

# ON THE QUADRATIC RANDOM MATCHING PROBLEM IN TWO-DIMENSIONAL DOMAINS

LUIGI AMBROSIO, MICHAEL GOLDMAN, AND DARIO TREVISAN

**ABSTRACT.** We investigate the average minimum cost of a bipartite matching, with respect to the squared Euclidean distance, between two samples of  $n$  i.i.d. random points on a bounded Lipschitz domain in the Euclidean plane, whose common law is absolutely continuous with strictly positive Hölder continuous density. We confirm in particular the validity of a conjecture by D. Benedetto and E. Caglioti stating that the asymptotic cost as  $n$  grows is given by the logarithm of  $n$  multiplied by an explicit constant times the volume of the domain. Our proof relies on a reduction to the optimal transport problem between the associated empirical measures and a Whitney-type decomposition of the domain, together with suitable upper and lower bounds for local and global contributions, both ultimately based on PDE tools. We further show how to extend our results to more general settings, including Riemannian manifolds, and also give an application to the asymptotic cost of the random quadratic bipartite travelling salesperson problem.

## 1. INTRODUCTION

The minimum weight perfect matching problem on bipartite graphs, also called assignment problem, is a combinatorial optimization problem which has been historically subject of intense research by several communities, well beyond operation research and algorithm theory, including combinatorics and graph theory [29], probability and statistics [44] and even theoretical physics [32, 31]. Applications related to planning and allocation of resources are classical, but recently they have seen an increased interest, e.g. in the online version of the problem, related to Internet advertising [30]. The assignment problem and its common linear programming relaxation, the optimal transport problem, provide also useful tools in machine learning and data science [5, 34], mostly because of its efficiency and versatility at discriminating between empirical distributions.

In this article, we focus on the Euclidean formulation of the problem, where we are given two finite families of points  $(x_i)_{i=1}^n, (y_j)_{j=1}^m \subseteq \mathbb{R}^d$ , with  $m \geq n$ , and the matching cost is defined, for a given parameter  $p > 0$ ,

$$\min_{\sigma \in \mathcal{S}_{n,m}} \sum_{i=1}^n |x_i - y_{\sigma(i)}|^p, \quad (1.1)$$

where  $\mathcal{S}_{n,m}$  denotes the set of injective maps  $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, m\}$ . This corresponds to a minimum weight perfect matching problem on the complete bipartite graph and edge weights  $w_{ij} = |x_i - y_j|^p$ . In particular, when  $n = m$ ,  $\mathcal{S}_{n,n} = \mathcal{S}_n$  is the symmetric group of permutations over  $n$  elements.

A linear programming approximation of (1.1) is given by the Monge-Kantorovich optimal transport problem, where  $\sigma$  is replaced by a transport plan or coupling, i.e., a matrix  $(\pi_{ij})_{i=1, \dots, n}^{j=1, \dots, m} \in [0, 1]^{n \times m}$  with  $\sum_i \pi_{ij} = 1/m$ ,  $\sum_j \pi_{ij} = 1/n$ . Minimization among all couplings yields a particular instance of the Wasserstein cost of order  $p > 0$  between the

---

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

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

L.A. thanks the support of the PRIN 2017 project “Gradient Flows, Optimal Transport and Metric Measure Structures”. M.G. was partially supported by the project ANR-18-CE40-0013 SHAPO financed by the French Agence Nationale de la Recherche (ANR). D.T. was partially supported by Gnampa project 2020 “Problemi di ottimizzazione con vincoli via trasporto ottimo e incertezza”.

empirical measures  $\frac{1}{n} \sum_{i=1}^n \delta_{x_i}$  and  $\frac{1}{m} \sum_{j=1}^m \delta_{y_j}$ ,

$$\min_{\pi} \sum_{i=1}^n \sum_{j=1}^m \pi_{ij} |x_i - y_j|^p = W_p^p \left( \frac{1}{n} \sum_{i=1}^n \delta_{x_i}, \frac{1}{m} \sum_{j=1}^m \delta_{y_j} \right). \quad (1.2)$$

The Birkhoff-von Neumann theorem [29] provides an exact correspondence between (1.1) and (1.2) when  $n = m$ , i.e., both the costs and optimizers are the same, an optimal plan is as a permutation matrix (both up to a factor  $1/n$ ). As a consequence, one can exploit the rich analytic structure of optimal transport and use tools from convex analysis and partial differential equations.

We further assume that the points  $x_i = X_i$ ,  $y_j = Y_j$  are samples of independent identically distributed (i.i.d.) random variables with a common law. This is the paradigm of many problems in geometric probability, in particular related to the theory of Euclidean additive functionals [40, 51]. The random Euclidean bipartite matching problem is in fact known to be quite challenging: general results and techniques that give precise results in similar problems (e.g. the non-bipartite matching problem) fail here, most notably when the dimension of the space is small.

The “critical” case turns out to be  $d = 2$ , which we also consider in this work. In this case, Ajtai, Komlós and Tusnády [1] first showed that for i.i.d. uniform samples on the square  $(0, 1)^2$  the asymptotics in the case  $n = m \rightarrow \infty$  reads, for<sup>1</sup>  $p \geq 1$ ,

$$\mathbb{E} \left[ W_p^p \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \frac{1}{n} \sum_{i=1}^n \delta_{Y_i} \right) \right] \sim \left( \frac{\log n}{n} \right)^{p/2}. \quad (1.3)$$

Talagrand [42] subsequently investigated the case of general laws, possibly not even absolutely continuous, providing a universal upper bound for  $p = 1$ , with the same rate.

The parameter  $p > 0$  plays a relevant role, since the solution to both problems (1.1) and (1.2) in general depends on  $p$ , and for  $d = 2$ ,  $p < 1$  the correct rate in (1.3) turns out to be  $n^{-p/2}$ , i.e. no logarithmic corrections appear [7]. The classical literature focused mostly on the case  $p = 1$  – but also on the case  $p = \infty$ , see [28, 46].

A recent breakthrough in the quadratic case  $p = 2$  was obtained by the statistical physics community, starting from the seminal work [16] and further developed in [17, 39]. By formally linearising the Monge-Ampère equation around the constant density to obtain the Poisson equation, they argued in particular that the optimal transport cost should be well-approximated by a negative Sobolev norm of the difference of the empirical measures. After a renormalization procedure to cut-off divergences, this led for uniform points on  $(0, 1)^2$  to the conjecture

$$\lim_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E} \left[ W_2^2 \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \frac{1}{n} \sum_{i=1}^n \delta_{Y_i} \right) \right] = \frac{1}{2\pi}. \quad (1.4)$$

A rigorous mathematical proof of (1.4) was later obtained in [4]. Further simplifications and improvements of this method have been done in [2], leading to quantitative bounds for optimizers [3]. See also [20] for a justification of the linearisation ansatz of [16] down to mesoscopic scales based on a large-scale regularity theory for the Monge-Ampère equation [21]. Let us point out that for  $p \geq 1$ , with the exception of  $p = 2$ , even in the case of uniform points on  $(0, 1)^2$  it is still an open problem [44, Section 4.3.3] to determine if the limit in (1.3) exists, i.e. prove the existence of

$$\lim_{n \rightarrow \infty} \left( \frac{n}{\log n} \right)^{p/2} \mathbb{E} \left[ W_p^p \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \frac{1}{n} \sum_{i=1}^n \delta_{Y_i} \right) \right].$$

---

<sup>1</sup>We use the notation  $A \sim B$  if both  $A \lesssim B$  and  $B \lesssim A$  hold, where in turn  $A \lesssim B$  means that there exists a global constant  $C = C(d, p, \Omega) > 0$  depending only on the dimension  $d$ ,  $p \geq 1$  and on the fixed domain  $\Omega$  such that  $A \leq CB$ .

A natural question is how (1.4) should be modified when the uniform law is replaced by a different density. As a first step in this direction, it is proven in [2, 4] that (1.4) holds on general closed compact Riemannian manifolds  $(M, g)$ , when the cost is the square of the Riemannian distance and the law of the samples is the normalized Riemannian volume measure. This however does not even cover the case of a uniform measure on a general convex set  $M = \Omega \subseteq \mathbb{R}^2$ , because of the presence of the boundary. For non-convex sets, an additional problem is that the Riemannian distance is different from the Euclidean one. On the other side, it was recently predicted in [9, Conjecture 2] that for every bounded connected open set  $\Omega \subset \mathbb{R}^2$  with smooth boundary and every smooth, uniformly positive and bounded density with respect to the Lebesgue measure, the limit (1.4) holds true with the right-hand side multiplied by  $|\Omega|$ . A rigorous proof of the upper bound

$$\limsup_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E} \left[ W_2^2 \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \frac{1}{n} \sum_{i=1}^n \delta_{Y_i} \right) \right] \leq \frac{|\Omega|}{2\pi}, \quad (1.5)$$

has been obtained in [9, Theorem 1] under the hypothesis that  $\Omega = (0, 1)^2$  and that the common law of the points is absolutely continuous with a uniformly strictly positive and bounded Lipschitz density. Further conjectures on higher order terms in the asymptotic expansion of (1.4) can be found in [10], relying on a clever comparison between the expansions in any two smooth compact manifolds. Unfortunately, a rigorous mathematical justification of these predictions seems currently out of reach.

**Main results.** Our main results fully settle the validity of [9, Conjecture 2], allowing in fact for densities that are not necessarily smooth, but only Hölder continuous and strictly positive on a bounded connected domain with Lipschitz boundary. We state and prove separately first the case of the quadratic Wasserstein distance between an empirical measure and the common law of the sampled points, and then that of optimal transport between two independent samples. In the former setting we have the following

**Theorem 1.1.** *Let  $\Omega \subseteq \mathbb{R}^2$  be a bounded connected domain, with Lipschitz boundary and  $\rho$  be a Hölder continuous probability density on  $\Omega$  uniformly strictly positive and bounded from above. Given i.i.d. random variables  $(X_i)_{i=1}^\infty$  with common distribution  $\rho$ , we have*

$$\lim_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E} \left[ W_2^2 \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \rho \right) \right] = \frac{|\Omega|}{4\pi}.$$

In the latter case, we also allow for possibly a different number of points  $n, m$  among two (jointly) i.i.d. families of random variables, extending [2, Remark 7.1].

**Theorem 1.2.** *Let  $\Omega \subseteq \mathbb{R}^2$  be a bounded connected domain with Lipschitz boundary and  $\rho$  be a Hölder continuous probability density on  $\Omega$  uniformly strictly positive and bounded from above. Given i.i.d. random variables  $(X_i, Y_i)_{i=1}^\infty$  with common distribution  $\rho$ , for every  $q \in [1, \infty]$ ,*

$$\lim_{\substack{n, m \rightarrow \infty \\ m/n \rightarrow q}} \frac{n}{\log n} \mathbb{E} \left[ W_2^2 \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \frac{1}{m} \sum_{j=1}^m \delta_{Y_j} \right) \right] = \frac{|\Omega|}{4\pi} \left( 1 + \frac{1}{q} \right).$$

Building on these results for Euclidean planar domains, it is possible using similar sub-additivity arguments to generalize them to some metric measure spaces, including e.g., smooth connected domains of two dimensional Riemannian manifolds with densities, see Section 6 for more precise statements.

Notice that when  $q \rightarrow \infty$  we obtain in Theorem 1.2 the same limit as in Theorem 1.1. When  $n = m$ , Theorem 1.2 gives the asymptotic value of the minimum bipartite matching. In fact, using simple upper and lower bounds we can obtain a related result for  $n, m \rightarrow \infty$  with  $m - n$  that does not grow too fast.

**Corollary 1.3.** *Let  $\Omega \subseteq \mathbb{R}^2$  be a bounded connected domain, with Lipschitz boundary and  $\rho$  be a Hölder continuous probability density on  $\Omega$  uniformly strictly positive and bounded from above. Given i.i.d. random variables  $(X_i, Y_i)_{i=1}^\infty$  with common distribution  $\rho$ , for any sequence  $m = m(n) \geq n$  with  $\lim_{n \rightarrow \infty} (m - n) / \log n = 0$ , it holds*

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \mathbb{E} \left[ \min_{\sigma \in \mathcal{S}_{n,m}} \sum_{i=1}^n |X_i - Y_{\sigma(i)}|^2 \right] = \frac{|\Omega|}{2\pi}.$$

We further deduce a result for the average cost of the random quadratic Euclidean bipartite travelling salesperson problem. The Euclidean travelling salesperson problem searches for the shortest cycle that visits a given set of points in  $\mathbb{R}^d$ , and the study of its random version is classical [8, 36, 50, 49]. Its random bipartite variant requires that the cycle must alternatively connect points from two given sets of points. By extending the argument from [14] from the case of the square to a general domain, we deduce that the cost of the random quadratic Euclidean bipartite travelling salesperson problem in two dimensions is asymptotically twice that of the bipartite matching problem.

**Corollary 1.4.** *Let  $\Omega \subseteq \mathbb{R}^2$  be a bounded connected domain, with Lipschitz boundary and  $\rho$  be a Hölder continuous probability density on  $\Omega$  uniformly strictly positive and bounded from above. Given i.i.d. random variables  $(X_i, Y_i)_{i=1}^\infty$  with common distribution  $\rho$ , then*

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \mathbb{E} \left[ \min_{\sigma, \tau \in \mathcal{S}_n} \sum_{i=1}^n |X_{\sigma(i)} - Y_{\tau(i)}|^2 + |Y_{\tau(i)} - X_{\sigma(i+1)}|^2 \right] = \frac{|\Omega|}{\pi},$$

where we use the convention  $\sigma(n+1) = \sigma(1)$ , for  $\sigma \in \mathcal{S}_n$ .

Finally, let us comment that our main focus on limit results for the expected costs is justified by general concentration arguments that can improve to convergence in probability of the sequence of renormalized cost – see [4, Remark 4.7], where the only assumption is the validity of a  $L^2$ -Poincaré-Wirtinger inequality, satisfied on such regular domains.

**Idea of the proof.** We follow the sub-additivity method introduced in this context in [7, 18] and later improved in [9, 22]. This splits the energy in a local and a global part. A central point is to prove that the global part, which encodes the defect in sub-additivity, has asymptotically only a vanishing contribution. While this method has already been implemented in [9] to obtain the upper bound (1.5) one of our main achievements is to prove that it may indeed be used also to get the corresponding lower bound. To this aim, we rely on a “boundary” variant of the optimal transport which has super-additivity properties. Similar functionals have been widely used in the theory of Euclidean additive functionals [7, 51]. The main point is to prove that when  $\Omega = (0, 1)^2$  and the law is the uniform one, the asymptotic costs for the usual optimal transport and the “boundary” versions are equal (see Proposition 3.1 and Proposition 3.2). Our proof relies on the PDE approach from [2, 4]. To the best of our knowledge, this is the first example where the so-called Dirichlet-Neumann bracketing method is used in the context of optimal transport (see [6] for a recent application of these ideas for Coulomb gases).

On a more technical side, we introduce two main ideas. First, in order to deal with domains which are not  $(0, 1)^2$  or a finite union of cube, we consider a Whitney partition of our domain. While this does not affect too much the treatment of the local term, it requires a finer estimate of the global one. Indeed, as in [9, 22] (see also [35, 4, 21, 26]) we first rely on the estimate

$$W_2^2(f, 1) \lesssim \|f - 1\|_{H^{-1}}^2$$

to reduce ourselves to an estimate in  $H^{-1}$ . In [9, 22] (and also [21]) the right-hand side is then estimated thanks to Poincaré inequality by an  $L^2$  norm. A quick look at [9] shows that for fixed number of points  $n$ , this yields an error which is proportional to the number of cubes in the partition of our domain. Since a Whitney partition is made of infinitely

many such cubes (or in any case a very large number of them, see Lemma 5.1) this would lead to a very bad estimate. This is due to the fact that a function with rapid oscillations typically has large  $L^2$  norm but small  $H^{-1}$  norm. To capture this, we prove in Lemma 2.1 a finer estimate for functions with “small” support. This gives an error term which, up to logarithm, does not depend on the mesh-size of our partition (see (4.10)). In order to deal with Hölder continuous densities instead of Lipschitz ones as in [9], the second idea is to replace the Knothe map used in [9, Lemma 1] by a transport map constructed via heat flow interpolation in the spirit of the Dacorogna-Moser construction (see Lemma 2.5). This might be of independent interest and may be related to similar constructions in the literature, e.g. [25, 33].

**Further questions.** We mention here some open problems related to our results:

- (1) In [3] and [20] quantitative rates of convergence for the optimal transport map have been obtained. It may be worth exploring whether similar quantitative rates still hold in the general setting addressed in this article.
- (2) We expect that our results may be extended to the case of less regular domains and densities, with a similar asymptotic cost. However, the validity of (1.5) must require some condition on the support of  $\rho$ , such as connectedness [9, Remark 2], which is in contrast with the universal upper bounds obtained by Talagrand [43] for the case  $p = 1$ .
- (3) The case of unbounded domains may be treated with similar techniques, possibly leading to sharper limit results for a wide class densities, e.g. Gaussian ones [26, 27, 45], where existence of a limit is an open question. We mention in this regard that the case  $d = 1$  already rises yields some interesting problems, see [12, 15, 11].
- (4) Finally, it may be worth exploring how large the difference  $m - n$  can be in Corollary 1.3 so that (1.3) still holds true.

**Structure of the paper.** In Section 2 we fix our notation and provide some general results about negative Sobolev norms, classical optimal transport and its “boundary” version and finally a construction of a transport map from a Hölder density to the uniform one via heat flow on the cube. Section 3 is devoted to the extension of the PDE approach from [2, 4] to the “boundary” transport cost on the square. In Section 4 and Section 5 we prove respectively Theorem 1.1 and Theorem 1.2. Section 6 describes how similar techniques allow to consider more general settings, including Riemannian manifolds. Section 7 and Section 8 contain respectively the proof of Corollary 1.3 and Corollary 1.4. Two appendices include technical results about decomposition of a domain into sets with certain properties that may be well-known in the literature, but we did not find exactly stated in a version useful for our purposes.

## 2. NOTATION AND PRELIMINARY RESULTS

We write  $|A|$  for the Lebesgue measure of a Borel set  $A \subseteq \mathbb{R}^d$ , and  $\int_A f$  for the integral of an integrable function  $f$  on  $A$ . If a measure is absolutely continuous with respect to the Lebesgue measure, we always identify it with its density. A cube  $Q \subseteq \mathbb{R}^d$  of side length  $\ell$  is a set of the form  $Q = \prod_{i=1}^d (v_i, v_i + \ell)$  with  $(v_i)_{i=1}^d \in \mathbb{R}^d$  and  $\ell > 0$ . We note  $\ell(Q)$  the sidelength of a cube  $Q$ . By a partition of  $\Omega \subseteq \mathbb{R}^d$ , we always mean in fact that  $\Omega$  is covered up to Lebesgue negligible sets. We denote by  $|v|$  the Euclidean norm of a vector  $v \in \mathbb{R}^d$ . We use the letters  $C, c$  for positive constants whose value may vary from one line to the next and  $\omega$  for a generic decreasing rate function with  $\lim_{t \rightarrow \infty} \omega(t) = 0$ . In this paper the domain  $\Omega$  is fixed and we write  $A \lesssim B$  if there is  $C > 0$  depending only on  $\Omega$  (and potentially on  $d$  and  $p$  if we consider the  $p$ -Wasserstein distance in  $\mathbb{R}^d$ ) such that  $A \leq CB$ . For a function  $f$ , we use the notation  $\nabla f$  for its gradient,  $\nabla \cdot f$  for its divergence,  $\Delta f = \nabla \cdot \nabla f$  for its Laplacian. The push-forward of a measure  $\mu$  with respect to a map  $f$  is denoted  $f_\# \mu$ , i.e.,  $f_\# \mu(A) = \mu(f^{-1}(A))$ .

For  $f : \Omega \subseteq \mathbb{R}^d \rightarrow \mathbb{R}^k$ , we write  $\|f\|_{L^p(\Omega)} = (\int_{\Omega} |f|^p)^{1/p}$ ,  $\|f\|_{L^\infty(\Omega)} = \sup_{x \in \Omega} |f(x)|$  and

$$\|f\|_{C^\alpha(\Omega)} = \|f\|_{L^\infty(\Omega)} + \sup_{\substack{x, y \in \Omega \\ x \neq y}} \frac{|f(y) - f(x)|}{|x - y|^\alpha},$$

for  $\alpha \in (0, 1]$  for the  $\alpha$ -Hölder norm of  $f$  (when  $\alpha = 1$  we obtain its Lipschitz norm). We also write

$$\text{Lip}_\Omega f = \sup_{\substack{x, y \in \Omega \\ x \neq y}} \frac{|f(y) - f(x)|}{|x - y|}.$$

We omit to specify  $\Omega$  when it is clear from the context and write  $\|f\|_p$ ,  $\|f\|_\infty$ ,  $\|f\|_{C^\alpha}$  and  $\text{Lip } f$ .

We collect below some general facts useful to prove our main results. For possible future reference we state and prove them in a slightly more general form, e.g., in general dimensions or for general, non quadratic, costs.

**2.1. Negative Sobolev norms.** Let  $\Omega \subseteq \mathbb{R}^d$  be a bounded connected open set with Lipschitz boundary. We denote the negative Sobolev norm by

$$\|f\|_{W^{-1,p}(\Omega)} = \sup_{|\nabla \phi|_{L^q(\Omega)} \leq 1} \int_{\Omega} f \phi,$$

where  $q$  is the Hölder conjugate of  $p$  (i.e.  $1 = \frac{1}{p} + \frac{1}{q}$ ). For  $p = 2$  we simply write  $H^{-1}(\Omega)$ . Notice in particular that in order to have  $\|f\|_{W^{-1,p}(\Omega)} < \infty$  we must have  $\int_{\Omega} f = 0$ . In this case we may also restrict the supremum to functions  $\phi$  having also average zero. When it is clear from the context, we will drop the explicit dependence on  $\Omega$  in the norms.

The heat semi-group with null Neumann boundary conditions on  $\Omega$  is well-defined as the symmetric Markov semi-group of operators  $(P_t)_{t \geq 0}$  arising as the  $L^2$  gradient flow of the Dirichlet energy  $\|\nabla f\|_2^2$  on the Sobolev space  $f \in H^1(\Omega)$ . It is well-known [47] that the validity of  $L^2$ -Poincaré-Wirtinger inequality on  $\Omega$  is equivalent to a spectral gap, i.e., for some constant  $c > 0$ , for every  $f \in L^2(\Omega)$  with  $\int_{\Omega} f = 0$ ,

$$\|P_t f\|_2 \leq e^{-ct} \|f\|_2, \quad \text{for } t \geq 0. \quad (2.1)$$

In particular, one has the integral representation

$$\Delta^{-1} f = \int_0^\infty P_t f dt$$

for the solution to the elliptic PDE,  $\Delta u = f$  in  $\Omega$  with null Neumann boundary conditions and also, for the negative Sobolev norm,

$$\|f\|_{H^{-1}}^2 = \int_{\Omega} f \Delta^{-1} f = \int_0^\infty \int_{\Omega} f P_t f dt = \int_0^\infty \int_{\Omega} (P_{t/2} f)^2 dt. \quad (2.2)$$

Sobolev inequality instead is equivalent to ultracontractivity: for every  $t \in [0, 1]$  and  $f \in L^1$  with  $\int_{\Omega} f = 0$ ,

$$\|P_t f\|_\infty \lesssim t^{-d/2} \|f\|_1. \quad (2.3)$$

In terms of the symmetric heat kernel  $p_t(x, y) = p_t(y, x)$  (defined by the equality  $P_t f(x) = \int_{\Omega} p_t(x, y) f(y) dy$ ), it reads  $\|p_t\|_\infty \lesssim t^{-d/2}$ . This inequality can be easily combined with (2.1) to obtain the stronger inequality

$$\|P_t f\|_\infty \lesssim t^{-d/2} e^{-ct} \|f\|_1. \quad (2.4)$$

The following lemma will be used to bound the negative Sobolev norm of functions supported on subsets  $A \subseteq \Omega$ .



**Lemma 2.1.** *Let  $d \geq 2$ ,  $\Omega \subseteq \mathbb{R}^d$  be a bounded connected Lipschitz domain. For every  $f \in L^\infty(\Omega)$  with  $\int_\Omega f = 0$ ,*

$$\|f\|_{H^{-1}} \lesssim \begin{cases} \|f\|_{L^1} \sqrt{\left| \log \frac{\|f\|_1}{\|f\|_\infty} \right| + 1} & \text{if } d = 2, \\ \|f\|_1^{\frac{1}{2} + \frac{1}{d}} \|f\|_\infty^{\frac{1}{2} - \frac{1}{d}} & \text{if } d > 2. \end{cases}$$

*Proof.* By (2.2)

$$\|f\|_{H^{-1}}^2 = \int_0^\infty \int_\Omega f P_t f dt.$$

By (2.4) and  $\|P_t f\|_\infty \leq \|f\|_\infty$ , for every  $t > 0$ ,

$$\int_\Omega f P_t f \leq \|f\|_1 \|P_t f\|_\infty \lesssim \|f\|_1 \min(\|f\|_\infty, t^{-\frac{d}{2}} \exp(-ct) \|f\|_1).$$

By integration we obtain for every  $t_0 > 0$ ,

$$\|f\|_{H^{-1}}^2 \lesssim \|f\|_1 \left( t_0 \|f\|_\infty + \|f\|_1 \begin{cases} |\log t_0| & \text{if } d = 2, \\ t_0^{1-d/2} & \text{if } d > 2 \end{cases} \right).$$

Optimizing in  $t_0$  by setting  $t_0 = \|f\|_1^{d/2} \|f\|_\infty^{-d/2}$  concludes the proof.  $\square$

**2.2. Optimal transport.** 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 [48, 38].

Let  $p \geq 1$  and  $\mu, \lambda$  be positive Borel measures with finite  $p$ -th moments and equal mass  $\mu(\mathbb{R}^d) = \lambda(\mathbb{R}^d)$ . The Wasserstein distance of order  $p$  between  $\mu$  and  $\lambda$  is defined as

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

where  $\mathcal{C}(\mu, \lambda)$  is the set of couplings between  $\mu$  and  $\lambda$ . For a Borel subset  $\Omega \subseteq \mathbb{R}^d$ , we also write

$$W_{\Omega, p}(\mu, \lambda) = W_p(\mu \llcorner \Omega, \lambda \llcorner \Omega),$$

which implicitly assumes that  $\mu(\Omega) = \lambda(\Omega)$ .

Let us recall that  $W_p$  is a distance, in particular the triangle inequality holds:

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

We will also use the classical sub-additivity inequality

$$W_p^p \left( \sum_k \mu_k, \sum_k \lambda_k \right) \leq \sum_k W_p^p(\mu_k, \lambda_k), \quad (2.5)$$

for a finite set of positive measures  $\mu_k, \lambda_k$ . A simple combination of the two properties above and Young inequality yield a geometric subadditivity property [22, Lemma 3.1], that we report here for the reader's convenience: there exists a constant  $C = C(p) > 0$  such that, for every Borel partition  $(\Omega_k)_{k \in \mathbb{N}}$  of  $\Omega$ , and every  $\varepsilon \in (0, 1)$ ,

$$W_{\Omega, p}^p \left( \mu, \frac{\mu(\Omega)}{\lambda(\Omega)} \lambda \right) \leq (1 + \varepsilon) \sum_k W_{\Omega_k, p}^p \left( \mu, \frac{\mu(\Omega_k)}{\lambda(\Omega_k)} \lambda \right) + \frac{C}{\varepsilon^{p-1}} W_{\Omega, p}^p \left( \sum_k \frac{\mu(\Omega_k)}{\lambda(\Omega_k)} \chi_{\Omega_k} \lambda, \frac{\mu(\Omega)}{\lambda(\Omega)} \lambda \right). \quad (2.6)$$

If we assume that  $\Omega$  is sufficiently regular, then we may use the Benamou-Brenier formula in a similar fashion as in [22, Lemma 3.4] (see also [35, Corollary 3]) to bound from above the Wasserstein distance by the negative Sobolev norm. Notice that we use the fact that the Euclidean distance is bounded from above by the geodesic distance in  $\Omega$  (they coincide if  $\Omega$  is convex).

**Lemma 2.2.** *Assume that  $\Omega \subseteq \mathbb{R}^d$  is a bounded connected open set with Lipschitz boundary. If  $\mu$  and  $\lambda$  are measures on  $\Omega$  with  $\mu(\Omega) = \lambda(\Omega)$ , absolutely continuous with respect to the Lebesgue measure and  $\inf_{\Omega} \lambda > 0$ , then, for every  $p \geq 1$ ,*

$$W_{\Omega,p}^p(\mu, \lambda) \lesssim \frac{1}{\inf_{\Omega} \lambda^{p-1}} \|\mu - \lambda\|_{W^{-1,p}(\Omega)}^p. \quad (2.7)$$

To deal with lower bounds, we rely on a “boundary” variant of the optimal transport, introduced in [19], but also independently studied in the setting of Euclidean bipartite matching problems, see [7, 18]. Here, one is allowed to transport any amount of mass from and to the boundary of an open set  $\Omega \subseteq \mathbb{R}^d$ . We write, for finite positive measures  $\mu$  and  $\nu$  with finite  $p$ -th moment,

$$Wb_{\Omega,p}^p(\mu, \lambda) = \left( \min_{\pi \in \mathcal{Cb}_{\Omega}(\mu, \lambda)} \int_{\overline{\Omega} \times \overline{\Omega}} |x - y|^p d\pi(x, y) \right)^{\frac{1}{p}},$$

where  $\mathcal{Cb}_{\Omega}(\mu, \lambda)$  is the set of positive measures  $\pi$  on  $\overline{\Omega} \times \overline{\Omega}$  such that  $\pi_1 \llcorner \Omega = \mu \llcorner \Omega$  and  $\pi_2 \llcorner \Omega = \lambda \llcorner \Omega$  ( $\pi_1, \pi_2$  denoting respectively the first and second marginals of  $\pi$ ).

The triangle inequality holds,

$$Wb_{\Omega,p}(\mu, \nu) \leq Wb_{\Omega,p}(\mu, \lambda) + Wb_{\Omega,p}(\nu, \lambda).$$

A geometric superadditivity property also holds: for every disjoint family  $(\Omega_k)_{k \in \mathbb{N}}$  of open subsets of  $\Omega$ ,

$$Wb_{\Omega,p}^p(\mu, \nu) \geq \sum_k Wb_{\Omega_k,p}^p(\mu, \nu).$$

The intuition behind this property is quite clear. Given any plan on  $\Omega$ , one should be able to suitably restrict it on each  $\Omega_k$  by stopping the transport at each boundary. To prove it rigorously, it is sufficient to argue in the case of discrete measures  $\mu = \sum_i \mu_i \delta_{x_i}$ ,  $\nu = \sum_i \nu_i \delta_{y_i}$ . The general case follows by approximation. Given  $\pi \in \mathcal{Cb}_{\Omega,p}(\mu, \nu)$ , we define  $(\pi_k)_{k \in \mathbb{N}}$  with  $\pi_k \in \mathcal{Cb}_{\Omega_k,p}(\mu, \nu)$  such that

$$\int_{\overline{\Omega} \times \overline{\Omega}} |x - y|^p d\pi(x, y) \geq \sum_k \int_{\overline{\Omega_k} \times \overline{\Omega_k}} |x - y|^p d\pi_k(x, y). \quad (2.8)$$

Again by approximation, we may also assume that  $\pi = \sum_{\ell} \pi_{\ell} \delta_{(x_{\ell}, y_{\ell})}$  is discrete, with  $(x_{\ell}, y_{\ell}) \in \overline{\Omega} \times \overline{\Omega}$ . The following algorithm can be used to define the sequence  $(\pi_k)_k$ . Set initially  $\pi_k = 0$  for every  $k$ . For every  $\ell$ , consider the pair  $(x_{\ell}, y_{\ell})$ . If  $x_{\ell}, y_{\ell} \in \Omega_k$ , then add  $\pi_{\ell} \delta_{(x_{\ell}, y_{\ell})}$  to  $\pi_k$ . If  $x_{\ell} \in \Omega_k$  and  $y_{\ell} \in \Omega_j$  with  $k \neq j$ , let

$$\begin{aligned} t^- &= \inf \{t \in [0, 1] : (1 - t)x_{\ell} + ty_{\ell} \notin \Omega_k\} \\ t^+ &= \inf \{t \in [0, 1] : tx_{\ell} + (1 - t)y_{\ell} \notin \Omega_j\}, \end{aligned}$$

and define  $y_{\ell}^- = (1 - t^-)x_{\ell} + t^-y_{\ell} \in \partial\Omega_k$ ,  $x_{\ell}^+ = t^+x_{\ell} + (1 - t^+)y_{\ell} \in \partial\Omega_j$ . Then, add respectively  $\pi_{\ell} \delta_{(x_{\ell}, y_{\ell}^-)}$  to  $\pi_k$  and  $\pi_{\ell} \delta_{(x_{\ell}^+, y_{\ell})}$  to  $\pi_j$ . If  $x \in \Omega_k$  for some  $k$  and  $y \in \overline{\Omega} \setminus \bigcup_j \Omega_j$ , set  $t^-$  and  $y_{\ell}^-$  as the previous case and add  $\pi_{\ell} \delta_{(x_{\ell}, y_{\ell}^-)}$  to  $\pi_k$ . Similarly, if  $x \in \overline{\Omega} \setminus \bigcup_j \Omega_j$  and  $y \in \Omega_k$  then add  $\pi_{\ell} \delta_{(x_{\ell}^+, y)}$  to  $\pi_k$ . In all the other cases, i.e., if  $x_{\ell}, y_{\ell} \in \overline{\Omega} \setminus \bigcup_k \Omega_k$ , do nothing. It is not difficult to check that  $(\pi_k)_k$  thus defined indeed satisfies  $\pi_k \in \mathcal{Cb}_{\Omega_k}$  and (2.8) holds (using in particular that  $p \geq 1$  to argue that  $|x_{\ell} - y_{\ell}|^p \geq |x_{\ell} - y_{\ell}^-|^p + |x_{\ell}^+ - y_{\ell}|^p$ ).

Arguing similarly as in the proof of (2.6) we obtain a symmetric inequality: there exists a constant  $C(p) > 0$  such that, for every Borel partition  $(\Omega_k)_{k \in \mathbb{N}}$  of  $\Omega$ , and every  $\varepsilon \in (0, 1)$ ,

$$\begin{aligned} Wb_{\Omega,p}^p\left(\mu, \frac{\mu(\Omega)}{\lambda(\Omega)}\lambda\right) &\geq (1 - \varepsilon) \sum_k Wb_{\Omega_k,p}^p\left(\mu, \frac{\mu(\Omega_k)}{\lambda(\Omega_k)}\lambda\right) \\ &\quad - \frac{C}{\varepsilon^{p-1}} Wb_{\Omega,p}^p\left(\sum_k \frac{\mu(\Omega_k)}{\lambda(\Omega_k)}\chi_{\Omega_k}\lambda, \frac{\mu(\Omega)}{\lambda(\Omega)}\lambda\right). \end{aligned} \quad (2.9)$$



To obtain lower bounds, we will also rely on the dual formulation of  $Wb_{\Omega,p}$  that reads

$$Wb_{\Omega,p}^p(\mu, \nu) = \sup \left\{ \int_{\Omega} f d\mu + \int_{\Omega} g d\nu : f(x) + g(y) \leq |x - y|^p \text{ for } x, y \in \Omega, f = g = 0 \text{ on } \partial\Omega \right\}. \quad (2.10)$$

In fact, only the inequality  $\geq$  will be used, which is immediate to check.

**Remark 2.3.** Many of the above properties, in particular (2.6) and (2.9), hold for the Wasserstein distance defined over any length metric space, in particular on a Riemannian manifold.

**2.3. A transport map via heat flow.** The following result yields a Lipschitz map on the cube  $(0, 1)^d$  transporting a given Hölder probability density to the uniform one. The main point here is that the Lipschitz constant of the map is very close to 1. This construction, possibly interesting on its own, allows us to avoid the use of general boundary regularity theory for the optimal transport map with respect to the squared distance cost [13], as it does not seem to be applicable to the case of the cube (it requires strictly convex and  $C^2$  boundary). Notice that for Hölder continuous densities, counterexamples can be constructed if we work with the Knothe map as in [9].

**Proposition 2.4.** *For  $d \geq 1$ ,  $\alpha \in (0, 1]$ , there exists  $C > 0$  depending on  $d$  and  $\alpha$  only such that the following holds: for any  $\rho : (0, 1)^d \rightarrow (0, \infty)$  with*

$$\int_{(0,1)^d} \rho = 1 \quad \text{and} \quad \|\rho - 1\|_{C^\alpha} \leq 1/2,$$

*there exists  $T : (0, 1)^d \rightarrow (0, 1)^d$  such that  $T_\# \rho = 1$ , with*

$$\text{Lip } T, \text{Lip } T^{-1} \leq 1 + C \|\rho - 1\|_{C^\alpha} \quad \text{and} \quad T(\partial(0, 1)^d) = \partial(0, 1)^d.$$

*Proof.* Using the Neumann heat semi-group, we define  $\rho_t = P_t \rho$ , so that, for  $t \geq 0$ , the weak formulation of the heat equation reads, for every  $f \in H^1((0, 1)^d)$ ,

$$\int_{(0,1)^d} f \rho_t = \int_{(0,1)^d} f \rho - \int_0^t \int_{(0,1)^d} \nabla f \nabla \rho_t dt, \quad \text{for every } t \geq 0.$$

We notice first that the assumption gives  $\inf_{(0,1)^d} \rho \geq 1/2$  so that, for every  $t \geq 0$ ,  $\inf_{(0,1)^d} \rho_t \geq 1/2$  as well, hence

$$\|\rho_t^{-1}\|_{\infty} \lesssim 1.$$

By standard heat kernel estimates (one has in fact an explicit representation of  $p_t(x, y)$ , see [2, Appendix A] for the case  $d = 2$ ) we have, for every  $t > 0$ ,

$$\|\nabla \rho_t\|_{\infty} \lesssim t^{-\frac{1}{2} + \frac{\alpha}{2}} e^{-ct} \|\rho - 1\|_{C^\alpha} \quad \text{and} \quad \|\nabla^2 \rho_t\|_{\infty} \lesssim t^{-1 + \frac{\alpha}{2}} e^{-ct} \|\rho - 1\|_{C^\alpha}.$$

We thus define the time-dependent vector field  $(b_t)_{t>0}$ , on  $(0, 1)^d$ ,

$$b_t(x) = -\nabla \log \rho_t(x) = -\frac{\nabla \rho_t(x)}{\rho_t(x)}, \quad \text{with} \quad \nabla b_t(x) = -\frac{\nabla^2 \rho_t(x)}{\rho_t(x)} + \frac{\nabla \rho_t(x) \otimes \nabla \rho_t(x)}{\rho_t^2(x)}.$$

Using the previous estimates, it follows that

$$\|b_t\|_{\infty} \lesssim t^{-\frac{1}{2} + \frac{\alpha}{2}} e^{-ct} \|\rho - 1\|_{C^\alpha} \quad \text{and} \quad \|\nabla b_t\|_{\infty} \lesssim t^{-1 + \frac{\alpha}{2}} e^{-ct} \|\rho - 1\|_{C^\alpha}.$$

Since the right-hand sides are integrable functions with respect to  $t \in (0, \infty)$ , standard arguments yield that the associated flow  $X(t, x)$ , i.e., the solution to

$$X(t, x) = x + \int_0^t b_s(X(s, x)) ds \quad \text{for } t \geq 0, x \in (0, 1)^d$$

is well-defined and Lipschitz continuous:

$$\text{Lip } X(t, \cdot) \leq \exp \left( \int_0^t \|\nabla b_s\|_\infty ds \right) \leq \exp \left( Ct^{\frac{\alpha}{2}} e^{-ct} \|\rho - 1\|_{C^\alpha} \right). \quad (2.11)$$

Because of the homogeneous Neumann boundary condition,  $b_t$  is tangent to the boundary  $\partial(0, 1)^d$ , hence  $X(t, x) \in (0, 1)^d$  for every  $t \geq 0$ . Moreover,  $\lim_{t \rightarrow \infty} X(t, x) \in (0, 1)^d$  exists for every  $x \in (0, 1)^d$ , since, for  $s \leq t$ ,

$$|X(t, x) - X(s, x)| \leq \int_s^t \|b_r\|_\infty dr \lesssim s^{\frac{\alpha}{2}} e^{-cs} \|\rho - 1\|_{C^\alpha},$$

which yields that  $t \mapsto X(t, x)$  is Cauchy as  $t \rightarrow \infty$ . We thus define  $T(x) = \lim_{t \rightarrow \infty} X(t, x)$ . Letting  $t \rightarrow \infty$  in (2.11), we have

$$\text{Lip } T \leq \exp(C \|\rho - 1\|_{C^\alpha}) \leq 1 + C \|\rho - 1\|_{C^\alpha}.$$

To prove that  $T$  is invertible, we simply notice that for every  $t > 0$  the inverse map of  $x \mapsto X(t, x)$  is given by  $y \mapsto Y^t(t, y)$ ,  $Y^t$  being the flow of the “backward” in time vector field  $(b_s^t)_{s \in [0, t]}$ ,  $b_s^t(y) = -b_{(t-s)}(y)$ . Writing the analogue of (2.11) for  $y \mapsto Y^t(t, y)$  and letting  $t \rightarrow \infty$  yields the claimed bound on  $\text{Lip } T^{-1}$ .

To conclude, we need to argue that  $T_\# \rho = 1$ . This follows quite classically from the fact that  $\tilde{\rho}_t = X(t, \cdot)_\# \rho$  and  $\rho_t$  both solve the same continuity equation for the vector field  $(b_t)_{t \geq 0}$ . Since  $b_t$  is locally Lipschitz continuous, with Lipschitz norm that is integrable in time, uniqueness holds for the Cauchy problem (this follows for example by a slight adaptation of [38, Theorem 4.4]). Therefore,  $\rho_t = \tilde{\rho}_t$  and  $T_\# \rho = \lim_{t \rightarrow \infty} \tilde{\rho}_t = \lim_{t \rightarrow \infty} \rho_t = 1$ .  $\square$

In general, if  $\Omega \subseteq \mathbb{R}^d$  is open and  $T : \Omega \rightarrow \Omega$  is Lipschitz, then for any pair of measures  $\mu, \lambda$  with  $\mu(\Omega) = \lambda(\Omega)$ ,

$$W_{\Omega, p}(T_\#(\mu \llcorner \Omega), T_\#(\lambda \llcorner \Omega)) \leq (\text{Lip } T) W_{\Omega, p}(\mu, \lambda), \quad (2.12)$$

since any coupling  $\pi$  between  $\mu \llcorner \Omega$  and  $\lambda \llcorner \Omega$  induces the coupling  $(T, T)_\# \pi$ . For the “boundary” optimal transport a similar inequality holds provided that  $T : \bar{\Omega} \rightarrow \bar{\Omega}$  is such that  $T(\Omega) \subseteq \Omega$  and  $T(\partial\Omega) \subseteq \partial\Omega$ :

$$Wb_{\Omega, p}(T_\#(\mu \llcorner \Omega), T_\#(\lambda \llcorner \Omega)) \leq (\text{Lip } T) Wb_{\Omega, p}(\mu, \lambda). \quad (2.13)$$

Combining these observations with Proposition 2.4, we obtain the following result for the bipartite matching problem, extending [9, Lemma 1] to the case of Hölder continuous densities.

**Lemma 2.5.** *Let  $d \geq 1$ ,  $\alpha \in (0, 1]$ ,  $p \geq 1$  and let  $\rho$  be a uniformly positive and  $\alpha$ -Hölder continuous function on a domain  $\Omega \subseteq \mathbb{R}^d$ . Then, there exists a constant  $c = C(p, \rho) > 0$  such that, for every cube  $Q \subseteq \Omega$  with side length  $r < c$ , the following holds:*

- (1) *if  $(X_i)_{i=1}^n$  are i.i.d. on  $Q$  with common density  $\rho/\rho(Q)$  and  $(X_i^r)_{i=1}^n$  are i.i.d. uniformly distributed on  $[0, r]^d$ , then*

$$\left| \mathbb{E} \left[ W_*^p \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \frac{\rho}{\rho(Q)} \right) \right] - \mathbb{E} \left[ W_*^p \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i^r}, \frac{1}{|Q|} \right) \right] \right| \lesssim r^\alpha \mathbb{E} \left[ W_*^p \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i^r}, \frac{1}{|Q|} \right) \right], \quad (2.14)$$

where  $W_* \in \{W_{Q, p}, Wb_{Q, p}\}$ ;

(2) with the same notation, if  $(X_i)_{i=1}^n, (Y_j)_{j=1}^m$  are (jointly) i.i.d. on  $Q$  with common density  $\rho/\rho(Q)$  and  $(X_i^r)_{i=1}^n, (Y_j^r)_{j=1}^m$  are (jointly) i.i.d. and uniformly distributed on  $[0, r]^d$ , then

$$\left| \mathbb{E} \left[ W_*^p \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \frac{1}{m} \sum_{j=1}^m \delta_{Y_j} \right) \right] - \mathbb{E} \left[ W_*^p \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i^r}, \frac{1}{m} \sum_{j=1}^m \delta_{Y_j^r} \right) \right] \right| \lesssim r^\alpha \mathbb{E} \left[ W_*^p \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i^r}, \frac{1}{m} \sum_{j=1}^m \delta_{Y_j^r} \right) \right]. \quad (2.15)$$

*Proof.* The proof is identical to the proof of [9, Lemma 1] but we include it for the reader's convenience. We only prove that for  $r$  small enough,

$$(1 - Cr^\alpha) \mathbb{E} \left[ W_*^p \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i^r}, \frac{1}{|Q|} \right) \right] \leq \mathbb{E} \left[ W_*^p \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \frac{\rho}{\rho(Q)} \right) \right]$$

since the other inequalities may be proven in the same way. Without loss of generality we assume  $Q = (0, r)^d$ . Let also  $\bar{\rho} = \min_Q \rho$ . We define  $\rho^r(x) = \rho(rx)r^d/\rho(Q)$  for  $x \in (0, 1)^d$ , so that  $\int_{(0,1)^d} \rho^r = 1$ , and for every  $x, y \in (0, 1)^d$ ,

$$\rho^r(x) - \rho^r(y) \leq \frac{\|\rho\|_{C^\alpha}}{\bar{\rho}} r^\alpha |x - y|^\alpha,$$

thus  $\|\rho^r - 1\|_{C^\alpha} \leq 1/2$  if  $r$  is sufficiently small. We define  $S : \bar{Q} \rightarrow \bar{Q}$  as  $S(x) = rT(x/r)$ , where  $T$  is the map provided by Proposition 2.4. It holds  $\text{Lip } S = \text{Lip } T$  and  $\text{Lip } S^{-1} = \text{Lip } T^{-1}$ ,  $S_\# \rho/\rho(Q) = 1/|Q|$  is uniform, hence the variables  $(S(X_i))_{i=1}^n$  have the same law as  $(X_i^r)_{i=1}^n$ . Moreover,

$$S_\# \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i} \right) = \frac{1}{n} \sum_{i=1}^n \delta_{S(X_i)}$$

has the same law as  $\frac{1}{n} \sum_{i=1}^n \delta_{X_i^r}$ . Since  $S(\partial Q) = \partial Q$  and  $S(Q) = Q$ , it follows from (2.12) and (2.13) that

$$\mathbb{E} \left[ W_*^p \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i^r}, \frac{1}{|Q|} \right) \right] \leq (\text{Lip } T)^p \mathbb{E} \left[ W_*^p \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \frac{\rho}{\rho(Q)} \right) \right],$$

This proves the claim provided  $r$  is small enough so that  $(\text{Lip } T)^p \leq (1 - Cr^\alpha)^{-1}$ .  $\square$

### 3. ASYMPTOTIC EQUIVALENCE OF THE USUAL AND BOUNDARY COSTS FOR UNIFORM POINTS ON THE SQUARE

In this section we show how to modify the proofs from [4], using also ideas from [2], to obtain the analogue for  $Wb$  of the limits

$$\lim_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E} \left[ W_2^2 \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, 1 \right) \right] = \frac{1}{4\pi}, \quad (3.1)$$

and

$$\lim_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E} \left[ W_2^2 \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \frac{1}{n} \sum_{i=1}^n \delta_{Y_i} \right) \right] = \frac{1}{2\pi},$$

where  $(X_i, Y_i)_{i=1}^\infty$  are i.i.d. uniformly distributed on  $(0, 1)^d$ . We deal first with the simpler case of matching to the reference measure.

**Proposition 3.1.** *Let  $(X_i)_{i=1}^\infty$  be i.i.d. uniformly distributed on  $(0, 1)^2$ . Then*

$$\lim_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E} \left[ Wb_{(0,1)^2,2}^2 \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, 1 \right) \right] = \frac{1}{4\pi}.$$

*Proof.* To simplify the notation, we write  $Wb$  instead of  $Wb_{(0,1)^2,2}$  and  $W$  instead of  $W_{(0,1)^2,2}$ . Since  $Wb \leq W$ , it is sufficient to prove that

$$\liminf_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E} \left[ Wb^2 \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, 1 \right) \right] \geq \frac{1}{4\pi}. \quad (3.2)$$

We do it in several steps.

*Step 1 (Regularization).* Write  $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ , and for  $t \geq 0$ ,

$$\mu_n^t(x) = P_t \mu_n = \frac{1}{n} \sum_{i=1}^n p_t(x, X_i),$$

where we recall that  $(P_t)_{t \geq 0}$  denotes the Neumann heat semi-group on  $(0, 1)^d$  with kernel  $p_t(x, y)$ . The triangle inequality for  $Wb$  and the inequality  $Wb \leq W$  give

$$\begin{aligned} Wb^2(\mu_n^t, 1) &\leq (Wb(\mu_n, 1) + W(\mu_n, \mu_n^t))^2 \\ &\leq Wb^2(\mu_n, 1) + W(\mu_n, \mu_n^t) (W(\mu_n, \mu_n^t) + 2W(\mu_n, 1)). \end{aligned}$$

We choose  $t = (\log n)^4/n$ , so that, the refined contractivity estimate [2, Theorem 5.2] gives

$$\mathbb{E} [W^2(\mu_n, \mu_n^t)] \lesssim \frac{\log \log n}{n}.$$

Together with the upper bound

$$\mathbb{E} [W^2(\mu_n, 1)] \lesssim \frac{\log n}{n},$$

which follows from (3.1), we obtain that for  $n$  large enough,

$$\mathbb{E} [Wb^2(\mu_n, 1)] \geq \mathbb{E} [Wb^2(\mu_n^t, 1)] - C \frac{\sqrt{\log n \log \log n}}{n}. \quad (3.3)$$

Thus, it is sufficient to prove (3.2) with  $\mu_n^t$  instead of  $\mu_n$ .

*Step 2 (PDE and energy estimates).* Let  $f_n^t$  be the solution to  $-\Delta f_n^t = \mu_n^t - 1$  with null Neumann boundary conditions, i.e., since  $P_s \mu_n^t = \mu_n^{n,t+s}$ ,

$$f_n^t = \int_t^\infty (\mu_n^s - 1) ds.$$

Recall that, as proved in [4] and [2, Lemma 3.14],

$$\mathbb{E} \left[ \int_{(0,1)^2} |\nabla f_n^t|^2 \right] = \frac{\log n}{4\pi n} + O\left(\frac{1}{n}\right) \quad (3.4)$$

and

$$\mathbb{E} \left[ \int_{(0,1)^2} |\nabla f_n^t|^4 \right] \lesssim \left( \frac{\log n}{n} \right)^2. \quad (3.5)$$

In addition to these gradient estimates, we will also need the upper bounds

$$\mathbb{E} \left[ \int_{(0,1)^2} |f_n^t|^2 \right] \lesssim \frac{1}{n}, \quad \text{and} \quad \mathbb{E} \left[ \int_{(0,1)^2} |f_n^t|^4 \right] \lesssim \frac{1}{n^2}, \quad (3.6)$$

which can be proved in a similar way. Indeed, for every  $x \in (0, 1)^2$ ,

$$\mu_n^s(x) - 1 = \frac{1}{n} \sum_{i=1}^n (p_s(x, X_i) - 1) = \frac{1}{n} \sum_{i=1}^n \xi_i$$

is a sum of centered i.i.d. random variables  $(\xi_i)_{i=1}^n$ . For  $s \in (0, 1)$ , we bound from above

$$\mathbb{E} [\xi_i^2] = \int_{(0,1)^2} p_s^2(x, y) - 1 = p_{2s}(x, x) - 1 \lesssim s^{-1}$$

where we used the ultracontractivity (2.3) for the heat kernel and similarly

$$\mathbb{E} [\xi_i^4] \lesssim \int_{(0,1)^2} p_s^4(x, y) + 1 \lesssim s^{-2} \int_{(0,1)^2} p_s^2(x, y) + 1 \lesssim s^{-3}.$$

This yields immediately that

$$\int_{(0,1)^2} \mathbb{E} [(\mu_n^s(x) - 1)^2] \lesssim \frac{1}{ns}.$$

Using Rosenthal's inequality [37] (see also [26]) we also get

$$\int_{(0,1)^2} \mathbb{E} [(\mu_n^s(x) - 1)^4] \lesssim \frac{1}{n^3 s^3} + \frac{1}{n^2 s^2}.$$

For  $s > 1$ , we use (2.4), and  $p \in \{2, 4\}$ ,

$$\begin{aligned} \|\mu_n^s - 1\|_p &\leq \|\mu_n^s - 1\|_\infty \lesssim e^{-cs/2} s^{-d/2} \left\| \mu_n^{s/2} - 1 \right\|_1 \leq e^{-cs/2} s^{-d/2} \left\| \mu_n^{s/2} - 1 \right\|_p \\ &\lesssim e^{-cs/2} \|\mu_n^{1/2} - 1\|_p. \end{aligned}$$

Taking expectation and using also the upper bounds above for  $s = 1/2$ , it follows that for some constant  $c > 0$ ,

$$\int_{(0,1)^2} \mathbb{E} [(\mu_n^s(x) - 1)^p] \lesssim \frac{e^{-cs}}{n^{p/2}}.$$

For  $p = 2$ , we obtain

$$\int_t^\infty \mathbb{E} [(\mu_n^s(x) - 1)^2]^{1/2} \lesssim \int_t^1 \frac{1}{\sqrt{ns}} ds + \int_1^\infty \frac{e^{-cs/2}}{\sqrt{n}} ds \lesssim \frac{1}{\sqrt{n}},$$

and, for  $p = 4$ ,

$$\int_t^\infty \mathbb{E} [(\mu_n^s(x) - 1)^4]^{1/4} \lesssim \int_t^1 \left( \frac{1}{n^3 s^3} + \frac{1}{n^2 s^2} \right)^{1/4} ds + \int_1^\infty \frac{e^{-cs/4}}{\sqrt{n}} ds \lesssim \frac{1}{\sqrt{n}}.$$

We conclude, for  $p = 2$  that

$$\begin{aligned} \mathbb{E} \left[ \int_{(0,1)^2} |f_n^t|^2 \right] &= \int_t^\infty \int_t^\infty \mathbb{E} \left[ \int_{(0,1)^2} (\mu^{n,s_1} - 1)(\mu^{n,s_2} - 1) \right] ds_1 ds_2 \\ &\leq \int_t^\infty \int_t^\infty \mathbb{E} \left[ \int_{(0,1)^2} (\mu^{n,s_1}(x) - 1)^2 \right]^{1/2} \mathbb{E} \left[ \int_{(0,1)^2} (\mu^{n,s_2}(x) - 1)^2 \right]^{1/2} ds_1 ds_2, \\ &\leq \left( \int_t^\infty \mathbb{E} \left[ \int_{(0,1)^2} (\mu_n^s(x) - 1)^2 \right]^{1/2} ds \right)^2 \lesssim \frac{1}{n}, \end{aligned}$$

and similarly for  $p = 4$ ,

$$\mathbb{E} \left[ \int_{(0,1)^2} |f_n^t|^4 \right] \leq \left( \int_t^\infty \mathbb{E} \left[ \int_{(0,1)^2} (\mu_n^s(x) - 1)^4 \right]^{1/4} ds \right)^4 \lesssim \frac{1}{n^2}.$$

*Step 3 (Dual potential).* For  $\eta \in (0, 1)$  we introduce a smooth cut-off function  $\chi_\eta : (0, 1)^2 \rightarrow [0, 1]$  such that  $\chi_\eta(x) = 1$  on  $[\eta, 1 - \eta]^2$ ,  $\chi_\eta(x) = 0$  on  $(0, 1)^2 \setminus [\eta/2, 1 - \eta/2]^2$ , with

$$\|\nabla \chi_\eta\|_\infty \lesssim \eta^{-1}, \quad \|\nabla^2 \chi_\eta\|_\infty \lesssim \eta^{-2}.$$

The function  $f_n = f_n^t \chi_\eta$  can be extended by periodicity to a function on the flat torus  $\mathbb{T}^2$ . By [4, Corollary 3.3] applied to  $\mathbb{T}^2$ , there exists  $g_n \in C(\mathbb{T}^2)$  such that for  $x, y \in \mathbb{T}^2$ ,

$$f_n(x) + g_n(y) \leq \frac{d_{\mathbb{T}^2}(x, y)^2}{2}, \quad \text{and} \quad \int_{\mathbb{T}^2} (f_n + g_n) \geq -e^{\|\Delta f_n\|_\infty} \int_{\mathbb{T}^2} |\nabla f_n|^2.$$

We may naturally interpret  $g_n$  as being defined on  $(0, 1)^2$ , so that using  $d_{\mathbb{T}^2}(x, y) \leq |x - y|^2$ , we have for  $x, y \in (0, 1)^2$ ,

$$f_n(x) + g_n(y) \leq \frac{|x - y|^2}{2}.$$

To prove that  $g_n(x) = 0$  on  $\partial(0, 1)^2$ , we recall that [4, Remark 3.4]  $g_n$  is also the unique viscosity solution to the Hamilton-Jacobi equation  $\partial_t u + \frac{1}{2}|\nabla u|^2 = 0$  with initial condition  $-f^n$ , hence, it coincides with the Hopf-Lax solution

$$g_n(x) = \inf_{y \in (0, 1)^2} \left\{ -f_n(y) + \frac{d_{\mathbb{T}^2}(x, y)^2}{2} \right\}.$$

From this representation, we see at once that, if

$$\|f_n\|_\infty \leq \frac{\eta^2}{8} \tag{3.7}$$

then  $g_n(x) = 0$  on  $\partial(0, 1)^2$ , because that  $f_n(y) = f_n^t(y) \chi_\eta(y) = 0$  if  $x \in \partial(0, 1)^2$  and  $d_{\mathbb{T}^2}(x, y) < \eta/2$ .

Let us also notice that, since  $\nabla f_n = \nabla f_n^t \chi_\eta + f_n^t \nabla \chi_\eta$ , we have by Young and Cauchy-Schwarz

$$\begin{aligned} \int_{(0, 1)^2} |\nabla f_n - \nabla f_n^t|^2 &\lesssim \int_{(0, 1)^2} |\nabla f_n^t|^2 (1 - \chi_\eta)^2 + \int_{(0, 1)^2} |f_n^t \nabla \chi_\eta|^2 \\ &\lesssim \left( \int_{(0, 1)^2} |\nabla f_n^t|^4 \right)^{1/2} \eta^{1/2} + \int_{(0, 1)^2} |f_n^t|^2 \eta^{-2}, \end{aligned}$$

which in expectation gives by (3.5) and (3.6)

$$\mathbb{E} \left[ \int_{(0, 1)^2} |\nabla f_n - \nabla f_n^t|^2 \right] \lesssim \frac{\log n}{n} \eta^{1/2} + \frac{1}{n} \eta^{-2}. \tag{3.8}$$

Using (3.4) and Young inequality we find for every  $\varepsilon > 0$ ,

$$\limsup_{n \rightarrow \infty} \left| \frac{n}{\log n} \mathbb{E} \left[ \int_{(0, 1)^2} |\nabla f_n|^2 \right] - \frac{1}{4\pi} \right| \lesssim \varepsilon + \frac{C}{\varepsilon} \eta^{1/2}.$$

Choosing  $\varepsilon = \eta^{1/4}$  we conclude

$$\limsup_{n \rightarrow \infty} \left| \frac{n}{\log n} \mathbb{E} \left[ \int_{(0, 1)^2} |\nabla f_n|^2 \right] - \frac{1}{4\pi} \right| \lesssim \eta^{1/4}. \tag{3.9}$$

Arguing similarly, we have

$$\left( \frac{n}{\log n} \right)^2 \mathbb{E} \left[ \int |\nabla f_n|^4 \right] \lesssim 1 + \frac{1}{\eta^4 (\log n)^2}. \tag{3.10}$$

Let us finally argue that

$$\|f_n^t\|_\infty + \|\nabla f_n^t\|_\infty + \|\Delta f_n^t\|_\infty \lesssim \|\nabla^2 f_n^t\|_\infty. \tag{3.11}$$

Since the estimate  $\|\Delta f_n^t\|_\infty \lesssim \|\nabla^2 f_n^t\|_\infty$  is clear and  $\|f_n^t\|_\infty \lesssim \|\nabla f_n^t\|_\infty$  follows from  $\int_{(0, 1)^2} f_n^t = 0$ , we are left with the proof of  $\|\nabla f_n^t\|_\infty \lesssim \|\nabla^2 f_n^t\|_\infty$ . Fix  $x_2 \in (0, 1)$  and consider the function  $\phi(x_1) = f_n^t(x_1, x_2)$ . We have  $\phi'(x_1) = \nabla f_n^t(x_1, x_2) \cdot e_1$  (here  $(e_1, e_2)$  is the canonical basis of  $\mathbb{R}^2$ ). By the Neumann boundary conditions,  $\phi'(0) = \phi'(1) = 0$ , so that  $\|\phi'\|_\infty \leq \|\phi''\|_\infty \lesssim \|\nabla^2 f_n^t\|_\infty$  and thus  $\|\nabla f_n^t \cdot e_1\|_\infty \lesssim \|\nabla^2 f_n^t\|_\infty$ . Exchanging the



roles of  $x_1$  and  $x_2$  concludes the proof of the claim.

*Step 4 (Conclusion)* In the event  $A_n = \{\|\nabla^2 f_n^t\|_\infty < 1/\log n\}$  we use  $(f_n, g_n)$ , as defined in the previous step, in the duality (2.10). The event  $A_n$  has large probability, i.e.

$$\mathbb{P}(A_n^c) \lesssim n^{-k}, \quad (3.12)$$

for every  $k > 0$  (with an implicit constant depending on  $k$  only, see [2]), hence it is sufficient to bound from below the expectation of  $Wb^2(\mu_n^t, 1)$  on  $A_n$  (using the trivial bound  $Wb^2(\mu_n^t, 1) \leq 2$  on  $A_n^c$ ). If  $A_n$  occurs, we have by (3.11),

$$\|f_n^t\|_\infty + \|\nabla f_n^t\|_\infty + \|\Delta f_n^t\|_\infty \lesssim \frac{1}{\log n},$$

so that (3.7) holds if  $n$  is sufficiently large (for fixed  $\eta$ ) and

$$\|\Delta f_n\|_\infty \lesssim \|\nabla^2(f_n^t \chi_\eta)\|_\infty \lesssim \frac{1}{\eta^2 \log n} = \omega_\eta(n),$$

where for fixed  $\eta$ ,  $\lim_{n \rightarrow \infty} \omega_\eta(n) = 0$ . Thus,

$$\begin{aligned} \frac{1}{2} Wb^2(\mu_n^t, 1) &\geq \int_{(0,1)^2} (f_n + g_n) + \int f_n (\mu_n^t - 1) \\ &\geq -e^{\omega_\eta(n)} \int_{(0,1)^2} \frac{|\nabla f_n|^2}{2} - \int_{(0,1)^2} f_n \Delta f_n^t \\ &= \left(1 - \frac{e^{\omega_\eta(n)}}{2}\right) \int_{(0,1)^2} |\nabla f_n|^2 + \int_{(0,1)^2} \nabla f_n \cdot (\nabla f_n^t - \nabla f_n). \end{aligned}$$

Taking expectations, we bound the second term in the right-hand side using (3.8) and (3.9),

$$\limsup_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E} \left[ \int_{(0,1)^2} |\nabla f_n \cdot (\nabla f_n^t - \nabla f_n)| \right] \lesssim \eta^{1/4}.$$

For the first term, we have

$$\begin{aligned} &\liminf_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E} \left[ (2 - e^{\omega_\eta(n)}) \chi_{A_n} \int_{(0,1)^2} |\nabla f_n|^2 \right] \\ &\geq \liminf_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E} \left[ \int_{(0,1)^2} |\nabla f_n|^2 \right] - \limsup_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E} \left[ \chi_{A_n^c} \int_{(0,1)^2} |\nabla f_n|^2 \right] \\ &\stackrel{(3.9)}{=} \frac{1}{4\pi} - C\eta^{1/4}, \end{aligned}$$

where we used in the last line that

$$\limsup_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E} \left[ \chi_{A_n^c} \int_{(0,1)^2} |\nabla f_n|^2 \right] \leq \limsup_{n \rightarrow \infty} \mathbb{P}(A_n^c)^{1/2} \frac{n}{\log n} \mathbb{E} \left[ \int_{(0,1)^2} |\nabla f_n|^4 \right]^{1/2} \stackrel{(3.12) \& (3.10)}{=} 0.$$

Letting  $\eta \rightarrow 0$  we finally obtain the thesis.  $\square$

Next, we deal with the analogue result for the bipartite matching. In fact, we even consider the case of different number of points. Let us point out that we only prove a lower bound since this is what we will later use in the proof of Theorem 1.2. The corresponding upper bound may be obtained as a corollary of that theorem or more elementarily by arguing as in (5.2).

**Proposition 3.2.** *Let  $(X_i, Y_i)_{i=1}^\infty$  be i.i.d uniformly distributed in  $(0, 1)^2$ , then there exists a rate function<sup>2</sup>  $\omega$  such that for every  $n \leq m$ ,*

$$\mathbb{E} \left[ Wb_{(0,1)^2,2}^2 \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \frac{1}{m} \sum_{i=1}^m \delta_{Y_i} \right) \right] \geq \frac{\log n}{4\pi n} \left( 1 + \frac{n}{m} - \omega(n) \right).$$

*Proof.* The proof is very similar to the lower bound in the previous Proposition 3.1 so we only sketch it. Define  $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$  and  $\lambda_m = \frac{1}{m} \sum_{i=1}^m \delta_{Y_i}$ , and write  $W$  for  $W_{(0,1)^2,2}$  and similarly for  $Wb$ . We denote by  $\mu_n^t$  and  $\lambda_m^t$  the Neumann heat-kernel regularizations of  $\mu_n$  and  $\lambda_m$ . We define  $f_n^t$  as the solution to  $-\Delta f_n^t = \mu_n^t - 1$  and  $f_m^t$  the solution of  $-\Delta f_m^t = \lambda_m^t - 1$ . We finally set  $f_{n,m}^t = f_n^t - f_m^t$ , and choose  $t = (\log n)^4/n$ . Notice that if  $t_m = (\log m)^4/m$  then  $t_m \leq t$ .

Repeating *Step 1* of the proof of Proposition 3.1 we have, by the triangle inequality and  $Wb \leq W$ ,

$$\begin{aligned} Wb^2(\mu_n^t, \lambda_m^t) - Wb^2(\mu_n, \lambda_m) &\leq (W(\mu_n^t, \mu_n) + Wb(\mu_n, \lambda_m) + W(\lambda_m^t, \lambda_m))^2 - Wb^2(\mu_n, \lambda_m) \\ &\lesssim W^2(\mu_n^t, \mu_n) + W^2(\lambda_m^t, \lambda_m) + W(\mu_n, \lambda_m) (W(\mu_n^t, \mu_n) + W(\lambda_m^t, \lambda_m)). \end{aligned}$$

The refined contractivity estimate [2, Theorem 5.2] gives

$$\mathbb{E} [W^2(\mu_n^t, \mu_n)] \lesssim \frac{\log \log n}{n},$$

and moreover, since  $t = (\log n)^4/n = \gamma/m$  with  $\gamma = (\log n)^4(m/n) \gg \log m$ , applying again [2, Theorem 5.2] we have

$$\mathbb{E} [W^2(\lambda_m^t, \lambda_m)] \lesssim \frac{\log \gamma}{m} \lesssim \frac{\log \log n + \log(\frac{m}{n})}{m} \lesssim \frac{\log \log n}{n},$$

where we used that  $m \geq n$  and thus  $\frac{n}{m} \log(\frac{m}{n})$  is bounded. Using also that

$$\mathbb{E} [W^2(\mu_n, \lambda_m)] \lesssim \mathbb{E} [W^2(\mu_n, 1)] + \mathbb{E} [W^2(\lambda_m, 1)] \lesssim \frac{\log n}{n} + \frac{\log m}{m} \lesssim \frac{\log n}{n}$$

we find the analogously to (3.3) that for some constant  $C > 0$ ,

$$\mathbb{E} [Wb^2(\mu_n, \lambda_m)] \geq \mathbb{E} [Wb^2(\mu_n^t, \lambda_m^t)] - C \frac{\sqrt{\log n \log \log n}}{n}.$$

Turning now to *Steps 2-4*, we see that the proof can be repeated verbatim using  $f_{n,m}^t$  instead of  $f_n^t$  and the triangle inequality to obtain the various estimates from the corresponding ones on  $f_n^t$  and  $f_m^t$ . The only additional ingredient is that using integration by parts and independence, we have the following orthogonality property

$$\begin{aligned} \mathbb{E} \left[ \int_{(0,1)^2} \nabla f_n^t \cdot \nabla f_m^t \right] &= -\mathbb{E} \left[ \int_{(0,1)^2} f_n^t \Delta f_m^t \right] = - \int_{(0,1)^2} \mathbb{E} [f_n^t] \mathbb{E} [\lambda_m^t] \\ &= 0. \end{aligned}$$

---

<sup>2</sup>recall that a rate function is a generic decreasing function  $\omega$  such that  $\lim_{t \rightarrow \infty} \omega(t) = 0$ .

Therefore, appealing once again to [2, Lemma 3.14],

$$\begin{aligned}
\mathbb{E} \left[ \int_{(0,1)^2} |\nabla f_{n,m}^t|^2 \right] &= \mathbb{E} \left[ \int_{(0,1)^2} |\nabla f_n^t|^2 \right] + \mathbb{E} \left[ \int_{(0,1)^2} |\nabla f_m^t|^2 \right] + 2\mathbb{E} \left[ \int_{(0,1)^2} \nabla f_n^t \cdot \nabla f_m^t \right] \\
&= \mathbb{E} \left[ \int_{(0,1)^2} |\nabla f_n^t|^2 \right] + \mathbb{E} \left[ \int_{(0,1)^2} |\nabla f_m^t|^2 \right] \\
&= \frac{|\log t|}{4\pi n} + \frac{|\log t|}{4\pi m} + O\left(\frac{1}{n}\right) \\
&= \frac{\log n}{4\pi n} \left(1 + \frac{n}{m} + \omega(n)\right). \quad \square
\end{aligned}$$

**Remark 3.3.** Arguing as in [4, Theorem 5.2], we can also provide a lower bound for the Wasserstein distance of order 1:

$$\liminf_{n \rightarrow \infty} \frac{1}{\sqrt{\log n}} \mathbb{E} \left[ Wb_{(0,1)^2,1} \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \frac{1}{n} \sum_{i=1}^n \delta_{Y_i} \right) \right] > 0.$$

The proof is exactly as in [4, Theorem 5.2]. The only difference lies in the choice of the Lipschitz function in the dual description of the Wasserstein distance. The function  $\phi$  from [4, Theorem 5.2] must be replaced by  $\phi\chi_\eta$  where  $\chi_\eta$  is defined in the proof above.

#### 4. PROOF OF THEOREM 1.1

We recall that for a given Lipschitz and connected domain  $\Omega$  and a Hölder continuous and uniformly strictly positive probability density  $\rho$  on  $\Omega$ , we consider  $(X_i)_{i=1}^\infty$  i.i.d. random variables with common distribution  $\rho$ . Letting

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

we want to prove that

$$\lim_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E} [W_2^2(\mu_n, \rho)] = \frac{|\Omega|}{4\pi}. \quad (4.1)$$

Without loss of generality, we assume that  $|\Omega| = 1$ . We also omit to specify  $p = 2$  and simply write  $W_Q = W_{Q,2}$  and  $Wb_Q = Wb_{Q,2}$  for  $Q \subseteq \Omega$ . Let us introduce some further notation: for  $r > 0$  we consider  $(X_i^r)_{i=1}^\infty$ , i.i.d. uniformly distributed random variables in  $[0, r]^2$ . We then define the average cost functions

$$F_r(n) = \mathbb{E} \left[ W_{[0,r]^2}^2 \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i^r}, \frac{1}{r^2} \right) \right], \quad Fb_r(n) = \mathbb{E} \left[ Wb_{[0,r]^2}^2 \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i^r}, \frac{1}{r^2} \right) \right]$$

By scaling, (3.1) and Proposition 3.1, we have

$$F_r(n) = r^2 F_1(n) \leq r^2 \frac{\log n}{4\pi n} (1 + \omega(n)), \quad \text{and} \quad Fb_r(n) = r^2 Fb_1(n) \geq r^2 \frac{\log n}{4\pi n} (1 - \omega(n)). \quad (4.2)$$

*Step 1 (Whitney decomposition).* We consider  $\{Q_k\}_k$  a Whitney decomposition of  $\Omega$ , see [41, Chapter 6]. For  $\delta = \delta(n) > 0$  to be chosen below we set  $U_\delta = \{Q_k : \text{diam}(Q_k) \geq \delta\}$  and  $V_\delta = \{Q_k : \text{diam}(Q_k) < \delta\}$ . For  $r > 0$  dyadic and small enough so that Lemma 2.5 applies with  $r \text{diam}(\Omega)$ , we split each  $Q_k$  into  $r^{-2}$  sub-cubes. We let  $V_\delta^r$  be the union of these cubes. Up to relabeling we have  $U_\delta \cup V_\delta^r = \{\Omega_k\}_k$ . We finally define  $\kappa_k = \mu_n(\Omega_k)/\rho(\Omega_k)$ .

For  $\varepsilon > 0$ , we estimate by the subadditivity inequality (2.5)

$$\mathbb{E} [W_\Omega^2(\mu_n, \rho)] \leq (1 + \varepsilon) \sum_k \mathbb{E} [W_{\Omega_k}^2(\mu_n, \kappa_k \rho)] + \frac{C}{\varepsilon} \mathbb{E} \left[ W^2 \left( \sum_k \kappa_k \rho \chi_{\Omega_k}, \rho \right) \right] \quad (4.3)$$

and the superadditivity inequality (2.9)

$$\mathbb{E} [Wb_{\Omega}^2(\mu_n, \rho)] \geq (1 - \varepsilon) \sum_k \mathbb{E} [Wb_{\Omega_k}^2(\mu_n, \kappa_k \rho)] - \frac{C}{\varepsilon} \mathbb{E} \left[ Wb_{\Omega}^2 \left( \sum_k \kappa_k \rho \chi_{\Omega_k}, \rho \right) \right]. \quad (4.4)$$

In both expressions, we recognize in the right-hand sides a sum of “local” contributions and an additional “global” term. We consider these contributions separately in the next two steps and show in particular that the global term does not contribute in the limit.

*Step 2 (Local term).* We let  $N_k = n\mu_n(\Omega_k)$  be the number of points  $X_i$  which belong to the cube  $\Omega_k$ . This is a random variable with binomial law and parameters  $(n, \rho(\Omega_k))$ . We decompose the sum in the right-hand side of (4.3) as

$$\sum_k \mathbb{E} [W_{\Omega_k}^2(\mu_n, \kappa_k \rho)] = \sum_{U_{\delta}} \mathbb{E} [W_{\Omega_k}^2(\mu_n, \kappa_k \rho)] + \sum_{V_{\delta}^r} \mathbb{E} [W_{\Omega_k}^2(\mu_n, \kappa_k \rho)].$$

For the first term we use the naive estimate

$$\mathbb{E} [W_{\Omega_k}^2(\mu_n, \kappa_k \rho)] \lesssim \text{diam}^2(\Omega_k) \mathbb{E} [\mu_n(\Omega_k)] \leq \delta^2 \rho(\Omega_k),$$

to get

$$\sum_{U_{\delta}} \mathbb{E} [W_{\Omega_k}^2(\mu_n, \kappa_k \rho)] \lesssim \delta^2 \sum_{U_{\delta}} \rho(\Omega_k) \lesssim \delta^2 |d(\cdot, \partial\Omega)| \lesssim \delta \lesssim \delta^3. \quad (4.5)$$

As for the second term, we use (2.14) from Lemma 2.5 to infer that for every  $\Omega_k \in V_{\delta}^r$ ,

$$\begin{aligned} \mathbb{E} [W_{\Omega_k}^2(\mu_n, \kappa_k \rho)] &= \mathbb{E} [\mathbb{E} [W_{\Omega_k}^2(\mu_n, \kappa_k \rho) | N_k]] \\ &\leq (1 + Cr^{\alpha}) \mathbb{E} \left[ \frac{N_k}{n} F_{\sqrt{|\Omega_k|}}(N_k) \right] \\ &\stackrel{(4.2)}{\leq} (1 + Cr^{\alpha}) \frac{|\Omega_k|}{4\pi n} \mathbb{E} [\log(N_k)(1 + \omega(N_k))]. \end{aligned}$$

Using the concentration properties of the binomial random variable  $N_k$  and the fact that  $\delta = \delta(n)$  satisfies  $\lim_{n \rightarrow \infty} n\delta^2 = \infty$ , we see that

$$\mathbb{E} [\log(N_k)(1 + \omega(N_k))] = \log(n\rho(\Omega_k))(1 + \omega(nr^2\delta^2)) \quad (4.6)$$

and we find

$$\mathbb{E} [W_{\Omega_k}^2(\mu_n, \kappa_k \rho)] \leq (1 + Cr^{\alpha}) \frac{\log(n\rho(\Omega_k))}{4\pi n} |\Omega_k| (1 + \omega(nr^2\delta^2)). \quad (4.7)$$

We now claim that

$$\sum_k |\Omega_k| |\log |\Omega_k|| \lesssim |\log r|. \quad (4.8)$$

Indeed,

$$\begin{aligned} \sum_k |\Omega_k| |\log |\Omega_k|| &= \sum_{U_{\delta}} |\Omega_k| |\log |\Omega_k|| + \sum_{V_{\delta}^r} |\Omega_k| |\log |\Omega_k|| \\ &= \sum_{U_{\delta}} |Q_k| |\log |Q_k|| + \sum_{V_{\delta}} |Q_k| |\log |Q_k|| + \log r \\ &\lesssim \sum_k |Q_k| |\log |Q_k|| + \sum_{V_{\delta}} |Q_k| |\log r| \lesssim 1 + |\log r| \lesssim |\log r|, \end{aligned}$$

where we used that (notice that the finiteness of the integral below is ensured, for instance, by the finiteness of the Minkowski content of  $\Omega$ , in turn ensured by Lipschitz regularity)

$$\sum_k |Q_k| |\log |Q_k|| \lesssim \int_{\Omega} |\log d(\cdot, \partial\Omega)| < \infty.$$

Thus, from (4.7) and (4.8),

$$\sum_k \mathbb{E} [W_{\Omega_k}^2(\mu_n, \kappa_k \rho)] \leq (1 + Cr^\alpha) \frac{\log n}{n} \frac{1}{4\pi} + C \left( \frac{|\log r|}{n} + \frac{\log n}{n} \omega(nr^2 \delta^2) + \delta^3 \right). \quad (4.9)$$

For the analogue term in (4.4), we simply discard the terms with  $\Omega_k \in U_\delta$  and for the remaining ones use again (2.14) from Lemma 2.5 and the function  $Fb_r$  instead of  $F_r$  to obtain

$$\sum_k \mathbb{E} [Wb_{\Omega_k}^2(\mu_n, \kappa_k \rho)] \geq (1 - Cr^\alpha) \frac{\log n}{n} \frac{1 - C\delta}{4\pi} - C \left( \frac{|\log r|}{n} + \frac{\log n}{n} \omega(nr^2 \delta^2) \right).$$

*Step 3 (Global term).* We now turn to the second term in the right-hand side of (4.3), for which we show that

$$\mathbb{E} \left[ W_\Omega^2 \left( \sum_k \kappa_k \rho \chi_{\Omega_k}, \rho \right) \right] \lesssim \frac{|\log r|}{n}. \quad (4.10)$$

Notice that the inequality  $Wb \leq W$  gives an inequality also for the corresponding term in (4.4).

We first use Lemma 2.2 to deduce

$$\begin{aligned} W_\Omega^2 \left( \sum_k \kappa_k \rho \chi_{\Omega_k}, \rho \right) &\lesssim \left\| \sum_k (\kappa_k - 1) \rho \chi_{\Omega_k} \right\|_{H^{-1}}^2 \\ &= \left\| \sum_k (\kappa_k - 1) (\rho \chi_{\Omega_k} - \rho(\Omega_k) \chi_\Omega) \right\|_{H^{-1}}^2 = \left\| \sum_k (\kappa_k - 1) f_k \right\|_{H^{-1}}^2, \end{aligned}$$

where

$$f_k = \rho \chi_{\Omega_k} - \rho(\Omega_k) \chi_\Omega.$$

Taking expectation and expanding the squares we have

$$\left\| \sum_k (\kappa_k - 1) f_k \right\|_{H^{-1}}^2 = \sum_k \mathbb{E}[(\kappa_k - 1)^2] \|f_k\|_{H^{-1}}^2 + \sum_{j \neq k} \mathbb{E}[(\kappa_j - 1)(\kappa_k - 1)] \langle f_j, f_k \rangle_{H^{-1}}.$$

For the first sum, we use that

$$\mathbb{E}[(\kappa_k - 1)^2] = \frac{1}{\rho^2(\Omega_k)} \mathbb{E}[(\mu_n(\Omega_k) - \rho(\Omega_k))^2] \leq \frac{1}{n\rho(\Omega_k)} \lesssim \frac{1}{n|\Omega_k|}$$

and, by Lemma 2.1 with  $f = f_k$ , since  $\|f_k\|_\infty \lesssim 1$  and  $\|f_k\|_1 \sim \rho(\Omega_k)$ ,

$$\|f_k\|_{H^{-1}}^2 \lesssim \rho(\Omega_k)^2 |\log \rho(\Omega_k)| \lesssim |\Omega_k|^2 |\log |\Omega_k|| \quad (4.11)$$

to get

$$\sum_k \mathbb{E}[(\kappa_k - 1)^2] \|f_k\|_{H^{-1}}^2 \lesssim \frac{1}{n} \sum_k |\Omega_k| |\log |\Omega_k|| \lesssim \frac{1}{n} \sum_k |\Omega_k| |\log |\Omega_k|| \stackrel{(4.8)}{\lesssim} \frac{|\log r|}{n}.$$

For the second sum, we use that if  $j \neq k$ ,

$$\mathbb{E}[(\kappa_j - 1)(\kappa_k - 1)] = \frac{\mathbb{E}[\mu_n(\Omega_j) \mu_n(\Omega_k)]}{\rho(\Omega_j) \rho(\Omega_k)} - 1 = -\frac{1}{n}$$

together with (4.11) and Cauchy-Schwarz inequality to conclude

$$\begin{aligned}
\sum_{j \neq k} \mathbb{E}[(\kappa_j - 1)(\kappa_k - 1)] \langle f_j, f_k \rangle_{H^{-1}} &\lesssim \frac{1}{n} \sum_{j \neq k} \|f_j\|_{H^{-1}} \|f_k\|_{H^{-1}} \\
&\lesssim \frac{1}{n} \sum_{j \neq k} |\Omega_j| |\log |\Omega_j||^{1/2} |\Omega_k| |\log |\Omega_k||^{1/2} \\
&\lesssim \frac{1}{n} \left( \sum_k |\Omega_k| |\log |\Omega_k||^{1/2} \right)^2 \\
&\lesssim \frac{|\log r|}{n},
\end{aligned}$$

where we argued as above to bound  $\sum_k |\Omega_k| |\log |\Omega_k||^{1/2} \lesssim |\log r|^{1/2}$ . This proves (4.10).

*Step 4 (Conclusion).* Putting together (4.9) and (4.10), we conclude that

$$\mathbb{E}[W^2(\mu_n, \rho)] \leq (1 + \varepsilon)(1 + Cr^\alpha) \frac{\log n}{n} \frac{1}{4\pi} + C \left( \frac{\log n}{n} \omega(nr^2 \delta^2) + \delta^3 + \frac{1}{\varepsilon} \frac{|\log r|}{n} \right)$$

and similarly

$$\mathbb{E}[Wb^2(\mu_n, \rho)] \geq (1 - \varepsilon)(1 - Cr^\alpha) \frac{\log n}{n} \frac{1 - C\delta}{4\pi} - \left( \frac{\log n}{n} \omega(nr^2 \delta^2) + \frac{1}{\varepsilon} \frac{|\log r|}{n} \right).$$

We now choose  $\delta = \delta(n) = n^{-\beta}$  with  $\beta \in (\frac{1}{3}, \frac{1}{2})$ , so that the condition  $\lim_{n \rightarrow \infty} n\delta^2 = \infty$  holds but also  $\lim_{n \rightarrow \infty} n\delta^3 / \log n = 0$ . This yields

$$\begin{aligned}
(1 + \varepsilon)(1 + Cr^\alpha) \frac{1}{4\pi} &\geq \limsup_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E}[W^2(\mu_n, \rho)] \geq \liminf_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E}[Wb^2(\mu_n, \rho)] \\
&\geq (1 - \varepsilon)(1 - Cr^\alpha) \frac{1}{4\pi}.
\end{aligned}$$

Letting finally  $r \rightarrow 0$  and  $\varepsilon \rightarrow 0$  we conclude the proof of (4.1).

## 5. PROOF OF THEOREM 1.2

For  $(X_i, Y_i)_{i=1}^\infty$  i.i.d. random variables with common distribution  $\rho$  in  $\Omega$ , we write

$$\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i} \quad \text{and} \quad \lambda_m = \frac{1}{m} \sum_{j=1}^m \delta_{Y_j}.$$

We now prove that

$$\lim_{\substack{n, m \rightarrow \infty \\ m/n \rightarrow q}} \frac{n}{\log n} \mathbb{E}[W_2^2(\mu_n, \lambda_m)] = \frac{|\Omega|}{4\pi} \left( 1 + \frac{1}{q} \right).$$

As above, we may assume by scaling that  $|\Omega| = 1$  and we will omit to specify  $p = 2$ , simply writing  $W_Q = W_{Q,2}$  and  $Wb_Q = Wb_{Q,2}$  for  $Q \subseteq \Omega$ . For  $n \leq m$ , we will mostly focus on the lower bound,

$$\mathbb{E}[Wb_\Omega^2(\mu_n, \lambda_m)] \geq \frac{\log n}{4\pi n} \left( 1 + \frac{n}{m} - \omega(n) \right). \quad (5.1)$$

Indeed, we first show how the upper bound,

$$\mathbb{E}[W_\Omega^2(\mu_n, \lambda_m)] \leq \frac{\log n}{4\pi n} \left( 1 + \frac{n}{m} + \omega(n) \right), \quad (5.2)$$



can be then quickly obtained as a consequence of [4, Proposition 4.9]<sup>3</sup>. Letting  $T^{\mu_n}$  (respectively  $T^{\lambda_m}$ ) be the optimal transport maps between  $\rho$  and  $\mu_n$  (respectively  $\lambda_m$ ), by independence we have

$$\begin{aligned} \mathbb{E} [W_\Omega^2(\mu_n, \lambda_m)] &\leq \mathbb{E} \left[ \int_\Omega |T^{\mu_n} - T^{\lambda_m}|^2 \rho \right] \\ &= \mathbb{E} [W_\Omega^2(\mu_n, \rho)] + \mathbb{E} [W_\Omega^2(\lambda_m, \rho)] \\ &\quad - 2 \int_\Omega \mathbb{E} [(T^{\mu_n} - x)] \cdot \mathbb{E} [(T^{\lambda_m} - x)] \rho. \end{aligned} \quad (5.3)$$

We start with the case  $n = m$ . In this case  $\mu_n$  and  $\lambda_m$  have the same law, hence the last term above becomes

$$\int_\Omega \mathbb{E} [(T^{\mu_n} - x)] \cdot \mathbb{E} [(T^{\lambda_m} - x)] \rho = \int_{(0,1)^2} |\mathbb{E} [(T^{\mu_n} - x)]|^2 \rho.$$

By (5.1) and Theorem 1.1 (recall (4.1)) we get

$$\limsup_{n \rightarrow \infty} \frac{n}{\log n} \int_\Omega |\mathbb{E} [(T^{\mu_n} - x)]|^2 \rho \leq \lim_{n \rightarrow \infty} \frac{n}{\log n} \left( \mathbb{E} [W_\Omega^2(\mu_n, \rho)] - \frac{1}{2} \mathbb{E} [W_\Omega^2(\mu_n, \lambda_n)] \right) = 0.$$

Therefore,

$$\mathbb{E} [W_\Omega^2(\mu_n, \rho)] + \int_\Omega |\mathbb{E} [(T^{\mu_n} - x)]|^2 \rho \leq \frac{\log n}{4\pi n} (1 + \omega(n)).$$

We now turn to the general case  $n \leq m$ . Using the inequality  $-2ab \leq a^2 + b^2$  for the last term in (5.3), we obtain (here we use that  $\omega(m) \leq \omega(n)$  and  $\log m = \log \frac{m}{n} + \log n$ )

$$\begin{aligned} \mathbb{E} [W_\Omega^2(\mu_n, \lambda_m)] &\leq \frac{\log n}{4\pi n} (1 + \omega(n)) + \frac{\log m}{4\pi m} (1 + \omega(m)) \\ &= \frac{\log n}{4\pi n} \left( 1 + \frac{n}{m} + \left( \frac{\log m}{\log n} - 1 \right) \frac{n}{m} + \frac{n}{m} \frac{\log m}{\log n} \omega(m) + \omega(n) \right) \\ &\leq \frac{\log n}{4\pi n} \left( 1 + \frac{n}{m} + \frac{1}{\log n} \frac{\log \frac{m}{n}}{\frac{m}{n}} (1 + \omega(n)) + \left( \frac{n}{m} + 1 \right) \omega(n) \right) \\ &\leq \frac{\log n}{4\pi n} \left( 1 + \frac{n}{m} + \omega(n) \right). \end{aligned}$$

This proves (5.2).

We thus focus on the proof of (5.1). For this we follow the same steps as in Theorem 1.1. However, we need to be more careful when estimating the “global” term. Indeed, we cannot apply directly Lemma 2.2 since the measure  $\lambda_m$  is singular. To fix this issue, we first slightly modify the Whitney decomposition to avoid cubes whose measure is too small. This guarantees that with overwhelming probability, every element of the partition contains many points so that we may argue as in [22, Proposition 5.2]. Notice that of course we could have used Lemma 5.1 also in the proof of Theorem 1.1.

To simplify the exposition, we postpone the proof of the following geometric result to Appendix A.

**Lemma 5.1.** *Let  $\Omega \subset \mathbb{R}^d$  be a bounded open set with Lipschitz boundary, let  $\{Q_k\}_k$  be a Whitney decomposition of  $\Omega$ . For every  $\delta > 0$  small enough (depending on  $\Omega$ ), letting  $V_\delta = \{Q_k : \text{diam}(Q_k) \geq \delta\}$ , there exists a family  $U_\delta = \{\Omega_k\}_k$  of disjoint open sets such that  $\text{diam}(\Omega_k) \lesssim \delta$ ,  $|\Omega_k| \sim \delta^d$  and  $U_\delta \cup V_\delta$  is a partition of  $\Omega$ .*

*Step 1 (Whitney-type decomposition and reduction to a good event).* For  $\delta = \delta(n) > 0$  to be specified below, let  $U_\delta$  and  $V_\delta$  be given by Lemma 5.1. For  $r > 0$  dyadic we divide each

<sup>3</sup>in fact the statement there wrongly contains an equality that should be instead an inequality  $\leq$

cube  $Q_k \in V_\delta$  in  $r^{-2}$  equal sub-cubes of sidelength  $r\ell(Q_k)$ . We let  $V_\delta^r$  be their collection and relabel  $U_\delta \cup V_\delta^r = \{\Omega_k\}_k$ . We set

$$\theta = \frac{1}{\sqrt{\log n}} \quad (5.4)$$

and for every  $k$  such that  $\lambda_m(\Omega_k) \neq 0$

$$\kappa_k = \frac{\mu_n(\Omega_k)}{\lambda_m(\Omega_k)}.$$

If instead  $\lambda_m(\Omega_k) = 0$  we arbitrarily set  $\kappa_k = 0$ . Define the event

$$A = \left\{ \forall \Omega_k \in V_\delta^r \cup U_\delta, \left| 1 - \frac{\mu_n(\Omega_k)}{\rho(\Omega_k)} \right| + \left| 1 - \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} \right| \leq \frac{\theta}{4} \right\}.$$

Notice that if  $\left| 1 - \frac{\mu_n(\Omega_k)}{\rho(\Omega_k)} \right| + \left| 1 - \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} \right| \leq \frac{\theta}{4}$  then

$$|1 - \kappa_k| \leq \frac{\theta}{2}. \quad (5.5)$$

We claim that if

$$\delta = \frac{1}{n^\beta}, \quad \text{for some } \beta \in (0, 1/2), \quad r \sim \frac{1}{\log(n)},$$

then

$$\mathbb{P}(A^c) \lesssim \exp \left( -c \frac{n^{1-2\beta}}{(\log n)^3} \right), \quad (5.6)$$

so that we can restrict ourselves to the event  $A$  in the following steps. By a union bound, to prove the claim, we bound

$$\mathbb{P}(A^c) \leq \sum_k \mathbb{P} \left( \left| 1 - \frac{\mu_n(\Omega_k)}{\rho(\Omega_k)} \right| \geq \theta \right) + \sum_k \mathbb{P} \left( \left| 1 - \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} \right| \geq \theta \right).$$

We focus only on the terms involving  $\mu_n$  (those with  $\lambda_m$  are analogous, recalling that  $n \leq m$ ). For every  $k$ , the number of points  $n\mu_n(\Omega_k)$  is a binomial random variable with parameters  $(n, \rho(\Omega_k))$ . Hence, by Chernoff bounds

$$\mathbb{P} \left( \left| 1 - \frac{\mu_n(\Omega_k)}{\rho(\Omega_k)} \right| \geq \theta \right) \leq \exp \left( -\frac{n\rho(\Omega_k)}{2 \log n} \right).$$

Summing this over  $k$  and using that by definition of a Whitney partition, for every  $j \in \mathbb{N}$  such that  $2^{-j} \geq \delta$ ,

$$\#\{\Omega_k \in V_\delta^r : \ell(\Omega_k) \in r(2^{-(j+1)}, 2^{-j}]\} \lesssim r^{-2} 2^j,$$

we find

$$\begin{aligned} \sum_k \mathbb{P} \left( \left| 1 - \frac{\mu_n(\Omega_k)}{\rho(\Omega_k)} \right| \geq \theta \right) &\leq \sum_{V_\delta^r} \exp \left( -\frac{n\rho(\Omega_k)}{2 \log n} \right) + \sum_{U_\delta} \exp \left( -\frac{n\rho(\Omega_k)}{2 \log n} \right) \\ &\lesssim \sum_{2^{-j} \geq \delta} r^{-2} 2^j \exp \left( -c \frac{nr^2 2^{-2j}}{\log n} \right) + \delta^{-1} \exp \left( -c \frac{n\delta^2}{\log n} \right) \\ &\lesssim r^{-2} \delta^{-1} \exp \left( -c \frac{nr^2 \delta^2}{\log n} \right) + \delta^{-1} \exp \left( -c \frac{n\delta^2}{\log n} \right) \\ &\lesssim r^{-2} \delta^{-1} \exp \left( -c \frac{nr^2 \delta^2}{\log n} \right). \end{aligned}$$

Up to replacing the constant  $c > 0$  with a smaller one, to control the terms  $r^{-2}\delta^{-1} \lesssim n^\beta(\log n)^2$ , we obtain (5.6). We end this step applying (2.9) and using  $Wb \leq W$  to get for every  $\varepsilon > 0$ ,

$$\begin{aligned} Wb_\Omega^2(\mu_n, \lambda_m) &\geq (1 - \varepsilon) \sum_k Wb_{\Omega_k}^2(\mu_n, \kappa_k \lambda_m) - \frac{C}{\varepsilon} W_\Omega^2 \left( \sum_k \kappa_k \chi_{\Omega_k} \lambda_m, \lambda_m \right) \\ &\geq (1 - \varepsilon) \sum_{V_\delta^r} Wb_{\Omega_k}^2(\mu_n, \kappa_k \lambda_m) - \frac{C}{\varepsilon} W_\Omega^2 \left( \sum_k \kappa_k \chi_{\Omega_k} \lambda_m, \lambda_m \right). \end{aligned}$$

*Step 2 (Local term).* For  $r > 0$  let  $(X_i^r, Y_i^r)$  be i.i.d uniformly distributed random variables in  $[0, r]^2$  and for  $n \leq m$  let

$$Fb_r(n, m) = \mathbb{E} \left[ Wb_{[0, r]^2}^2 \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i^r}, \frac{1}{m} \sum_{i=1}^m \delta_{Y_i^r} \right) \right]$$

so that by scaling and Proposition 3.2,

$$Fb_r(n, m) = r^2 Fb_1(n, m) \geq r^2 \frac{\log n}{4\pi n} \left( 1 + \frac{n}{m} - \omega(n) \right).$$

Define  $N_k = n\mu_n(\Omega_k)$  (respectively  $M_k = m\lambda_m(\Omega_k)$ ) to be the number of points  $X_i$  (respectively  $Y_i$ ) in  $\Omega_k$  so that in particular  $\kappa_k = N_k/M_k$ . For every given  $k$ , the random variables  $N_k$  and  $M_k$  are independent binomial random variables with parameters respectively  $(n, \rho(\Omega_k))$  and  $(m, \rho(\Omega_k))$ . We then define the random variables  $N_k^* = \min(N_k, M_k)$  and  $M_k^* = \max(N_k, M_k)$ .

For every fixed  $\Omega_k \in V_\delta^r$ , recalling that  $\Omega_k$  is a cube of sidelength at most of order  $r$ , and using (2.15), we have

$$\begin{aligned} \mathbb{E} [Wb_{\Omega_k}^2(\mu_n, \kappa_k \lambda_m)] &\geq (1 - Cr^\alpha) \mathbb{E} \left[ \frac{N_k^*}{n} Fb_{\sqrt{|\Omega_k|}}(N_k^*, M_k^*) \right] \\ &\geq (1 - Cr^\alpha) \frac{1}{4\pi n} |\Omega_k| \mathbb{E} \left[ \log N_k^* \left( 1 + \frac{N_k^*}{M_k^*} - \omega(N_k^*) \right) \right] \\ &\geq (1 - Cr^\alpha) \frac{\log(n\rho(\Omega_k))}{4\pi n} |\Omega_k| \left( 1 + \frac{n}{m} - \omega(nr^2\delta^2) \right), \end{aligned}$$

where in the last line one can argue as for (4.6). Summing over  $k$  and using (4.8) we find

$$\mathbb{E} \left[ \sum_{V_\delta^r} Wb_{\Omega_k}^2(\mu_n, \kappa_k \lambda_m) \right] \geq (1 - Cr^\alpha) \frac{\log n}{n} \frac{1 - C\delta}{4\pi} \left( 1 + \frac{n}{m} - \omega(nr^2\delta^2) \right) - C \frac{|\log r|}{n}. \quad (5.7)$$

*Step 3 (Global term).* We claim that for  $\delta = n^{-\beta}$  with  $\beta \in (\frac{1}{3}, \frac{1}{2})$ ,

$$\frac{n}{\log n} \mathbb{E} \left[ W_\Omega^2 \left( \sum_k \kappa_k \chi_{\Omega_k} \lambda_m, \lambda_m \right) \chi_A \right] \lesssim \frac{\log \log n}{\sqrt{\log n}}. \quad (5.8)$$

For this we argue along the lines of [22, Proposition 5.2]. We define

$$\theta_k = 1 - \kappa_k + \theta$$

and recall that if (5.5) holds (which is the case if  $A$  occurs) then  $\frac{3}{2}\theta \geq \theta_k \geq \frac{1}{2}\theta$ . We now use triangle inequality to write

$$\begin{aligned} & W_\Omega^2 \left( \sum_k \kappa_k \chi_{\Omega_k} \lambda_m, \lambda_m \right) \\ & \lesssim W_\Omega^2 \left( \sum_k \kappa_k \chi_{\Omega_k} \lambda_m, \sum_k \left[ (1 - \theta_k) \lambda_m + \theta \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} \rho \right] \chi_{\Omega_k} \right) \\ & + W_\Omega^2 \left( \sum_k \left[ (1 - \theta_k) \lambda_m + \theta \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} \rho \right] \chi_{\Omega_k}, \sum_k \left[ (1 - \theta_k) \lambda_m + \theta_k \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} \rho \right] \chi_{\Omega_k} \right) \\ & + W_\Omega^2 \left( \sum_k \left[ (1 - \theta_k) \lambda_m + \theta_k \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} \rho \right] \chi_{\Omega_k}, \lambda_m \right). \end{aligned}$$

The first and last terms are estimated in a similar way so we only estimate the first one. We use the fact that  $1 - \theta_k = \kappa_k - \theta$  and subadditivity (2.5) to bound from above

$$\begin{aligned} & W_\Omega^2 \left( \sum_k \kappa_k \chi_{\Omega_k} \lambda_m, \sum_k \left[ (\kappa_k - \theta) \lambda_m + \theta \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} \rho \right] \chi_{\Omega_k} \right) \\ & \leq \sum_k W_{\Omega_k}^2 \left( ((\kappa_k - \theta) + \theta) \lambda_m, (\kappa_k - \theta) \lambda_m + \theta \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} \rho \right) \\ & \stackrel{(2.5)}{\leq} \theta \sum_k W_{\Omega_k}^2 \left( \lambda_m, \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} \rho \right). \end{aligned}$$

By (4.5) and (4.7), we get

$$\mathbb{E} \left[ W_\Omega^2 \left( \sum_k \kappa_k \chi_{\Omega_k} \lambda_m, \sum_k \left[ (\kappa_k - \theta) \lambda_m + \theta \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} \rho \right] \chi_{\Omega_k} \right) \right] \lesssim \theta \left( \frac{\log m}{m} + \delta^3 \right).$$

For the middle term, we use again subadditivity together with Lemma 2.2 and the fact that  $\lambda_m(\Omega_k)/\rho(\Omega_k) \sim 1$  in the event  $A$ , to infer

$$\begin{aligned} & W_\Omega^2 \left( \sum_k \left[ (1 - \theta_k) \lambda_m + \theta \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} \rho \right] \chi_{\Omega_k}, \sum_k \left[ (1 - \theta_k) \lambda_m + \theta_k \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} \rho \right] \chi_{\Omega_k} \right) \\ & \leq W_\Omega^2 \left( \sum_k \theta \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} \chi_{\Omega_k} \rho, \sum_k \theta_k \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} \chi_{\Omega_k} \rho \right) \\ & \lesssim \frac{1}{\theta} \left\| \sum_k (\theta - \theta_k) \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} \chi_{\Omega_k} \rho \right\|_{H^{-1}}^2 \\ & = \frac{1}{\theta} \left\| \sum_k (\kappa_k - 1) \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} \chi_{\Omega_k} \rho \right\|_{H^{-1}}^2 \\ & \lesssim \frac{1}{\theta} \left\| \sum_k \left( \frac{\mu_n(\Omega_k)}{\rho(\Omega_k)} - 1 \right) \chi_{\Omega_k} \rho \right\|_{H^{-1}}^2 + \frac{1}{\theta} \left\| \sum_k \left( \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} - 1 \right) \chi_{\Omega_k} \rho \right\|_{H^{-1}}^2, \end{aligned}$$

where in the last line we used that  $(\kappa_k - 1) \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} = \left( \frac{\mu_n(\Omega_k)}{\rho(\Omega_k)} - 1 \right) - \left( \frac{\lambda_m(\Omega_k)}{\rho(\Omega_k)} - 1 \right)$  and triangle inequality in  $H^{-1}$ . By (4.10) we obtain

$$\mathbb{E} \left[ \left\| \sum_i (1 - \kappa_i) \frac{\lambda_m(\Omega_i)}{\rho(\Omega_i)} \chi_{\Omega_i} \rho \right\|_{H^{-1}}^2 \right] \lesssim \frac{|\log r|}{n}.$$

This proves

$$\mathbb{E} \left[ W_{\Omega}^2 \left( \sum_i \kappa_k \chi_{\Omega_k} \lambda_m, \lambda_m \right) \chi_A \right] \lesssim \theta \left( \frac{\log n}{n} + \delta^3 \right) + \frac{1}{\theta} \frac{|\log r|}{n},$$

which recalling the choice (5.4) of  $\theta$ ,  $\delta$  and  $r$  concludes the proof of (5.8).

*Step 4 (Conclusion).* Putting together (5.7) and (5.8) and using (5.6), we find that

$$\begin{aligned} \frac{n}{\log n} \mathbb{E} [Wb_{\Omega}^2(\mu_n, \lambda_m)] &\geq (1 - \varepsilon)(1 - Cr^{\alpha}) \frac{1 - C\delta}{4\pi} \left( 1 + \frac{n}{m} - \omega(nr^2\delta^2) \right) \\ &\quad - \frac{C}{\varepsilon} \left( \frac{\log \log n}{\sqrt{\log n}} + \exp \left( -c \frac{n^{1-2\beta}}{(\log n)^3} \right) \right) \\ &\geq \frac{1}{4\pi} \left( 1 + \frac{m}{n} + \omega(n) \right), \end{aligned}$$

where in the last line we chose  $\varepsilon = \varepsilon(n) \rightarrow 0$  as  $n \rightarrow \infty$  with  $\varepsilon \gg \log \log n / \sqrt{\log n}$ .

## 6. EXTENSION TO GENERAL SETTINGS

In this section we show that the techniques developed above lead to extensions of Theorem 1.1 and Theorem 1.2 to more general settings, including Riemannian manifolds. Let us give the following definition (see also [46, Definition 3.1]).

**Definition 6.1.** We say that a metric measure space  $(\Omega, d, \mathbf{m})$  with  $\mathbf{m}(\Omega) = 1$  is well-decomposable if for every  $\varepsilon > 0$  there exist a *finite* family, that we call  $\varepsilon$ -decomposition,  $(U_k, \Omega_k, T_k)_k$ , such that, for every  $k$ ,

- (1)  $U_k \subseteq \Omega$  is open and  $\mathbf{m}(U_k \cap U_{k'}) = 0$  for  $k \neq k'$ ,  $\mathbf{m}(\bigcup_k U_k) = 1$ ,
- (2)  $\Omega_k \subseteq \mathbb{R}^2$  is open, bounded and connected, with Lipschitz boundary,
- (3)  $T_k : \overline{U}_k \rightarrow \overline{\Omega}_k$ , is invertible with  $T_k(U_k) = \Omega_k$  and  $T_k(\partial U_k) = \partial \Omega_k$  and

$$\text{Lip } T_k, \text{Lip } T_k^{-1} \leq (1 + \varepsilon),$$

- (4)  $(T_k)_{\#}(\mathbf{m} \llcorner U_k)$  has Hölder continuous density, uniformly positive and bounded.

If  $(\Omega, d, \mathbf{m})$  is well-decomposable we define

$$|\Omega| = \inf_{\varepsilon > 0} \left\{ \sum_k |\Omega_k| : (U_k, \Omega_k, T_k)_k \text{ is an } \varepsilon\text{-decomposition of } \Omega \right\}.$$

**Remark 6.2.** Notice that the quantity inside the brackets is a decreasing function of  $\varepsilon$  and that if  $\Omega$  is a Riemannian manifold it coincides with its (Riemannian) volume.

In addition to such decomposability assumption, we need to assume the validity of the inequality:

$$W_2^2(f\mathbf{m}, \mathbf{m}) \lesssim \|f - 1\|_{L^2(\mathbf{m})}^2, \quad \text{for every probability density } f. \quad (6.1)$$

This estimate is analog to the weaker form of the combination of (2.7) with Lemma 2.1 used in [9, 22] and is intimately connected to the validity of an  $L^2$ -Poincaré-Wirtinger inequality (see e.g. [22, Lemma 3.4]).

**Theorem 6.3.** *Let  $(M, d)$  be a length space and  $\Omega \subseteq M$  be open with  $\text{diam}(\Omega) < \infty$ . Let  $\mathbf{m}$  be a probability measure on  $M$  with  $\mathbf{m}(\Omega) = 1$ , such that (6.1) holds and  $(\Omega, d, \mathbf{m})$  is well-decomposable and consider  $(X_i, Y_i)_{i=1}^{\infty}$  be i.i.d. with common density  $\mathbf{m}$ . Then, for every  $q \in [1, \infty]$ ,*

$$\lim_{\substack{n, m \rightarrow \infty \\ m/n \rightarrow q}} \frac{n}{\log n} \mathbb{E} \left[ W_2^2 \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \frac{1}{m} \sum_{j=1}^m \delta_{Y_j} \right) \right] = \frac{|\Omega|}{4\pi} \left( 1 + \frac{1}{q} \right), \quad (6.2)$$

and

$$\lim_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E} \left[ W_2^2 \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \mathbf{m} \right) \right] = \frac{|\Omega|}{4\pi}. \quad (6.3)$$

Examples of metric measure spaces for which the result apply include smooth domains in two-dimensional Riemannian manifolds with boundary, as shown in Appendix B. But also other, less regular cases, may be included, e.g. polyhedral surfaces.

*Proof of Theorem 6.3.* As usual, we simply write  $W = W_2$  and  $Wb = Wb_2$ . We write  $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ ,  $\lambda_m = \frac{1}{m} \sum_{j=1}^m \delta_{Y_j}$ . For  $\varepsilon > 0$ , let  $(U_k, \Omega_k, T_k)_k$  be an  $\varepsilon$ -decomposition and let  $\kappa_k = \mu_n(U_k)/\lambda_m(U_k)$ . In both cases, after an application of the sub-additivity and super-additivity inequalities (2.6), (2.9) we are reduced to estimate separately a finite sum of “local” terms, one for each  $U_k$ , and the global terms (using that  $Wb \leq W$ ) that are respectively

$$\mathbb{E} \left[ W_2^2 \left( \sum_k \kappa_k \chi_{U_k} \lambda_m, \lambda_m \right) \right] \quad \text{and} \quad \mathbb{E} \left[ W_2^2 \left( \sum_k \frac{\mu_n(U_k)}{\mathbf{m}(U_k)} \chi_{U_k} \mathbf{m}, \mathbf{m} \right) \right]. \quad (6.4)$$

*Local terms.* Using (2.12) with  $T_k^{-1}$  and (2.12) with  $T_k$  and taking expectation give respectively

$$\mathbb{E} [W^2 (\mu_n \llcorner U_k, \kappa_k \lambda_m \llcorner U_k)] \leq (1 + \varepsilon)^2 \mathbb{E} \left[ NW^2 \left( \frac{1}{N} \sum_{i=1}^N \delta_{X_{k,i}}, \frac{1}{M} \sum_{j=1}^M \delta_{Y_{k,j}} \right) \right],$$

and

$$\mathbb{E} [Wb_{U_i}^2 (\mu_n \llcorner U_k, \kappa_k \lambda_m \llcorner U_k)] \geq (1 + \varepsilon)^{-2} \mathbb{E} \left[ NWb_{\Omega_i}^2 \left( \frac{1}{N} \sum_{i=1}^N \delta_{X_{k,i}}, \frac{1}{M} \sum_{j=1}^M \delta_{Y_{k,j}} \right) \right],$$

where, for every  $k$ , we consider i.i.d. random variables  $(X_{k,i}, Y_{k,i})_{i=1}^\infty$  with common law  $\rho_k$  given by  $(T_k)_\# \mathbf{m}$  normalized to a probability measure on  $\Omega_k$  and  $N, M$  are further independent random variables with binomial laws of parameters respectively  $(n, \mathbf{m}(U_k))$ ,  $(m, \mathbf{m}(U_k))$ . For each  $k$ , by conditioning upon  $N, M$ , and by (5.1), (5.2) with  $(X_{k,i}, Y_{k,i})_{i=1}^\infty$ , we obtain that, letting  $N^* = \min \{N, M\}$ ,  $M^* = \max \{N, M\}$

$$\mathbb{E} \left[ NW^2 \left( \frac{1}{N} \sum_{i=1}^N \delta_{X_{k,i}}, \frac{1}{M} \sum_{j=1}^M \delta_{Y_{k,j}} \right) \right] \leq \mathbb{E} \left[ N \frac{|\Omega_k| \log N^*}{4\pi N^*} \left( 1 + \frac{N^*}{M^*} + \omega_k(N^*) \right) \right],$$

and similarly

$$\mathbb{E} \left[ NWb_{\Omega_k}^2 \left( \frac{1}{N} \sum_{i=1}^N \delta_{X_{k,i}}, \frac{1}{M} \sum_{j=1}^M \delta_{Y_{k,j}} \right) \right] \geq \mathbb{E} \left[ N \frac{|\Omega_k| \log N^*}{4\pi N^*} \left( 1 + \frac{N^*}{M^*} - \omega_k(N^*) \right) \right].$$

As  $n \rightarrow \infty$  with  $m/n \rightarrow q \in [1, \infty]$ , we have

$$N/n \rightarrow \mathbf{m}(U_k), \quad M/m \rightarrow \mathbf{m}(U_k) \quad \text{and} \quad N^*/N \rightarrow 1, \quad M^*/N \rightarrow q,$$

with the usual concentration inequalities which eventually give

$$\lim_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E} \left[ N \frac{|\Omega_k| \log N^*}{4\pi N^*} \left( 1 + \frac{N^*}{M^*} - \omega_k(N^*) \right) \right] = \frac{|\Omega_k|}{4\pi} \left( 1 + \frac{1}{q} \right).$$

A similar argument gives

$$\lim_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E} \left[ W_{U_k}^2 \left( \frac{\mu_n(U_k)}{\mathbf{m}(U_k)} \mathbf{m}, \mathbf{m} \right) \right] = \frac{|\Omega_k|}{4\pi}. \quad (6.5)$$



*Global term.* We consider the second term in (6.4), for which we apply (6.1) obtaining

$$\mathbb{E} \left[ W_2^2 \left( \sum_k \frac{\mu_n(U_k)}{\mathbf{m}(U_k)} \chi_{U_k} \mathbf{m}, \mathbf{m} \right) \right] \lesssim \sum_k \mathbb{E} \left[ \left( \frac{\mu_n(U_k)}{\mathbf{m}(U_k)} - 1 \right)^2 \right] \mathbf{m}(U_k) \lesssim \frac{C(\varepsilon)}{n}.$$

After multiplying by  $n/\log n$  and letting first  $n \rightarrow \infty$  and then  $\varepsilon \rightarrow 0$ , the validity of (6.2) is thus settled.

To bound (6.4), we let  $\kappa = \min_k \kappa_k - \varepsilon$ , so that, as  $n \rightarrow \infty$ ,  $\kappa \rightarrow 1 - \varepsilon$  and using union and Chernoff bounds we have the crude estimate (but sufficient for our purposes)

$$P(\forall k, |\kappa - \kappa_k| > 2\varepsilon) \lesssim \frac{C(\varepsilon)}{n}. \quad (6.6)$$

We bound from above, using the subadditivity inequality (2.5) and the triangle inequality

$$\begin{aligned} \mathbb{E} \left[ W_2^2 \left( \sum_k \kappa_k \chi_{U_k} \lambda_m, \lambda_m \right) \right] &\leq \mathbb{E} \left[ W_2^2 \left( \sum_k (\kappa_k - \kappa) \chi_{U_k} \lambda_m, (1 - \kappa) \lambda_m \right) \right] \\ &\leq \mathbb{E} \left[ W_2^2 \left( \sum_k (\kappa_k - \kappa) \chi_{U_k} \lambda_m, \sum_k (\kappa_k - \kappa) \frac{\lambda_m(U_k)}{\mathbf{m}(U_k)} \chi_{U_k} \mathbf{m} \right) \right] \\ &\quad + \mathbb{E} \left[ W_2^2 \left( \sum_k (\kappa_k - \kappa) \frac{\lambda_m(U_k)}{\mathbf{m}(U_k)} \chi_{U_k} \mathbf{m}, (1 - \kappa) \mathbf{m} \right) \right] \\ &\quad + \mathbb{E} \left[ W_2^2 ((1 - \kappa) \mathbf{m}, (1 - \kappa) \lambda_m) \right]. \end{aligned} \quad (6.7)$$

For the first term, we use again subadditivity (2.5),

$$\begin{aligned} &\mathbb{E} \left[ W_2^2 \left( \sum_k (\kappa_k - \kappa) \chi_{U_k} \lambda_m, \sum_k (\kappa_k - \kappa) \frac{\lambda_m(U_k)}{\mathbf{m}(U_k)} \chi_{U_k} \mathbf{m} \right) \right] \\ &\leq \sum_k \mathbb{E} \left[ W_{U_k}^2 \left( (\kappa_k - \kappa) \lambda_m, (\kappa_k - \kappa) \frac{\lambda_m(U_k)}{\mathbf{m}(U_k)} \mathbf{m} \right) \right] \\ &\leq \sum_k C(\varepsilon) \frac{\text{diam}(E)^2}{n^4} + \varepsilon \mathbb{E} \left[ W_{U_k}^2 \left( \lambda_m, \frac{\lambda_m(U_k)}{\mathbf{m}(U_k)} \mathbf{m} \right) \right] \\ &\lesssim C(\varepsilon) \frac{\text{diam}(E)^2}{n} + \varepsilon \sum_k |\Omega_k| \frac{\log m}{m}, \end{aligned}$$

having used (6.6) and (6.5) with  $\lambda_m$  instead of  $\mu_n$ . For the second term in (6.7), we use (6.1),

$$\begin{aligned} &\mathbb{E} \left[ W_2^2 \left( \sum_k (\kappa_k - \kappa) \frac{\lambda_m(U_k)}{\mathbf{m}(U_k)} \chi_{U_k} \mathbf{m}, (1 - \kappa) \mathbf{m} \right) \right] \\ &\lesssim \sum_k \mathbb{E} \left[ \left( (\kappa_k - \kappa) \frac{\lambda_m(U_k)}{\mathbf{m}(U_k)} - (1 - \kappa) \right)^2 \right] \mathbf{m}(U_k) \\ &\lesssim \frac{C(\varepsilon)}{n}, \end{aligned}$$

having used the variance bounds

$$\mathbb{E} \left[ \left( \frac{\lambda_m(U_k)}{\mathbf{m}(U_k)} - 1 \right)^2 \right] + \mathbb{E} [(\kappa_k - 1)^2] \lesssim \frac{C(\varepsilon)}{n}.$$

For the third term in (6.7), we use that  $1 - \kappa > 2\varepsilon$  with probability smaller than  $C(\varepsilon)/n$  and then, on the complementary event, the already settled (6.3), to obtain

$$\mathbb{E} [W_2^2((1 - \kappa)\mathbf{m}, (1 - \kappa)\lambda_m)] \lesssim \frac{C(\varepsilon)}{n} + \varepsilon \frac{m}{\log m}.$$

After collecting all these bounds, we multiply by  $n/\log n$  and let first  $n \rightarrow \infty$  and then  $\varepsilon \rightarrow 0$  to conclude.  $\square$

## 7. PROOF OF COROLLARY 1.3

We recall that with our usual notation, we want to prove that for any sequence  $m = m(n) \geq n$  with  $\lim_{n \rightarrow \infty} (m - n)/\log n = 0$ , it holds

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \mathbb{E} \left[ \min_{\sigma \in \mathcal{S}_{n,m}} \sum_{i=1}^n |X_i - Y_{\sigma(i)}|^2 \right] = \frac{|\Omega|}{2\pi}.$$

The inclusion  $\mathcal{S}_n \subseteq \mathcal{S}_{n,m}$  yields immediately

$$\min_{\sigma \in \mathcal{S}_{n,m}} \sum_{i=1}^n |X_i - Y_{\sigma(i)}|^2 \leq \min_{\sigma \in \mathcal{S}_n} \sum_{i=1}^n |X_i - Y_{\sigma(i)}|^2,$$

so that it always holds, for any sequence  $m = m(n) \geq n$ ,

$$\limsup_{n \rightarrow \infty} \frac{1}{\log n} \mathbb{E} \left[ \min_{\sigma \in \mathcal{S}_{n,m}} \sum_{i=1}^n |X_i - Y_{\sigma(i)}|^2 \right] \leq \lim_{n \rightarrow \infty} \frac{1}{\log n} \mathbb{E} \left[ \min_{\sigma \in \mathcal{S}_n} \sum_{i=1}^n |X_i - Y_{\sigma(i)}|^2 \right] = \frac{|\Omega|}{2\pi}.$$

To show the converse inequality, given  $\sigma \in \mathcal{S}_{n,m}$  we induce a matching between  $(X_i)_{i=1}^m$  and  $(Y_j)_{j=1}^m$  by pairing the points  $(X_i)_{i=n+1}^m$  to those in  $(Y_j)_{j=1}^m \setminus (Y_{\sigma(i)})_{i=1}^n$ . Since  $\Omega$  is bounded, we obtain the inequality

$$\min_{\sigma \in \mathcal{S}_m} \sum_{i=1}^m |X_i - Y_{\sigma(i)}|^2 \leq \min_{\sigma \in \mathcal{S}_{n,m}} \sum_{i=1}^n |X_i - Y_{\sigma(i)}|^2 + \text{diam}(\Omega)^2(m - n).$$

Taking expectation and using the assumption  $(m - n)/\log n \rightarrow 0$ , which in particular gives  $\log n/\log m \rightarrow 1$  yields

$$\frac{|\Omega|}{2\pi} = \lim_{m \rightarrow \infty} \frac{1}{\log m} \mathbb{E} \left[ \min_{\sigma \in \mathcal{S}_m} \sum_{i=1}^m |X_i - Y_{\sigma(i)}|^2 \right] \leq \liminf_{n \rightarrow \infty} \frac{1}{\log n} \mathbb{E} \left[ \min_{\sigma \in \mathcal{S}_{n,m}} \sum_{i=1}^n |X_i - Y_{\sigma(i)}|^2 \right].$$

## 8. PROOF OF COROLLARY 1.4

Given  $(x_i)_{i=1}^n, (y_i)_{i=1}^n \subseteq \mathbb{R}^2$ , we introduce the notation

$$\mathcal{C}_{TSP}^2((x_i)_{i=1}^n) = \min_{\tau \in \mathcal{S}_n} \sum_{i=1}^n |x_{\tau(i)} - x_{\tau(i+1)}|^2$$

for the costs of the Euclidean travelling salesperson problem and

$$\mathcal{C}_{bTSP}^2((x_i)_{i=1}^n, (y_i)_{i=1}^n) = \min_{\tau, \tau' \in \mathcal{S}_n} \sum_{i=1}^n |x_{\tau(i)} - y_{\tau'(i)}|^2 + |y_{\tau'(i)} - x_{\tau(i+1)}|^2$$

for the cost of its bipartite variant (writing  $\tau(n+1) = \tau(1)$  for permutations  $\tau \in \mathcal{S}_n$ ).

The proof of the lower bound follows straightforwardly taking expectation in the inequality

$$\mathcal{C}_{bTSP}^2((X_i)_{i=1}^n, (Y_i)_{i=1}^n) \geq 2 \min_{\sigma \in \mathcal{S}_n} \sum_{i=1}^n |X_i - Y_{\sigma(i)}|^2. \quad (8.1)$$

Using Theorem 1.2, we have then

$$\liminf_{n \rightarrow \infty} \frac{1}{\log n} \mathbb{E} [\mathcal{C}_{b\text{TSP}}^2((X_i)_{i=1}^n, (Y_i)_{i=1}^n)] \geq \frac{|\Omega|}{\pi}.$$

To prove the converse inequality, let  $\tau \in \mathcal{S}_n$  be a minimizer for  $\mathcal{C}_{\text{TSP}}^2((X_i)_{i=1}^n)$  and let  $\sigma \in \mathcal{S}_n$  be an optimal matching between  $(X_i)_{i=1}^n$  and  $(Y_i)_{i=1}^n$ , with respect to the quadratic cost. We let  $\tau' = \sigma \circ \tau \in \mathcal{S}_n$ , so that

$$\mathcal{C}_{b\text{TSP}}^2((X_i)_{i=1}^n, (Y_i)_{i=1}^n) \leq \sum_{i=1}^n |X_{\tau(i)} - Y_{\sigma(\tau(i))}|^2 + |Y_{\sigma(\tau(i))} - X_{\tau(i+1)}|^2.$$

For every  $\varepsilon > 0$ , using the inequality  $|a + b|^2 \leq (1 + \varepsilon)|a|^2 + (1 + \varepsilon^{-1})|b|^2$ , we bound from above, for every  $i \in \{1, \dots, n\}$ ,

$$\begin{aligned} |Y_{\sigma(\tau(i))} - X_{\tau(i+1)}|^2 &= |(Y_{\sigma(\tau(i))} - X_{\tau(i)}) + (X_{\tau(i)} - X_{\tau(i+1)})|^2 \\ &\leq (1 + \varepsilon)|Y_{\sigma(\tau(i))} - X_{\tau(i)}|^2 + (1 + \varepsilon^{-1})|X_{\tau(i)} - X_{\tau(i+1)}|^2. \end{aligned}$$

Summing upon  $i$  gives

$$\begin{aligned} \mathcal{C}_{b\text{TSP}}^2((X_i)_{i=1}^n, (Y_i)_{i=1}^n) &\leq (2 + \varepsilon) \min_{\sigma \in \mathcal{S}_n} \sum_{i=1}^n |X_i - Y_{\sigma(i)}|^2 \\ &\quad + (1 + \varepsilon^{-1}) \mathcal{C}_{\text{TSP},2}^2((X_i)_{i=1}^n). \end{aligned}$$

We claim that

$$\limsup_{n \rightarrow \infty} \frac{1}{\log n} \mathbb{E} [\mathcal{C}_{\text{TSP}}^2((X_i)_{i=1}^n)] = 0, \quad (8.2)$$

so that we obtain, again by Theorem 1.2,

$$\limsup_{n \rightarrow \infty} \frac{1}{\log n} \mathbb{E} [\mathcal{C}_{b\text{TSP}}^2((X_i)_{i=1}^n, (Y_i)_{i=1}^n)] \leq (2 + \varepsilon) \frac{|\Omega|}{2\pi}.$$

The thesis follows letting  $\varepsilon \rightarrow 0$ .

To prove (8.2), we rely on the known (deterministic) bound

$$\sup_{(x_i)_{i=1}^n \subseteq \Omega} \mathcal{C}_{\text{TSP}}^2((x_i)_{i=1}^n) \lesssim 1,$$

where the implicit constant depends on  $\Omega$  only. We briefly recall here a proof, for the reader convenience, based on the space-filling curve heuristic [40, Section 2.6]. Without loss of generality, we consider the case  $\Omega = Q$  a square. Consider a Peano curve  $\psi : [0, 1] \rightarrow \overline{Q}$ , that is surjective and  $1/2$ -Hölder continuous. Let  $t_i \in [0, 1]$  be such that  $\psi(t_i) = x_i$  and choose  $\tau \in \mathcal{S}_n$  such that  $(t_{\tau(i)})_{i=1}^n$  are in increasing order. Then,

$$\mathcal{C}_{\text{TSP}}^2((x_i)_{i=1}^n) \leq \sum_{i=1}^n |\psi(t_{\tau(i)}) - \psi(t_{\tau(i+1)})|^2 \leq \|\psi\|_{C^{1/2}}^2 \sum_{i=1}^n |t_{\tau(i)} - t_{\tau(i+1)}| \leq 2 \|\psi\|_{C^{1/2}}.$$

**Remark 8.1.** The argument above is in fact rather general and it applies on general metric spaces provided that the cost of the quadratic bipartite matching problem is asymptotically larger than that of the travelling salesperson problem. Also, already in the Euclidean setting, it seems possible to adapt it to the case  $p \neq 2$ , but existence of the limit for the asymptotic cost of the bipartite matching problem is not known. Finally, exactly as in [14], the argument applies as well for the random bipartite quadratic 2-factor problem, i.e., for the relaxation where the single tour is replaced by a collection of disjoint ones: indeed, since the cost becomes smaller, it is sufficient to notice that the lower bound (8.1) still holds.

## APPENDIX A. PROOF OF LEMMA 5.1

Let us recall that given a bounded open set  $\Omega$  with Lipschitz boundary, and  $\{Q_k\}_k$  a Whitney decomposition of  $\Omega$  we want to construct for every  $\delta > 0$  small enough (depending on  $\Omega$ ), a partition of  $U_\delta \cup V_\delta$  of  $\Omega$  such that  $V_\delta = \{Q_k : \text{diam}(Q_k) \geq \delta\}$ , and  $U_\delta = \{\Omega_k\}_k$  with  $\Omega_k$  open,  $\text{diam}(\Omega_k) \lesssim \delta$  and  $|\Omega_k| \sim \delta^d$ .

We start by constructing a partition of the set  $A_\delta = \{d(\cdot, \Omega^c) < \sqrt{d}\delta\}$  by open sets  $\tilde{\Omega}_k$  satisfying  $\text{diam}(\tilde{\Omega}_k) \lesssim \delta$ ,  $|\tilde{\Omega}_k| \sim \delta^d$ . For this we first consider a  $\delta$ -net  $\{x_k\}_k$  of  $\partial\Omega$ , that is a family satisfying  $\partial\Omega \subset \cup_i B_\delta(x_i)$  and  $\min_{i \neq j} |x_i - x_j| \geq \frac{\delta}{2}$ . Such a family can be for instance constructed by choosing any starting point  $x_1 \in \partial\Omega$  and then setting  $x_k \in \arg\max_{\partial\Omega} d(x, \cup_{i=1}^{k-1} \{x_i\})$  as long as  $\partial\Omega$  is not covered by  $\cup_{i=1}^{k-1} B_\delta(x_i)$ . We then set

$$\tilde{\Omega}_k = \{x \in \Omega : |x - x_k| < |x - x_i| \quad \forall i \neq k\} \cap A_\delta.$$

Notice that since  $\{|x - x_k| = |x - x_i|\}$  is a hyperplane and thus of Lebesgue measure zero,  $\{\tilde{\Omega}_k\}_k$  is indeed a partition of  $A_\delta$  in disjoint open sets. Notice also that by definition of a Whitney partition, if  $Q_k \in V_\delta$  then  $Q_k \cap A_\delta = \emptyset$ . Now on the one hand, by triangle inequality it is immediate that  $\text{diam}(\tilde{\Omega}_k) \lesssim \delta$  and thus also  $|\tilde{\Omega}_k| \lesssim \delta^d$ . On the other hand, still by triangle inequality,  $B_{\frac{\delta}{4}}(x_k) \cap \Omega \subset \tilde{\Omega}_k$  so that by Lipschitz regularity of  $\Omega$ , we also obtain  $|\tilde{\Omega}_k| \gtrsim \delta^d$ .

We finally define  $U_\delta = \cup_k \{Q_k \notin V_\delta : d(Q_k, \Omega^c) > \sqrt{d}\delta\} \cup_k \{\Omega_k\}$  where

$$\Omega_k = \tilde{\Omega}_k \cup \bigcup_j \left\{ Q_j : d(Q_j, \Omega^c) < \sqrt{d}\delta, Q_j \cap A_\delta^c \neq \emptyset \text{ and } k = \arg\min\{i : \tilde{\Omega}_i \cap Q_j \neq \emptyset\} \right\}.$$

In words this means that we consider in  $U_\delta$  either cubes which are at distance greater than  $\sqrt{d}\delta$  from  $\Omega^c$  but which are not in  $V_\delta$  or we combine the sets  $\tilde{\Omega}_k$  with the cubes which intersect both  $A_\delta$  and its complement. The choice to which  $\tilde{\Omega}_k$  we associate  $Q_i$  is arbitrary (as long as they intersect). The sets  $\Omega_k$  satisfy both  $\text{diam}(\Omega_k) \lesssim \delta$  and  $|\Omega_k| \sim \delta^d$  since for every  $j$  such that  $d(Q_j, \Omega^c) < \sqrt{d}\delta$  and  $Q_j \cap A_\delta^c \neq \emptyset$ , we have  $\ell(Q_j) \sim \delta$  and thus the number of such cubes which can intersect  $\tilde{\Omega}_k$  is uniformly bounded with respect to  $\delta$ .

## APPENDIX B. DECOMPOSITION OF RIEMANNIAN MANIFOLDS

We show that Theorem 6.3 applies to compact connected smooth Riemannian manifolds  $(M, g)$ , possibly with boundary. In fact, we argue in the case of bounded connected domains  $\Omega \subseteq M$  with smooth boundary (so that  $M$  itself does not need to be compact). Notice that  $\Omega$  with the restriction of the metric  $g$  is also a compact connected smooth Riemannian manifold with boundary, but the induced distance would then be larger (and different, in general) than the restriction of the one on  $M$ , exactly as in the Euclidean case.

**Lemma B.1.** *Let  $(M, g)$  be a two-dimensional Riemannian manifold with (possibly empty) boundary  $\partial M$  and let  $\Omega \subseteq M$  be open, bounded, connected, with smooth boundary. Let also  $\rho$  be a Hölder continuous probability density on  $\Omega$  (with respect to the Riemannian volume) uniformly strictly positive and bounded from above. Then,  $(\Omega, d, \rho)$ , where  $d$  is the restriction of the Riemannian distance on  $M$ , is well-decomposable and (6.1) holds.*

*Proof.* To show that  $\Omega$  is well-decomposable, we modify the well-known argument to obtain a decomposition of  $M$  into geodesic triangles, see e.g. [24, Theorem 2.3.A.1], to take into account also the presence of the boundary  $\partial\Omega$ . To simplify the exposition, assume first that  $\partial M = \emptyset$ . Then, by compactness,  $\partial\Omega$  is the union of a finite family of closed simple  $C^1$  curves, and to simplify again let us assume that there is only one such curve  $\gamma : \mathbb{S}^1 \rightarrow M$  parametrized so that  $\dot{\gamma}(t) \neq 0$  for every  $t \in \mathbb{S}^1$ . Consider also a parametrization of the

tubular neighbourhood of  $\gamma(\mathbb{S}^1)$ , i.e., a map  $\Gamma : \mathbb{S}^1 \times (-r_0, r_0) \rightarrow M$ ,

$$\Gamma(t, r) = \exp_{\gamma(t)}(r\dot{\gamma}(t)^\perp),$$

where  $\dot{\gamma}(t)^\perp$  denotes the inward unit normal to  $\partial\Omega$  at  $\gamma(t)$ . By compactness, if  $r_0$  is sufficiently small, such a parametrization exist, is smooth (in particular Lipschitz) and

$$\Gamma(\mathbb{S}^1 \times (0, r_0)) \subseteq \Omega \quad \text{while} \quad \Gamma(\mathbb{S}^1 \times (-r_0, 0)) \subseteq \Omega^c.$$

Let us also assume that  $r_0$  is sufficiently small so that  $\exp_p$  is well-defined and invertible on a ball  $B_p(r_0)$  for every  $p \in \bar{\Omega}$  and with Lipschitz constant smaller than  $1 + \varepsilon$  (together with its inverse). Then, normal coordinates  $T_p = \exp_p^{-1}$  are well defined on the image  $\exp_p(B_p(r_0))$ .

Consider a finite mesh  $\{t_i\}_i \subseteq \mathbb{S}^1$  with  $d(t_i, t_{i+1}) < r_0/\text{Lip}(\Gamma)$  and let  $p_i = \gamma(t_i)$  and  $q_i = \Gamma(t_i, r_0/2) \in \Omega$ . Then,  $d(q_i, q_{i+1}) \leq \text{Lip}(\Gamma)|t_i - t_{i+1}| < r_0/4$ . On the other side,  $d(q_i, \partial\Omega) \geq r_0/2$ , so that  $B(q_i, r_0/4) \cap \partial\Omega = \emptyset$  and in particular the geodesic segment connecting  $q_i$  to  $q_{i+1}$  does not intersect  $\partial\Omega$ . We define  $U_i$  to be the rectangle-like region with boundaries given by the geodesic segments connecting  $p_i$  to  $q_i$ , then  $q_i$  to  $q_{i+1}$  and  $q_{i+1}$  and  $p_i$ , together with the piece of boundary between  $p_i$  and  $p_{i+1}$  (i.e.,  $\Gamma(0, s)$  with  $s \in [t_i, t_{i+1}]$ ). We also notice that the boundary is piecewise  $C^1$  (with non-tangential intersections of different pieces), hence when transformed by  $T_{p_i}$  it is Lipschitz regular.

This settles the tubular neighbourhood of  $\partial\Omega$ . To complete the construction we then consider a finite set of points  $\{p_j\} \subseteq \Omega \setminus \Gamma(\mathbb{S}^1, (0, r_0/2))$  such that for every  $q \in \Omega$  there exist  $p_j$  with  $d(p_j, q) < r_0$  and for every  $p_j$  there are distinct points  $p_k, p_\ell$  such that

$$\max\{d(p_j, p_k), d(p_k, p_\ell), d(p_j, p_\ell)\} < r_0/2.$$

We connect with geodesic segments all points  $p_j, p_k$  with  $d(p_j, p_k) < r_0/2$  as well as  $p_j, q_i^+$  with  $d(p_j, q_i^+) < r_0/2$ . Notice that none of these segments intersects  $\partial\Omega$ . As in the proof of [24, Theorem 2.3.A.1], up to enlarging the family to include intersections of these segments, we obtain a decomposition into geodesic polygons. To conclude the construction of the  $\varepsilon$ -decomposition it is sufficient to add these polygons to the sets  $U_i$  obtained above.

In the case of non-empty  $\partial M$ , the only difference is that the notion of tubular neighbourhood must take into account points  $p = \gamma(t) \in \Omega \cap \partial M$ , hence  $\Gamma(t, r)$  is defined only for  $r \in [0, r_0)$ .

Finally, inequality (6.1) follows from the validity of an  $L^2$ -Poincaré-Wirtinger inequality on  $\Omega$ , adapting the proof of [22, Lemma 3.4] to the Riemannian setting. In turn, the validity of Poincaré-Wirtinger inequality seems to be folklore on smooth connected compact Riemannian manifolds and a full proof, also in presence of a smooth boundary, may be given by gluing local inequalities [23, Section 10.1].  $\square$

## REFERENCES

- [1] AJTAI, M., KOMLÓS, J., AND TUSNÁDY, G. On optimal matchings. *Combinatorica* 4 (1984), 259–264.
- [2] AMBROSIO, L., AND GLAUDO, F. Finer estimates on the 2-dimensional matching problem. *J. Éc. polytech. Math.* 6 (2019), 737–765.
- [3] AMBROSIO, L., GLAUDO, F., AND TREVISAN, D. On the optimal map in the 2-dimensional random matching problem. *Discrete & Continuous Dynamical Systems-A* 39, 12 (2019), 7291.
- [4] AMBROSIO, L., STRA, F., AND TREVISAN, D. A PDE approach to a 2-dimensional matching problem. *Probab. Theory Relat. Fields* 173, 1-2 (2019), 433–477.
- [5] ARJOVSKY, M., CHINTALA, S., AND BOTTOU, L. Wasserstein generative adversarial networks. In *International conference on machine learning* (2017), PMLR, pp. 214–223.
- [6] ARMSTRONG, S., AND SERFATY, S. Local laws and rigidity for Coulomb gases at any temperature. *Ann. Probab.* 49, 1 (2021), 46–121.
- [7] BARTHE, F., AND BORDENAVE, C. Combinatorial optimization over two random point sets. In *Séminaire de probabilités XLV*. Cham: Springer, 2013, pp. 483–535.
- [8] BEARDWOOD, J., HALTON, J. H., AND HAMMERSLEY, J. M. The shortest path through many points. In *Mathematical Proceedings of the Cambridge Philosophical Society* (1959), vol. 55, Cambridge University Press, pp. 299–327.

- [9] BENEDETTO, D., AND CAGLIOTI, E. Euclidean random matching in 2d for non-constant densities. *Journal of Statistical Physics* 181, 3 (2020), 854–869.
- [10] BENEDETTO, D., CAGLIOTI, E., CARACCILO, S., D’ACHILLE, M., SICURO, G., AND SPORTIELLO, A. Random assignment problems on 2d manifolds. *Journal of Statistical Physics* 183, 2 (2021), 1–40.
- [11] BERTHET, P., AND FORT, J. C. Exact rate of convergence of the expected  $W_2$  distance between the empirical and true gaussian distribution. *Electronic Journal of Probability* 25 (2020), 1–16.
- [12] BOBKOV, S., AND LEDOUX, M. One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *Mem. Amer. Math. Soc.* 261, 1259 (2019), v+126.
- [13] CAFFARELLI, L. A. Boundary regularity of maps with convex potentials–II. *Annals of mathematics* 144, 3 (1996), 453–496.
- [14] CAPELLI, R., CARACCILO, S., DI GIOACCHINO, A., AND MALATESTA, E. M. Exact value for the average optimal cost of the bipartite traveling salesman and two-factor problems in two dimensions. *Physical Review E* 98, 3 (2018), 030101.
- [15] CARACCILO, S., D’ACHILLE, M., AND SICURO, G. Anomalous scaling of the optimal cost in the one-dimensional random assignment problem. *Journal of Statistical Physics* 174, 4 (2019), 846–864.
- [16] CARACCILO, S., LUCIBELLO, C., PARISI, G., AND SICURO, G. Scaling hypothesis for the Euclidean bipartite matching problem. *Physical Review E* 90, 1 (2014).
- [17] CARACCILO, S., AND SICURO, G. Scaling hypothesis for the euclidean bipartite matching problem. ii. correlation functions. *Physical Review E* 91, 6 (2015).
- [18] DEREICH, S., SCHEUTZOW, M., AND SCHOTTSTEDT, R. Constructive quantization: approximation by empirical measures. *Ann. Inst. Henri Poincaré Probab. Stat.* 49, 4 (2013), 1183–1203.
- [19] FIGALLI, A., AND GIGLI, N. A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions. *Journal de mathématiques pures et appliquées* 94, 2 (2010), 107–130.
- [20] GOLDMAN, M., AND HUESMANN, M. A fluctuation result for the displacement in the optimal matching problem. *arXiv preprint arXiv:2105.02915* (2021).
- [21] GOLDMAN, M., HUESMANN, M., AND OTTO, F. Quantitative linearization results for the monge-ampère equation. *Communications on Pure and Applied Mathematics* 74, 12 (2021), 2483–2560.
- [22] GOLDMAN, M., AND TREVISAN, D. Convergence of asymptotic costs for random euclidean matching problems. *Probability and Mathematical Physics* 2, 2 (2021), 341–362.
- [23] HAJLASZ, P., AND KOSKELA, P. Sobolev met Poincaré. *Mem. Amer. Math. Soc.* 145, 688 (2000).
- [24] JOST, J. *Compact Riemann surfaces: an introduction to contemporary mathematics*. Springer Science & Business Media, 2013.
- [25] KIM, Y.-H., AND MILMAN, E. A generalization of caffarelli’s contraction theorem via (reverse) heat flow. *Mathematische Annalen* 354, 3 (2012), 827–862.
- [26] LEDOUX, M. On optimal matching of Gaussian samples. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* 457, Veroyatnost’ i Statistika. 25 (2017), 226–264.
- [27] LEDOUX, M. On optimal matching of Gaussian samples II. *preprint* (2018).
- [28] LEIGHTON, T., AND SHOR, P. Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. *Combinatorica* 9, 2 (1989), 161–187.
- [29] LOVÁSZ, L., AND PLUMMER, M. D. *Matching theory*, vol. 367. American Mathematical Soc., 2009.
- [30] MEHTA, A. Online matching and ad allocation. *Foundations and Trends in Theoretical Computer Science* 8 (4) (2013), 265–368.
- [31] MEZARD, M., AND MONTANARI, A. *Information, physics, and computation*. Oxford University Press, 2009.
- [32] MÉZARD, M., PARISI, G., AND VIRASORO, M. A. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, vol. 9. World Scientific Publishing Company, 1987.
- [33] NEEMAN, J. Lipschitz changes of variables via heat flow. *arXiv preprint arXiv:2201.03403* (2022).
- [34] PEYRÉ, G., CUTURI, M., ET AL. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning* 11, 5-6 (2019), 355–607.
- [35] PEYRE, R. Comparison between  $W_2$  distance and  $\dot{H}^{-1}$  norm, and Localization of Wasserstein distance. *ESAIM: Control, Optimisation and Calculus of Variations* 24, 4 (2018), 1489–1501.
- [36] REDMOND, C., AND YUKICH, J. Asymptotics for Euclidean functionals with power-weighted edges. *Stochastic processes and their applications* 61, 2 (1996), 289–304.
- [37] ROSENTHAL, H. P. On the subspaces of  $L^p$  ( $p > 2$ ) spanned by sequences of independent random variables. *Israel Journal of Mathematics* 8, 3 (1970), 273–303.
- [38] SANTAMBROGIO, F. *Optimal transport for applied mathematicians*, vol. 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.
- [39] SICURO, G. *The Euclidean Matching Problem*. Springer Theses. Springer International Publishing, 2017.
- [40] STEELE, J. M. *Probability theory and combinatorial optimization*. SIAM, 1997.



- [41] STEIN, E. M. *Singular integrals and differentiability properties of functions (PMS-30), volume 30*. Princeton university press, 2016.
- [42] TALAGRAND, M. The Ajtai-Komlós-Tusnády matching theorem for general measures. In *Probability in Banach Spaces, 8: Proceedings of the Eighth International Conference (1992)*, Springer, pp. 39–54.
- [43] TALAGRAND, M. Matching random samples in many dimensions. *Ann. Appl. Probab.* 2, 4 (11 1992), 846–856.
- [44] TALAGRAND, M. *Upper and lower bounds for stochastic processes: modern methods and classical problems*, vol. 60. Springer Science & Business Media, 2014.
- [45] TALAGRAND, M. Scaling and non-standard matching theorems. *Comptes Rendus Mathématique* 356, 6 (2018), 692–695.
- [46] TRILLOS, N. G., AND SLEPČEV, D. On the rate of convergence of empirical measures in  $\infty$ -transportation distance. *Canadian Journal of Mathematics* 67, 6 (2015), 1358–1383.
- [47] VAROPOULOS, N. T., SALOFF-COSTE, L., AND COULHON, T. *Analysis and geometry on groups*. No. 100. Cambridge university press, 2008.
- [48] VILLANI, C. *Topics in optimal transportation*, vol. 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
- [49] YUKICH, J. Asymptotics for the Euclidean TSP with power weighted edges. *Probability theory and related fields* 102, 2 (1995), 203–220.
- [50] YUKICH, J. Asymptotics for weighted minimal spanning trees on random points. *Stochastic Processes and their applications* 85, 1 (2000), 123–138.
- [51] YUKICH, J. E. *Probability theory of classical Euclidean optimization problems*. Berlin: Springer, 1998.

L.A.: SCUOLA NORMALE SUPERIORE, 56126 PISA, ITALY  
*Email address:* `luigi.ambrosio@sns.it`

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`