{aligned} W_{Q_1}^p \left( \sum_{i=1}^n \delta_{X_i}, 1 \right) &= W_{Q_1}^p \left( \binom{n}{m}^{-1} \sum_{\substack{I \subseteq \{1, \dots, n\} \\ |I|=m}} \frac{1}{m} \sum_{i \in I} \delta_{X_i}, \binom{n}{m}^{-1} \sum_{\substack{I \subseteq \{1, \dots, n\} \\ |I|=m}} 1 \right) \\ &\leq \binom{n}{m}^{-1} \sum_{\substack{I \subseteq \{1, \dots, n\} \\ |I|=m}} W_{Q_1}^p \left( \frac{1}{m} \sum_{i \in I} \delta_{X_i}, 1 \right). \end{aligned}$$

Taking expectation yields  $f(1|n) \leq f(1|m)$ , since for  $I \subseteq \{1, \dots, n\}$  with  $|I| = m$ ,  $\{X_i\}_{i \in I}$  have the same law as  $\{X_i\}_{i=1}^m$ .

Let  $\mu$  be a Poisson point process of intensity one on  $\mathbb{R}^d$  and for  $L > 1$  let  $N_L = \mu(Q_L)$  be a Poisson random variable of parameter  $L^d$ . For  $n \geq 1$ , we notice that

$$\begin{aligned} f(L|n) &= \mathbb{E} \left[ W_{Q_L}^p \left( \frac{\mu}{\mu(Q_L)}, \frac{1}{|Q_L|} \right) \middle| N_L = n \right] \\ &= \frac{L^d}{n} f^{\text{ref}}(L|n), \end{aligned} \tag{6.3}$$

where we write, extending the notation from Section 4,

$$f^{\text{ref}}(L|n) = \mathbb{E} \left[ \frac{1}{|Q_L|} W_{Q_L}^p(\mu, \kappa) \middle| N_L = n \right],$$

with  $\kappa = \mu(Q_L)/|Q_L|$  (and  $f^{\text{ref}}(L|0) = 0$ ). Notice that

$$f^{\text{ref}}(L) = \mathbb{E}[f^{\text{ref}}(L|N_L)].$$

In order to apply Proposition 6.1 and conclude the proof, we argue that

$$\lim_{L \rightarrow \infty} \mathbb{E} \left[ \left| f(L|N_L) - f^{\text{ref}}(L|N_L) \right| \right] = 0. \tag{6.4}$$



To this aim, we first bound  $\mathbb{E}[f(L|N_L)]$  uniformly from above as  $L \rightarrow \infty$ . For this we combine the following two simple facts. First, since  $f(L|n) \lesssim L^p$ , letting  $A_L = \{N_L \geq |Q_L|/2\}$ , we have

$$\mathbb{E}[f(L|N_L)|A_L^c] \lesssim L^p.$$

Second, since by (2.3)  $\mathbb{P}[A_L] \gtrsim 1$ ,

$$\mathbb{E}[f(L|N_L)|A_L] \stackrel{(6.3)}{=} \mathbb{E}\left[\frac{L^d}{N_L} f^{\text{ref}}(L|N_L) \Big| A_L\right] \lesssim \mathbb{E}[f^{\text{ref}}(L|N_L)] = f^{\text{ref}}(L) \lesssim 1,$$

where in the last inequality we used that  $f^{\text{ref}}$  is bounded as a consequence of Theorem 4.1. Therefore,

$$\begin{aligned} \mathbb{E}[f(L|N_L)] &= \mathbb{E}[f(L|N_L)|A_L^c]\mathbb{P}[A_L^c] + \mathbb{E}[f(L|N_L)|A_L]\mathbb{P}[A_L] \\ &\lesssim L^p\mathbb{P}[A_L^c] + 1 \\ &\stackrel{(2.3)}{\lesssim} L^p \exp(-cL^d) + 1 \lesssim 1. \end{aligned}$$

Using Hölder inequality with  $\frac{1}{q} + \frac{1}{q'} = 1$ , and the fact that  $f(L|N_L) \lesssim L^p$ ,

$$\begin{aligned} \mathbb{E}\left[\left|f(L|N_L) - f^{\text{ref}}(L|N_L)\right|\right] &\stackrel{(6.3)}{=} \mathbb{E}\left[\left|\frac{N_L}{|Q_L|} - 1\right| f(L|N_L)\right] \\ &\leq \mathbb{E}\left[\left|\frac{N_L}{|Q_L|} - 1\right|^{q'}\right]^{\frac{1}{q'}} \mathbb{E}[f(L|N_L)^q]^{\frac{1}{q}}. \\ &\stackrel{(2.4)}{\lesssim_q} L^{-\frac{d}{2}} L^{\frac{p(q-1)}{q}} \mathbb{E}[f(L|N_L)]^{\frac{1}{q}} \lesssim_q L^{-\frac{d}{2} + \frac{p(q-1)}{q}}. \end{aligned}$$

Choosing  $q$  close enough to 1, in particular  $1 < q < \frac{2p}{2p-d}$  if  $p > \frac{d}{2}$ , we get (6.4).  $\square$

Arguing similarly, we obtain the corresponding result for the bi-partite matching on the unit cube, that is Theorem 1.1, which we restate for the reader's convenience.

**Theorem 6.3.** *Let  $d \geq 3$ ,  $p \geq 1$ , and  $(X_i)_{i \geq 1}$ ,  $(Y_i)_{i \geq 1}$ , be independent uniform random variables on  $[0, 1]^d$ . Then,*

$$\lim_{n \rightarrow \infty} n^{\frac{p}{d}} \mathbb{E}\left[W_{[0,1]^d}^p\left(\frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}\right)\right] = f_{\infty}^{\text{bi}},$$

with  $f_{\infty}^{\text{bi}}$  as in Theorem 5.1.

*Proof.* The proof is very similar to that of Theorem 6.2, but in this case we define the function

$$f(L|n_1, n_2) = L^p \mathbb{E}\left[W_{[0,1]^d}^p\left(\frac{1}{n_1} \sum_{i=1}^{n_1} \delta_{X_i}, \frac{1}{n_2} \sum_{i=1}^{n_2} \delta_{Y_i}\right)\right]$$

for  $n_1, n_2 \geq 1$ , and let  $f(L|n_1, n_2) = 0$  otherwise. It is clearly bounded and  $p$ -homogeneous. Using the identities

$$\begin{aligned} \frac{1}{n_1} \sum_{i=1}^{n_1} \delta_{X_i} &= \binom{n_1}{m_1}^{-1} \binom{n_2}{m_2}^{-1} \sum_{\substack{I_1 \subseteq \{1, \dots, n_1\} \\ |I_1|=m_1}} \sum_{\substack{I_2 \subseteq \{1, \dots, n_2\} \\ |I_2|=m_2}} \frac{1}{m_1} \sum_{i \in I_1} \delta_{X_i} \\ \frac{1}{n_2} \sum_{i=1}^{n_2} \delta_{Y_i} &= \binom{n_1}{m_1}^{-1} \binom{n_2}{m_2}^{-1} \sum_{\substack{I_1 \subseteq \{1, \dots, n_1\} \\ |I_1|=m_1}} \sum_{\substack{I_2 \subseteq \{1, \dots, n_2\} \\ |I_2|=m_2}} \frac{1}{m_2} \sum_{i \in I_2} \delta_{Y_i} \end{aligned}$$

we obtain monotonicity arguing analogously as in Theorem 6.2. The proof is then concluded as in Theorem 6.2 and we omit the details.  $\square$

**Remark 6.4** (Convergence rates). An inspection of the proof of Proposition 6.1 shows that one can transfer rates of convergence (even only one-sided) from the Poisson case to that of a fixed number of points. In the case of matching to the reference measure this leads to the inequality

$$f_\infty^{\text{ref}} \leq n^{\frac{p}{d}} \mathbb{E} \left[ W_{Q_1}^p \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, 1 \right) \right] + Cn^{\frac{2-d}{2d}},$$

while for the bi-partite matching we obtain

$$f_\infty^{\text{bi}} \leq n^{\frac{p}{d}} \mathbb{E} \left[ W_{Q_1}^p \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \frac{1}{n} \sum_{i=1}^n \delta_{Y_i} \right) \right] + Cn^{\frac{2-d}{2dp}},$$

for some constant  $C \geq 0$ . Besides being one-sided bounds, these are still far from the conjectured rates in [10], which for  $p = 2$  read

$$n^{\frac{2}{d}} \mathbb{E} \left[ W_{Q_1}^2 \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \frac{1}{n} \sum_{i=1}^n \delta_{Y_i} \right) \right] = f_\infty^{\text{bi}} + Cn^{\frac{2-d}{d}} + o(n^{\frac{2-d}{d}}),$$

with an explicit constant  $C$ .

**Remark 6.5** (Strong convergence). If  $p < d$ , standard concentration of measure arguments allow to obtain strong convergence from convergence of the expected values (see also [3] for a similar argument in the case  $p = d = 2$ ). Let us consider for example the case of bi-partite matching, and show that, both  $\mathbb{P}$ -a.s. and in  $L^1(\mathbb{P})$ ,

$$\lim_{n \rightarrow \infty} n^{\frac{p}{d}} W_{Q_1}^p \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \frac{1}{n} \sum_{i=1}^n \delta_{Y_i} \right) = f_\infty^{\text{bi}}, \quad (6.5)$$

with  $(X_i)_{i \geq 1}, (Y_i)_{i \geq 1}$  independent uniform random variables on  $Q_1 = [0, 1]^d$ .

For any  $n \geq 1$ , the function

$$[0, 1]^{d \times 2n} \ni (x_1, \dots, x_n, y_1, \dots, y_n) \mapsto n^{\frac{1}{d}} W_p \left( \mu_{(x_i)_{i=1}^n}, \mu_{(y_i)_{i=1}^n} \right),$$

is  $2n^{\frac{1}{d} - \min\{\frac{1}{2}, \frac{1}{p}\}}$ -Lipschitz with respect to the Euclidean distance, where we write

$$\mu_{(x_i)_{i=1}^n} = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}, \quad \text{for } x_i \in [0, 1]^d.$$

This relies on the triangle inequality (2.1) and the fact that

$$(x_i)_{i=1}^n \mapsto \mu_{(x_i)_{i=1}^n} \quad \text{is } n^{-\min\{\frac{1}{2}, \frac{1}{p}\}}\text{-Lipschitz}$$

if we endow the set of probability measures on  $[0, 1]^d$  with the Wasserstein distance of order  $p$ . Indeed, for  $(x_i)_{i=1}^n, (y_i)_{i=1}^n \in [0, 1]^{d \times n}$ , then

$$W_p \left( \mu_{(x_i)_{i=1}^n}, \mu_{(y_i)_{i=1}^n} \right) \leq \left( \frac{1}{n} \sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}} \leq n^{-\min\{\frac{1}{2}, \frac{1}{p}\}} \left( \sum_{i=1}^n |x_i - y_i|^2 \right)^{\frac{1}{2}}.$$

Gaussian concentration for the uniform measure on the unit cube [20, Prop. 2.8] yields that if

$$Z_n = n^{\frac{1}{d}} W_p \left( \mu_{(X_i)_{i=1}^n}, \mu_{(Y_i)_{i=1}^n} \right),$$

then, for  $r > 0$ ,

$$\mathbb{P} (|Z_n - \mathbb{E}[Z_n]| \geq r) \leq 2 \exp \left( -cn^{2(\min\{\frac{1}{2}, \frac{1}{p}\} - \frac{1}{d})} r^2 \right),$$

where  $c > 0$  is an absolute constant. A standard application of Borel-Cantelli Lemma gives that, if  $1 \leq p < d$  and  $d \geq 3$ , then

$$\lim_{n \rightarrow \infty} (Z_n - \mathbb{E}[Z_n]) = 0, \quad \mathbb{P}\text{-a.s.} \quad (6.6)$$

Moreover, using the layer-cake formula, we obtain the inequality

$$\mathbb{E}[|Z_n - \mathbb{E}[Z_n]|^p] \lesssim n^{\frac{p}{d} - \min\{\frac{p}{2}, 1\}}, \quad (6.7)$$

which is infinitesimal as  $n \rightarrow \infty$ . To conclude, it is sufficient to argue that  $\lim_{n \rightarrow \infty} \mathbb{E}[Z_n] = (f_\infty^{\text{bi}})^{1/p}$ , since it yields by (6.7) and (6.6) that  $\lim_{n \rightarrow \infty} Z_n = (f_\infty^{\text{bi}})^{1/p}$  in  $L^p(\mathbb{P})$  and  $\mathbb{P}$ -a.s. and hence (6.5). By Theorem 6.2,  $\lim_{n \rightarrow \infty} \mathbb{E}[Z_n^p] = f_\infty^{\text{bi}}$ . By Jensen's inequality,

$$\limsup_{n \rightarrow \infty} \mathbb{E}[Z_n]^p \leq \lim_{n \rightarrow \infty} \mathbb{E}[Z_n^p] = f_\infty^{\text{bi}}.$$

By the elementary inequality (3.2), we have, for any  $\varepsilon > 0$ ,

$$Z_n^p \leq (\mathbb{E}[Z_n] + |Z_n - \mathbb{E}[Z_n]|)^p \leq (1 + \varepsilon) \mathbb{E}[Z_n]^p + \frac{C}{\varepsilon^{p-1}} |Z_n - \mathbb{E}[Z_n]|^p.$$

Taking expectation and letting  $n \rightarrow \infty$ , using (6.7), we obtain

$$f_\infty^{\text{bi}} = \lim_{n \rightarrow \infty} \mathbb{E}[Z_n^p] \leq (1 + \varepsilon) \liminf_{n \rightarrow \infty} \mathbb{E}[Z_n]^p.$$

Letting  $\varepsilon \rightarrow 0$  we conclude.

#### ACKNOWLEDGMENTS

M.G was partially supported by the project ANR-18-CE40-0013 SHAPO financed by the French Agence Nationale de la Recherche (ANR) and by the LYSM LIA AMU CNRS ECM INdAM. M.G. also thanks the Centro de Giorgi in Pisa for its hospitality. D.T. was partially supported by the University of Pisa, Project PRA 2018-49, and Gnampa project 2019 ‘‘Proprietà analitiche e geometriche di campi aleatori’’.

#### REFERENCES

1. M. Ajtai, J. Komlós, and G. Tusnády, *On optimal matchings.*, *Combinatorica* **4** (1984), 259–264.
2. L. Ambrosio and F. Glaudo, *Finer estimates on the 2-dimensional matching problem*, *J. Éc. polytech. Math.* **6** (2019), 737–765.
3. L. Ambrosio, F. Stra, and D. Trevisan, *A PDE approach to a 2-dimensional matching problem*, *Probab. Theory Relat. Fields* **173** (2019), no. 1-2, 433–477.
4. F. Barthe and C. Bordenave, *Combinatorial optimization over two random point sets.*, *Séminaire de probabilités XLV*, Cham: Springer, 2013, pp. 483–535.
5. D. Benedetto and E. Caglioti, *Euclidean random matching in 2d for non-constant densities*, *Journal of Statistical Physics* **181** (2020), no. 3, 854–869.
6. S. Bobkov and M. Ledoux, *One-dimensional empirical measures, order statistics, and Kantorovich transport distances*, *Mem. Amer. Math. Soc.* **261** (2019), no. 1259, v+126.
7. S. Bobkov and M. Ledoux, *A simple Fourier analytic proof of the AKT optimal matching theorem*, 2019.
8. S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities*, Oxford University Press, Oxford, 2013.
9. J. H. Boutet de Monvel and O. C. Martin, *Almost sure convergence of the minimum bipartite matching functional in Euclidean space*, *Combinatorica* **22** (2002), no. 4, 523–530.
10. S. Caracciolo, C. Lucibello, G. Parisi, and G. Sicuro, *Scaling hypothesis for the Euclidean bipartite matching problem*, *Physical Review E* **90** (2014), no. 1.
11. S. Dereich, M. Scheutzow, and R. Schottstedt, *Constructive quantization: approximation by empirical measures*, *Ann. Inst. Henri Poincaré Probab. Stat.* **49** (2013), no. 4, 1183–1203.
12. V. Dobrić and J.E. Yukich, *Asymptotics for transportation cost in high dimensions.*, *J. Theor. Probab.* **8** (1995), no. 1, 97–118.
13. N. Fournier and A. Guillin, *On the rate of convergence in Wasserstein distance of the empirical measure*, *Probab. Theory Related Fields* **162** (2015), no. 3-4, 707–738.
14. S. J. Fromm, *Potential space estimates for Green potentials in convex domains*, *Proc. Amer. Math. Soc.* **119** (1993), no. 1, 225–233.

15. M. Giaquinta and L. Martinazzi, *An introduction to the regularity theory for elliptic systems, harmonic maps and minimal graphs*, second ed., Appunti. Scuola Normale Superiore di Pisa (Nuova Serie), vol. 11, Edizioni della Normale, Pisa, 2012.
16. M. Goldman, M. Huesmann, and F. Otto, *A large-scale regularity theory for the Monge-Ampère equation with rough data and application to the optimal matching problem*, Aug 2018.
17. ———, *Quantitative linearization results for the Monge-Ampère equation*, (2019).
18. M. Huesmann and K.-T. Sturm, *Optimal transport from Lebesgue to Poisson*, Ann. Probab. **41** (2013), no. 4, 2426–2478.
19. G. Last and M. Penrose, *Lectures on the Poisson process*, vol. 7, Cambridge University Press, 2017.
20. M. Ledoux, *The concentration of measure phenomenon*, no. 89, American Mathematical Soc., 2001.
21. ———, *On optimal matching of Gaussian samples*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **457** (2017), no. Veroyatnost' i Statistika. 25, 226–264.
22. M. Ledoux and J.-X. Zhu, *On optimal matching of Gaussian samples III*, arXiv preprint arXiv:1911.07579 (2019).
23. G. Leoni, *A first course in Sobolev spaces*, second ed., Graduate Studies in Mathematics, vol. 181, American Mathematical Society, Providence, RI, 2017.
24. G. Peyré, M. Cuturi, et al., *Computational optimal transport: With applications to data science*, Foundations and Trends® in Machine Learning **11** (2019), no. 5-6, 355–607.
25. F. Santambrogio, *Optimal transport for applied mathematicians*, Progress in Nonlinear Differential Equations and their Applications, vol. 87, Birkhäuser/Springer, Cham, 2015, Calculus of variations, PDEs, and modeling.
26. M. Talagrand, *Matching random samples in many dimensions*, Ann. Appl. Probab. **2** (1992), no. 4, 846–856.
27. ———, *Upper and lower bounds for stochastic processes: modern methods and classical problems*, vol. 60, Springer Science & Business Media, 2014.
28. C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.
29. J. E. Yukich, *Probability theory of classical Euclidean optimization problems.*, Berlin: Springer, 1998.

M.G.: UNIVERSITÉ DE PARIS, CNRS, SORBONNE-UNIVERSITÉ, LABORATOIRE JACQUES-LOUIS LIONS (LJLL), F-75005 PARIS, FRANCE

*Email address:* `goldman@u-paris.fr`

D.T.: DIPARTIMENTO DI MATEMATICA, UNIVERSITÀ DEGLI STUDI DI PISA, 56125 PISA, ITALY

*Email address:* `dario.trevisan@unipi.it`