On the Rate of Convergence of Empirical Measures in ∞-transportation Distance

Nicolás García Trillos and Dejan Slepčev

Abstract. We consider random i.i.d. samples of absolutely continuous measures on bounded connected domains. We prove an upper bound on the ∞-transportation distance between the measure and the empirical measure of the sample. The bound is optimal in terms of scaling with the number of sample points.

1 Introduction

Consider a bounded open set $D \subset \mathbb{R}^d$. Given two probability measures ν and μ on D, the *∞-transportation distance* between ν and μ is defined by

$$(1.1) \qquad d_\infty(\nu, \mu) := \inf \left\{ \text{esssup}_\gamma \left\{ |x - y| : (x, y) \in D \times D \right\} : \gamma \in \Gamma(\nu, \mu) \right\},$$

where $\Gamma(\nu, \mu)$ is the set of all *couplings* (transportation plans) between ν and μ, that is, the set of all probability measures on $D \times D$ for which the marginal on the first variable is ν and the marginal on the second variable is μ. More precisely,

$$\Gamma(\nu, \mu) = \left\{ \gamma \in \mathcal{P}(D \times D) : (\forall A - \text{Borel}) \gamma(A \times D) = \nu(A), \gamma(D \times A) = \mu(A) \right\}.$$

We consider the ∞-transportation distance between a given measure, ν, and the empirical measure associated with a random i.i.d. sample drawn from the measure ν. We consider ν to be absolutely continuous with respect to the Lebesgue measure. Our main result is the following upper bound.

Theorem 1.1 *Let $D \subseteq \mathbb{R}^d$ be a bounded, connected, open set with Lipschitz boundary. Let ν be a probability measure on D with density $\rho : D \to (0, \infty)$ such that there exists $\lambda \geq 1$ for which*

$$(1.2) \qquad (\forall x \in D) \quad \frac{1}{\lambda} \leq \rho(x) \leq \lambda.$$

Let $X_1, \ldots, X_n, \ldots$ be i.i.d. samples from ν. Consider ν_n the empirical measure

$$\nu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}.$$

1358

Then, for any fixed $\alpha > 2$, except on a set with probability $O(n^{-\alpha/2})$,

$$d_\infty(v, v_n) \leq C \begin{cases} \dfrac{\ln(n)^{3/4}}{n^{1/2}}, & \text{if } d = 2, \\[2mm] \dfrac{(\ln n)^{1/d}}{n^{1/d}}, & \text{if } d \geq 3, \end{cases}$$

where C depends only on α, D, and λ.

We also establish that the bound above is optimal in terms of scaling in n.

1.1 Background

The ∞-transportation distance $d_\infty(v, \mu)$ is the least possible maximal distance by which a transportation plan between v and μ has to move the mass. Related to it is the p-transportation distance (*i.e.*, Monge–Kantorovich–Rubinstein or p-Wasserstein distance), which measures the average of the power of the distance by which the mass is moved. For $1 \leq p < \infty$,

$$d_p(v, \mu) := \left(\inf \left\{ \int |x - y|^p d\gamma(x, y) : \gamma \in \Gamma(v, \mu) \right\} \right)^{\frac{1}{p}}.$$

The above function metrizes the weak convergence of measures on D (for D bounded). It follows from the work of Dudley [10] that for $1 \leq p < \infty$ and $d \geq 3$ and under rather general conditions on v (weaker than the ones assumed in Theorem 1.1) that the expected p-transportation distance between a measure v and the empirical measure v_n scales as $n^{-1/d}$, that is,

$$d_p(v, v_n) \sim n^{-1/d} \quad \text{for } d \geq 3.$$

A related problem consists of comparing two measures v_n and μ_n, both of which are discrete measures with the same number of points of the same mass. Then the ∞-transportation distance $d_\infty(v_n, \mu_n)$ is also known as the min-max matching distance. There are a number of works on the matchings in the case that μ_n and v_n are measures on the cube $(0,1)^d$ and v_n is the empirical measure associated with i.i.d. samples drawn from the Lebesgue measure and μ_n is either another empirical measure of another independent sample or a measure supported on a regular grid. It is worth remarking that the discrete matching results imply the estimates on the distance between v_n and v. The converse also holds. Ajtai, Komlós, and Tusnády in [1] showed optimal bounds on the p-transportation distance, for $1 \leq p < \infty$, between two empirical measures sampled from the Lebesgue measure on a square. That is, they showed that if $X_1, \ldots, X_n, \ldots \in (0,1)^2$ and $Y_1, \ldots, Y_n, \ldots \in (0,1)^2$ are two independent samples and $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$, while v_n is as before, then the minimum over all permutations π of $\{1, \ldots, n\}$ satisfies

$$\min_\pi \frac{1}{n} \sum_{i=1}^n |X_{\pi(i)} - Y_i| \leq C \sqrt{\frac{\ln n}{n}}$$

with probability $1 - o(1)$. They introduced the technique of obtaining probabilistic estimates by dyadically dividing the cube into 2^k subcubes, obtaining a matching estimate at the fine level, and estimating the transformations needed to bridge different scales to obtain an upper bound on the total distance. Our proof also relies on a similar decomposition of the domain. Dobrić and Yukich [9], Talagrand [15], Talagrand and Yukich,[16], Bolley, Guillin, and Villani [6], Boissard [5], and others later refined these results and obtained more precise information on the distribution of the p-transportation distance between a measure on a cube and the empirical measure.

For the ∞-transportation distance obtaining estimates is more delicate, since almost all of the mass needs to be matched within the desired distance to obtain the bound. Furthermore, the optimal scaling itself has a logarithmic correction compared to the case $p < \infty$.

The optimal scaling in dimension $d = 2$, for v being the Lebesgue measure, was obtained by Leighton and Shor [11]. They consider i.i.d. random samples $X_1, \ldots, X_n$ distributed according to the Lebesgue measure and points $Y_1, \ldots, Y_n$ on a regular grid. They showed that there exist $c > 0$ and $C > 0$ such that with very high probability:

$$\frac{c(\ln n)^{3/4}}{n^{1/2}} \leq \min_\pi \max_i |X_{\pi(i)} - Y_i| \leq \frac{C(\ln n)^{3/4}}{n^{1/2}},$$

where π ranges over all permutations of $\{1, \ldots, n\}$. In other words, when $d = 2$, with high probability the ∞-transportation distance between the measure μ_n and the measure v_n is of order $(\ln n)^{3/4}/n^{1/2}$. For $d \geq 3$, Shor and Yukich [12] proved the analogous result on $(0,1)^d$ with v being the Lebesgue measure restricted to $(0,1)^d$. They showed that there exist $c > 0$ and $C > 0$ such that with very high probability:

$$\frac{c(\ln n)^{1/d}}{n^{1/d}} \leq \min_\pi \max_i |X_{\pi(i)} - Y_i| \leq \frac{C(\ln n)^{1/d}}{n^{1/d}}.$$

The result in dimension $d \geq 3$ is based on the matching algorithm introduced by Ajtai, Komlós, and Tusnády in [1]. For $d = 2$ the AKT scheme still gives an upper bound, but not a sharp one. As remarked in [12], there is a crossover in the nature of the matching when $d = 2$. For $d \geq 3$, the matching length between the random points and the points in the grid is determined by the behavior of the points locally; for $d = 1$ on the other hand, the matching length is determined by the behavior of random points globally, and finally for $d = 2$ the matching length is determined by the behavior of the random points at all scales. At the level of the AKT scheme this means that for $d \geq 3$ the major source of the transportation distance is at the finest scale, for $d = 1$ at the coarsest scale, while for $d = 2$ distances at all scales are of the same size (in terms of how they scale with n). The sharp result in dimension $d = 2$ by Leighton and Shor required a more sophisticated matching procedure; an alternative proof was provided by Talagrand [15] who also provided more streamlined and conceptually clear proofs in [13, 14].

In this paper, our main contribution is that we extend the previous results to general domains and general densities.

1.2 Outline of the Approach

One of the main steps in the proof of Theorem 1.1 consists of establishing estimates for the ∞-transportation distance between two measures that are absolutely continuous with respect to the Lebesgue measure and whose densities are bounded from above and below by positive constants. We prove the following result, which is of interest on its own.

Theorem 1.2 *Let $D \subset \mathbb{R}^d$ be a bounded, connected, open set with Lipschitz boundary. Let v_1, v_2 be measures on D of the same total mass: $v_1(D) = v_2(D)$. Assume the measures are absolutely continuous with respect to Lebesgue measure and let ρ_1 and ρ_2 be their densities. Furthermore, assume that for some $\lambda > 1$, for all $x \in D$,*

$$\frac{1}{\lambda} \le \rho_i(x) \le \lambda \quad \text{for } i = 1, 2.$$

Then there exists a constant $C(\lambda, D)$ depending only on λ and D such that for all v_1, v_2 as above,

$$d_\infty(v_1, v_2) \le C(\lambda, D)\|\rho_1 - \rho_2\|_{L^\infty(D)}.$$

We use this result in proving Theorem 1.1. We first consider the domain $D = (0,1)^d$. Note that when the density ρ is constant the result was obtained by Shor and Yukich [12] in case $d \ge 3$ and by Leighton and Shor [11] in case $d = 2$. In dimensions $d \ge 3$ we use a dyadic decomposition similar to the one introduced by Ajtai, Komlós, and Tusnády (also used by Shor and Yukich). However, the fact that we adjust the densities and not the geometry of the subdomains makes it easier to handle general densities. We remark that the probabilistic estimates in [12] are similar to the ones we use in our proof of Theorem 1.1 for $d \ge 3$. To obtain the optimal scaling when $d = 2$ a more subtle approach is needed. Talagrand's [13] proof of Leighton and Shor's theorem provides flexible tools that we adapt to nonuniform densities.

We note that having the result on $(0,1)^d$ implies the result on any domain that is bi-Lipshitz homeomorphic to $(0,1)^d$.

To prove Theorem 1.1 on general domains we partition them into finite number of subdomains that can be transformed via a bi-Lipschitz map to the unit cube. The difficulty that arises is that the empirical measure on the subdomain may not have the same total mass as the restriction of the measure to the subdomain. The mass discrepancy is small, but since we seek estimates in ∞-transportation distance, the mass discrepancy needs to be carefully redistributed. Thus we introduce special ways to partition domains that enable the appropriate mass exchange between subdomains. We call the domains *well partitioned* (see Definition 3.1) if they admit the desired partitioning. We prove the Theorem 1.1 on well partitioned domains using induction on the number of subdomains needed in the partition.

We then show that all connected bounded domains with smooth boundary can be well partitioned by a careful geometric argument that uses Voronoi tessellation. The final step in proving Theorem 1.1 consists of reducing the problem to domains that are well partitioned by proving that it is possible to find a bi-Lipschitz homeomorphism between an arbitrary open, connected, bounded set D with Lipschitz boundary and

a domain that is well partitioned. In fact, by a result in [3] an open and bounded domain D with Lipschitz boundary is bi-Lipschitz homeomorphic to an open domain with smooth boundary. Finally, Proposition 3.2 says that open and bounded domains with smooth boundary are well partitioned.

The paper is organized as follows. In Subsection 1.3 we introduce notation and present some results related to the ∞-transportation distance. In Section 2 we start by proving Theorem 1.2 for $D = (0,1)^d$ and then prove Theorem 1.1 for $D = (0,1)^d$, in Subsection 2.1 when $d \geq 3$ and in Subsection 2.2 when $d = 2$. In Section 3 we define well partitioned domains and prove Theorem 1.2 in full generality; then we prove Theorem 1.1. Finally, in the Appendix we prove Proposition 3.2, which states that open and bounded domains with smooth boundary are well partitioned.

1.3   Preliminaries and Notation

Let D be an open and bounded domain in $\mathbb{R}^d$. Given a finite Borel measure ν and a Borel map $T: D \to D$, the *push-forward* of ν, denoted by $T_\sharp \nu$, is the measure such that for all $A \in \mathfrak{B}(D)$,

$$T_\sharp \nu(A) \coloneqq \nu(T^{-1}(A)).$$

We say that a Borel map $T: D \to D$ is a *transportation map* between ν and μ if $T_\sharp \nu = \mu$. Note that a transportation map T between ν and μ induces the coupling γ_T given by $\gamma_T \coloneqq (\mathrm{Id} \times T)_\sharp \nu$. A natural question that arises from the connection between transportation maps and transportation plans is the following: in the definition of $d_\infty(\nu, \mu)$ can we restrict our attention to couplings induced by transportation maps? The answer to this question is affirmative in the case where the measure ν is absolutely continuous with respect to the Lebesgue measure. In fact, this is one of the results in [7], where it is proved that there exist solutions to the problem (1.1) that are also ∞-*cyclically monotone*, that if $\nu \ll \mathcal{L}^d$, are induced by transportation maps. In this paper ν is taken to be $d\nu = \rho dx$, where ρ is bounded above and below by positive constants, and so in this setting the results in [7] can be stated as follows: if $\nu(D) = \mu(D)$, then there exists a transportation map $T^*: D \to D$ with $T_\sharp^* \nu = \mu$ and such that

$$(1.3) \qquad\qquad d_\infty(\nu, \mu) = \| T^* - \mathrm{Id} \|_{L^\infty(D)}.$$

The question of uniqueness of the optimal transportation map T^*, although interesting on its own, is not of importance for the results we present in this paper. Nevertheless, it is worth mentioning that if μ is concentrated on finitely many points, then the transportation map T^* for which (1.3) holds is unique. This is the content of [7, Theorem 5.4]. In particular, if μ is taken to be ν_n, where ν_n is the empirical measure associated with data points $X_1, \ldots, X_n$ sampled from ν, then the uniqueness of the optimal transportation map is guaranteed.

We remark that for any transportation map T_n between ν and ν_n, it holds that

$$d_\infty(\nu, \nu_n) \leq \| T_n - \mathrm{Id} \|_{L^\infty(D)}.$$

Thus, we can estimate $d_\infty(\nu, \nu_n)$ by estimating the right-hand side of the previous expression for some T_n.

2   The Matching Results for $(0,1)^d$

The first goal of this section is to prove Theorem 1.2 for $D = (0,1)^d$. In order to do this we need a few preliminary lemmas.

Lemma 2.1   *Let $Q \subseteq \mathbb{R}^d$ be a rectangular box (rectangular parallelepiped). Let Q_1, Q_2 be the rectangular boxes obtained from Q by bisecting one of its sides. Let $\rho: Q \to (0, \infty)$ be given by*

$$\rho(x) := \begin{cases} c_1, & \text{if } x \in Q_1, \\ c_2, & \text{if } x \in Q_2, \end{cases}$$

where $c_1, c_2 > 0$ are such that $1 = \frac{c_1}{2} + \frac{c_2}{2}$. Denote by ν the measure with $d\nu = \rho(x)dx$ and let ν_0 be the Lebesgue measure restricted to Q. Then

$$d_\infty(\nu_0, \nu) \le \frac{L}{2} \Big| \frac{\nu(Q_1)}{\nu_0(Q_1)} - 1 \Big|,$$

where L is the length of the side of Q bisected to generate Q_1 and Q_2.

Proof   Without the loss of generality we can assume that $Q = [0, L] \times \widehat{Q}$, where $\widehat{Q}$ is a $(d-1)$-dimensional rectangular box. Thus $Q_1 = [0, \frac{L}{2}] \times \widehat{Q}$ and $Q_2 = [\frac{L}{2}, L] \times \widehat{Q}$. Note that the condition $1 = \frac{c_1}{2} + \frac{c_2}{2}$ is equivalent to $\nu(Q) = \nu_0(Q)$. Let us introduce auxiliary functions

$$h(t) = c_1 \mathbb{1}_{[0, \frac{L}{2}]}(t) + c_2 \mathbb{1}_{(\frac{L}{2}, L]}(t) \quad \text{and} \quad f(t) = \mathbb{1}_{[0, L]}(t).$$

For $t \in [0, L]$ let $F(t) = \int_0^t ds = t$ and $H(t) = \int_0^t h(s)ds$; that is,

$$H(t) := \begin{cases} c_1 t & \text{if } 0 \le t \le \frac{L}{2}, \\ \frac{c_1}{2} L + c_2(t - \frac{L}{2}) & \text{if } \frac{L}{2} \le t \le L. \end{cases}$$

A direct computation shows that

$$H^{-1} \circ F(t) := \begin{cases} \frac{t}{c_1} & \text{if } 0 \le t \le \frac{c_1 L}{2}, \\ \frac{t}{c_2} + \frac{L}{2}(1 - \frac{c_1}{c_2}) & \text{if } \frac{c_1 L}{2} \le t \le L. \end{cases}$$

Notice that the map $T_1 := H^{-1} \circ F$ is a transportation plan between the measures dt and $h(t)dt$. Therefore, $T = T_1 \times I_{d-1}$ is a transportation plan between ν_0 and ν.

A direct computation shows that

$$|T(x) - x| = |H^{-1} \circ F(x_1) - x_1| \le \frac{L}{2}|c_1 - 1|,$$

for all $x \in Q$. Since $c_1 = \frac{\nu(Q_1)}{\nu_0(Q_1)}$, we conclude from the previous inequality that

$$\|T - \mathrm{Id}\|_{L^\infty(Q)} \le \frac{L}{2} \Big| \frac{\nu(Q_1)}{\nu_0(Q_1)} - 1 \Big|,$$

which implies the result. ∎

Lemma 2.2 *Let* $\rho\colon (0,1)^d \to (0,\infty)$ *be integrable and let* v *be the measure given by* $dv = \rho dx$. *Let* $a = \int_{(0,1)^d} \rho(x)dx$ *and denote by* v_0 *the measure on* $(0,1)^d$ *given by* $dv_0 = adx$. *Then*

$$d_\infty(v_0, v) \le \frac{\overline{C}(d)}{a} \|a - \rho\|_{L^\infty((0,1)^d)},$$

where $\overline{C}(d)$ *is a constant that depends on* d *only.*

Proof   Given that

$$d_\infty(v_0, v) = d_\infty\left(\frac{1}{a}v_0, \frac{1}{a}v\right),$$

by rescaling the densities, it is enough to prove the result for $a = 1$.

Consider first the case where $\|1 - \rho\|_{L^\infty((0,1)^d)} < 1/2$.

Step 1. For every $k \in \mathbb{N}$ we consider a partition of $[0,1]^d$ into a family $\mathcal{G}_k$ of 2^k rectangular boxes. The boxes are constructed recursively. Let $\mathcal{G}_0 = \{(0,1)^d\}$. Given the collection of boxes $\mathcal{G}_k$, the collection of rectangular boxes $\mathcal{G}_{k+1}$ is obtained by bisecting each of the rectangular boxes belonging to $\mathcal{G}_k$ through one of their longest sides. We note that all boxes in $\mathcal{G}_k$ have volume $\frac{1}{2^k}$ and have the same diameter (which depends only on k and d).

Consider $\rho_0 := 1$ and for all $k > 0$ and all $Q \in \mathcal{G}_k$, let

$$\rho_k(x) := \frac{1}{v_0(Q)} \int_Q \rho(z)dz = \frac{v(Q)}{v_0(Q)} \qquad \text{for all } x \in Q.$$

Let v_k be the measure on $(0,1)^d$ with density ρ_k. The assumption $\|1 - \rho\|_{L^\infty((0,1)^d)} < \frac{1}{2}$ implies $\frac{1}{2} \le \rho \le \frac{3}{2}$ and consequently for all k, $\frac{1}{2} \le \rho_k \le \frac{3}{2}$.

Note that for all $Q \in \mathcal{G}_k$ and all $j \ge k$, $v_j(Q) = v_k(Q) = v(Q)$. We denote by $v_{k \llcorner Q}$, the restriction of the measure v_k to Q. The relation of v to v_k on Q is similar to the one of v to v_0 on $(0,1)^d$, but the scale is smaller. We show that estimates on ∞-transportation distance on the finer scale lead to the desired estimates on the macroscopic scale. Note that

$$(2.1) \qquad\qquad d_\infty(v_k, v_{k+1}) \le \max_{Q \in \mathcal{G}_k} d_\infty(v_{k \llcorner Q}, v_{k+1 \llcorner Q})$$

and that

$$(2.2) \qquad d_\infty(v_k, v) \le \max_{Q \in \mathcal{G}_k} d_\infty(v_{k \llcorner Q}, v_{\llcorner Q}) \le \max_{Q \in \mathcal{G}_k} \operatorname{diam}(Q) \le \frac{C}{2^{k/d}},$$

where C is a constant only depending on d.

Step 2. Let $Q \in \mathcal{G}_k$ and let $Q_1, Q_2 \in \mathcal{G}_{k+1}$ be the two sub-boxes of Q. Then $v_k(Q_1) = \frac{1}{2}v_k(Q)$ and $v_0(Q_1) = \frac{1}{2}v_0(Q)$. It follows that

$$\begin{aligned}
|v(Q_1) - v_k(Q_1)| &\le |v(Q_1) - v_0(Q_1)| + |v_0(Q_1) - v_k(Q_1)| \\
&= \|\rho - 1\|_{L^\infty((0,1)^d)} v_0(Q_1) + \left|\frac{1}{2}v_0(Q) - \frac{1}{2}v_k(Q)\right| \\
&\le \|\rho - 1\|_{L^\infty((0,1)^d)} v_0(Q_1) + \frac{1}{2}\|\rho - 1\|_{L^\infty((0,1)^d)} v_0(Q) \\
&= 2\|\rho - 1\|_{L^\infty((0,1)^d)} v_0(Q_1).
\end{aligned}$$

Therefore,

$$(2.3) \qquad \frac{|v(Q_1) - v_k(Q_1)|}{v_k(Q_1)} \leq \frac{2\|\rho - 1\|_{L^\infty((0,1)^d)} v_0(Q_1)}{v_0(Q_1)/2} = 4\|\rho - 1\|_{L^\infty((0,1)^d)}.$$

Step 3. For a fixed cube $Q \in \mathcal{G}_k$, denote the value of ρ_k in Q by b. Then

$$d_\infty(v_{k \llcorner Q}, v_{k+1 \llcorner Q}) = d_\infty\left(\frac{1}{b} v_{k \llcorner Q}, \frac{1}{b} v_{k+1 \llcorner Q}\right).$$

By Lemma 2.1 and by (2.3) we have

$$d_\infty\left(\frac{1}{b} v_{k \llcorner Q}, \frac{1}{b} v_{k+1 \llcorner Q}\right) \leq \frac{1}{2^{k/d}} \left| \frac{v(Q_1)}{v_k(Q_1)} - 1 \right|$$

$$\leq \frac{4}{2^{k/d}} \|\rho - 1\|_{L^\infty((0,1)^d)}.$$

From (2.1) and the previous inequality it follows that for every $k \in \mathbb{N}$,

$$d_\infty(v_k, v_{k+1}) \leq \frac{4}{2^{k/d}} \|\rho - 1\|_{L^\infty((0,1)^d)}.$$

Choose $\widetilde{k}$ such that $2^{-\widetilde{k}/d} \leq \|\rho - 1\|_{L^\infty}$. From the previous inequality and (2.2) we deduce that

$$d_\infty(v_0, v) \leq \sum_{k=0}^{\widetilde{k}-1} d_\infty(v_k, v_{k+1}) + d_\infty(v_{\widetilde{k}}, v)$$

$$\leq 4\|\rho - 1\|_{L^\infty((0,1)^d)} \sum_{k=0}^{\widetilde{k}-1} \frac{1}{2^{k/d}} + C\frac{1}{2^{\widetilde{k}/d}}$$

$$\leq C(d)\|\rho - 1\|_{L^\infty((0,1)^d)},$$

which shows the desired result.

We now turn to the case where $\|\rho - 1\|_{L^\infty((0,1)^d)} \geq 1/2$. The desired estimate follows from

$$d_\infty(v_0, v) \leq \mathrm{diam}((0,1)^d) = \sqrt{d} \leq 2\sqrt{d}\|1 - \rho\|_{L^\infty((0,1)^d)}.$$

In conclusion, taking the larger of the constants of the cases above,

$$\overline{C} = \max\{C(d), 2\sqrt{d}\},$$

provides the desired estimate. ∎

Proof of Theorem 1.2 for $D = (0,1)^d$. Suppose first that $\|\rho_1 - \rho_2\|_{L^\infty((0,1)^d)} \leq \frac{1}{2\lambda}$. Let $g(x) = \rho_1(x) - \rho_2(x) + \frac{1}{\lambda}$. Note that $g \geq 0$ and that

$$\rho_1 = \left(\rho_2 - \frac{1}{\lambda}\right) + g, \quad \rho_2 = \left(\rho_2 - \frac{1}{\lambda}\right) + \frac{1}{\lambda}.$$

By Lemma 2.2 and by (1.3), there exists a transportation map T between the measures $g\,dx$ and $\frac{1}{\lambda}\,dx$ such that

$$\|T - \mathrm{Id}\|_{L^\infty((0,1)^d)} \leq \lambda C(d)\left\|g - \frac{1}{\lambda}\right\|_{L^\infty((0,1)^d)} = \lambda C(d)\|\rho_1 - \rho_2\|_{L^\infty((0,1)^d)}.$$

Note that

$$\gamma := (\mathrm{Id} \times \mathrm{Id})_\sharp \left(\rho_2 - \frac{1}{\lambda}\right) dx + (\mathrm{Id} \times T)_\sharp g\,dx \in \Gamma(v_1, v_2).$$

Moreover for γ-a.e. $(x, y) \in (0,1)^d \times (0,1)^d$,

$$|x - y| \leq \lambda C(d) \|\rho_1 - \rho_2\|_{L^\infty((0,1)^d)}.$$

Thus,

$$d_\infty(\nu_1, \nu_2) \leq \lambda C(d) \|\rho_1 - \rho_2\|_{L^\infty((0,1)^d)}.$$

To get our estimate in case $\|\rho_1 - \rho_2\|_{L^\infty} > \frac{1}{2\lambda}$, note that

$$d_\infty(\nu_1, \nu_2) \leq \mathrm{diam}((0,1)^d) = \sqrt{d} \leq 2\lambda\sqrt{d} \|\rho_1 - \rho_2\|_{L^\infty((0,1)^d)}. \qquad \blacksquare$$

Remark 2.3   Note that from the previous proof, Theorem 1.2 is true for any domain D of the form $D = (a_1, b_1) \times \cdots \times (a_d, b_d)$. To deduce this fact, it is enough to consider a translation and rescaling of the coordinate axes to transform the rectangular box D into the unit box $(0,1)^d$ and then use Theorem 1.2 for the unit cube.

2.1   The Matching Results for $(0,1)^d$: $d \geq 3$.

Now we prove Theorem 1.1 for $D = (0,1)^d$ when $d \geq 3$. To achieve this it is useful to consider a partition of the cube $(0,1)^d$ into rectangular boxes analogous to the ones used in the proof of Lemma 2.2. The main difference is that we divide rectangular boxes into sub-boxes of the same ν-measure, instead of the same Lebesgue measure.

Let $\rho: (0,1)^d \to (0, \infty)$ be a density function satisfying $1/\lambda \leq \rho \leq \lambda$. For every $k \in \mathbb{N}$ we construct a family $\mathcal{F}_k$ of 2^k rectangular boxes that partition the cube $(0,1)^d$ with each rectangular box having ν-volume equal to $\frac{1}{2^k}$ and aspect ratio (ratio between its longest side and its shortest side) controlled in terms of λ. We let $\mathcal{F}_0 = \{(0,1)^d\}$. For $k = 1$ we construct rectangular boxes Q_1 and Q_2 by bisecting one of the sides (say the one lying on the first coordinate) of the cube $(0,1)^d$ using the measure ν. That is, we define $Q_1 := (0, a) \times (0,1)^{d-1}$ and $Q_2 := [a, 1) \times (0,1)^{d-1}$ where $a \in (0,1)$ is such that $\nu_{Q_1} = 1/2\nu(Q)$. Recursively, the collection of rectangular boxes at level $k + 1$ is obtained by bisecting, according the measure ν, each rectangular box from level k through one of its longest sides.

Lemma 2.4   *The aspect ratio of every rectangular box in $\mathcal{F}_k$ is bounded by $2\lambda^2$.*

Proof   We show that for every $k \in \mathbb{N}$, every rectangular box in $\mathcal{F}_k$ has aspect ratio less than $2\lambda^2$. The proof is by induction on k.

Base Case: At level $k = 1$ we consider $Q_1 = (0, a) \times (0,1)^{d-1}$, a chosen so that $\nu(Q_1) = 1/2$. Note that the aspect ratio of Q_1 is equal to $1/a$. Notice that,

$$\frac{1}{2} = \int_{Q_1} \rho(x)dx \leq a\lambda.$$

From this we conclude that the aspect ratio of Q_1 is no larger than 2λ and in particular no larger than $2\lambda^2$. By symmetry, the aspect ratio of Q_2 is no larger than $2\lambda^2$.

Inductive Step. Suppose that the aspect ratio of every rectangular box in $\mathcal{F}_k$ is bounded by $2\lambda^2$. Let Q be a rectangular box in $\mathcal{F}_{k+1}$. Note that Q is obtained by bisecting (using the measure ν) the longest side of a rectangular box $Q' \in \mathcal{F}_k$. Without the

loss of generality we can assume that $\overline{Q'} = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d]$ and that $\overline{Q} = [a_1, c] \times [a_2, b_2] \times \cdots \times [a_d, b_d]$, where $a_1 < c < b_1$. If (a_1, c) is not the smallest side of Q, then the aspect ratio of Q is no greater than the aspect ratio of Q' and hence by the induction hypothesis is less than $2\lambda^2$. If on the other hand (a_1, c) is the smallest side of Q then we let (a_i, b_i) be the longest side of Q; the aspect ratio of Q is then equal to $\frac{b_i - a_i}{c - a_1}$. Since (a_1, b_1) is the longest side of $\widetilde{Q}$, we have

$$\frac{b_i - a_i}{c - a_1} = \frac{b_1 - a_1}{c - a_1} \frac{b_i - a_i}{b_1 - a_1} \leq \frac{b_1 - a_1}{c - a_1}.$$

Finally, since $v(Q) = \frac{1}{2} v(\widetilde{Q})$, we deduce that

$$(c - a_1)\lambda \geq \frac{1}{2\lambda}(b_1 - a_1).$$

This implies the desired result.   ∎

The proof of Theorem 1.1 requires estimating how many of the sampled points fall in certain rectangles. These estimates rely on two concentration inequalities for binomial random variables, which we now recall. Let $S_m \sim \text{Bin}(m, p)$ be a binomial random variable, with m trials and probability of success for each trial of p. Chernoff's inequality [8] states that

$$(2.4) \qquad P\Big(\Big| \frac{S_m}{m} - p \Big| \geq t \Big) \leq 2\exp(-2mt^2).$$

Bernstein's inequality [4], which is sharper for small values of p gives that

$$(2.5) \qquad P\Big(\Big| \frac{S_m}{m} - p \Big| \geq t \Big) \leq 2\exp\Big(-\frac{\frac{1}{2}m^2 t^2}{mp(1-p) + \frac{1}{3}mt} \Big).$$

Proof of Theorem 1.1 for $D = (0,1)^d$ when $d \geq 3$.

Step 1. Let $\rho_0 := \rho$ and let $\mu_0 := v$. For every $Q \in \mathcal{F}_k$, consider

$$(2.6) \qquad \rho_k(x) := \frac{v_n(Q)}{v(Q)}\rho(x) = \frac{v_n(Q)}{2^{-k}}\rho(x) \qquad \text{for all } x \in Q.$$

Let μ_k be the measure with density ρ_k. Note that for all $Q \in \mathcal{F}_k$, and all $j \geq k$, $\mu_j(Q) = \mu_k(Q) = v_n(Q)$. Since by construction $v(Q) = 2^{-k}$, $nv_n(Q)$ is a binomial random variable with n trials and probability of success for each trial of $p = 2^{-k}$. Fix $\alpha > 2$ and let

$$k_n := \log_2\Big(\frac{n}{10\alpha \ln n} \Big).$$

Consider $k \in \mathbb{N}$ with $k \leq k_n$. Using Bernstein's inequality (2.5) with $t = \frac{p}{2}$ we obtain

$$(2.7) \qquad P\Big(\Big| v_n(Q) - \frac{1}{2^k} \Big| \geq \frac{1}{2^{k+1}} \Big) \leq 2\exp\Big(-\frac{\frac{1}{2} \cdot \frac{1}{4} n^2 p^2}{np(1-p) + \frac{1}{3} \cdot \frac{1}{2} np} \Big)$$

$$\leq 2\exp\Big(-\frac{1}{10} np \Big)$$

$$\leq 2\exp\Big(-\frac{1}{10} n \frac{10\alpha \ln n}{n} \Big) = 2n^{-\alpha}.$$

Since the probability of the union of events is less or equal to the sum of the probability of the events, we obtain

$$P\Big(\max_{Q\in\mathcal{F}_k}\Big| v_n(Q) - \frac{1}{2^k} \Big| \geq \frac{1}{2^{k+1}} \Big) \leq 2^k 2 n^{-\alpha}.$$

Summing over all $k \leq k_n$, we deduce that with probability at least $1 - n^{-\alpha/2}$,

$$(2.8) \qquad\qquad \frac{1}{2\lambda} \leq \rho_k \leq \frac{3\lambda}{2} \qquad \text{on } (0,1)^d,$$

for every $k \leq k_n$.

Let $Q \in \mathcal{F}_k$ and let $Q_1, Q_2 \in \mathcal{F}_{k+1}$ be the sub-boxes of Q. Let $m = n v_n(Q)$. Since $v(Q_1) = 2^{-(k+1)} = \frac{1}{2}v(Q)$, $m\frac{v_n(Q_1)}{v_n(Q)} \sim \mathrm{Bin}(m, \frac{1}{2})$ given $v_n(Q)$. Using Chernoff's bound (2.4) and (2.7), we deduce that

$$P\Big(\Big| \frac{v_n(Q_1)}{v_n(Q)} - \frac{1}{2} \Big| \geq \sqrt{\frac{\alpha 2^k \ln n}{n}} \Big) \leq 4 n^{-\alpha}.$$

Using the previous inequality, (2.6), and a union bound, we conclude that

$$P\Big(\sup_{x\in(0,1)^d} \Big| \frac{\rho_{k+1}(x)}{\rho_k(x)} - 1 \Big| \geq 2\sqrt{\frac{\alpha 2^k \ln n}{n}} \Big) \leq 2^k 4 n^{-\alpha}.$$

Summing over all $k \leq k_n$, we deduce that with probability at least $1 - n^{-\alpha/2}$,

$$(2.9) \qquad\qquad \sup_{x\in(0,1)^d} \Big| \frac{\rho_{k+1}(x)}{\rho_k(x)} - 1 \Big| \leq 2\sqrt{\frac{\alpha 2^k \ln n}{n}}$$

for every $k \leq k_n$.

Notice that for all $Q \in \mathcal{F}_k$, and all $j \geq k$, $\mu_j(Q) = \mu_k(Q) = v_n(Q)$. Then

$$d_\infty(\mu_k, \mu_{k+1}) \leq \max_{Q\in\mathcal{F}_k} d_\infty(\mu_{k\llcorner Q}, \mu_{k+1\llcorner Q}),$$

and

$$(2.10) \qquad d_\infty(\mu_k, v_n) \leq \max_{Q\in\mathcal{F}_k} d_\infty(\mu_{k\llcorner Q}, v_{n\llcorner Q}) \leq \max_{Q\in\mathcal{F}_k} \mathrm{diam}(Q) \leq C(\lambda)\frac{1}{2^{k/d}},$$

where $C(\lambda)$ is a constant only depending on λ; the last inequality in the previous expression is obtained from Lemma 2.4 and from the fact that $v(Q) = 2^{-k}$.

Using estimates (2.8) and (2.9)

$$\|\rho_k - \rho_{k+1}\|_{L^\infty((0,1)^d)} \leq \|\rho_k\|_{L^\infty((0,1)^d)}\Big\| \frac{\rho_{k+1}}{\rho_k} - 1 \Big\|_{L^\infty((0,1)^d)} \leq 2\lambda\Big(\alpha 2^k \frac{\ln n}{n} \Big)^{1/2},$$

with probability at least $1 - n^{-\alpha/2}$. Hence from Lemma 2.4 and Remark 2.3, we deduce that for all $Q \in \mathcal{F}_k$

$$d_\infty(\mu_k|_Q, \mu_{k+1\llcorner Q}) \leq C(\lambda, d)\,\mathrm{diam}(Q)\Big(\alpha 2^k \frac{\ln n}{n} \Big)^{1/2} \leq C(\lambda, d)\frac{1}{2^{k/d}}\Big(\alpha 2^k \frac{\ln n}{n} \Big)^{1/2}.$$

Using Lemma 2.1 and the previous inequalities, we conclude that except on a set with probability $O(n^{-\alpha/2})$, for every $k = 0, \ldots, k_n$,

$$d_\infty(\mu_k, \mu_{k+1}) \le C \frac{1}{2^{k/d}} \left(2^k \frac{\ln n}{n} \right)^{1/2},$$

for some constant C depending only on λ, α, and d. From the triangle inequality and (2.10), we obtain

$$d_\infty(\nu, \nu_n) \le \sum_{k=1}^{k_n} d_\infty(\mu_{k-1}, \mu_k) + d_\infty(\mu_{k_n}, \nu_n)$$

$$\le C \left(\sum_{k=1}^{k_n} \frac{1}{2^{k/d}} \left(\alpha 2^k \frac{\ln n}{n} \right)^{1/2} + \frac{(\ln n)^{1/d}}{n^{1/d}} \right)$$

$$\le C \left(\left(\frac{\ln n}{n} \right)^{1/2} \sum_{k=1}^{k_n} 2^{k(1/2-1/d)} + \frac{(\ln n)^{1/d}}{n^{1/d}} \right).$$

Given that $d \ge 3$, the term $\sum_{k=1}^{k_n} 2^{k(1/2-1/d)}$ is $O(1)$ and thus the previous expression is $O((\ln n)^{1/d}/n^{1/d})$. In summary, except on a set with probability $O(n^{-\alpha/2})$

$$d_\infty(\nu, \nu_n) \le C \frac{(\ln n)^{1/d}}{n^{1/d}},$$

where C is a constant that depends only on α, λ, and d. ∎

2.2 The Matching Results for $(0,1)^2$.

Now we prove Theorem 1.1 for $D = (0,1)^2$. We actually state and prove a stronger result that is in agreement with the result by Talagrand in [13]. The improvement with respect to the statement of Theorem 1.1 has to do with the speed of decay of the tail probability of the transportation distance. Theorem 1.1 is an immediate consequence of the following theorem.

Theorem 2.5 *Suppose that $\rho: (0,1)^2 \to (0,\infty)$ is a density function satisfying*

$$(2.11) \qquad\qquad \frac{1}{\lambda} \le \rho \le \lambda$$

for some $\lambda > 1$. Let $X_1, \ldots, X_n$ be i.i.d. samples from ρ and denote by ν_n the empirical measure

$$\nu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}.$$

Then there is a constant $L > 0$ depending only on λ such that except on a set with probability $L \exp(-(\ln n)^{3/2}/L)$, we have

$$d_\infty(\nu, \nu_n) \le L \frac{(\ln n)^{3/4}}{n^{1/2}}.$$

In order to match the empirical measure ν_n with the measure ν, we consider a partition of $(0,1)^2$ into n rectangles $Q_1, \ldots, Q_n$, each of which has ν-measure equal

to $1/n$. We then look for a bijection between the set of points $X_1, \ldots, X_n$ and the set $\{Q_1, \ldots, Q_n\}$, in such a way that every data point is matched to a nearby rectangle. Note however, that in order to guarantee that all points within a rectangle are close to the corresponding data point, we should be able to control the diameter of all the Q_is. This is is important, since we want to obtain estimates on $d_\infty(v, v_n)$. With a slight modification to the construction preceding Lemma 2.4, we obtain the following lemma.

Lemma 2.6   *Let $\rho: (0,1)^2 \to (0, \infty)$ be a density function satisfying (2.11) and let v be the measure $dv = \rho \, dx$. Then for any $n \in \mathbb{N}$ there exists a collection of rectangles $\{Q_i : i = 1, \ldots, n\}$ that partitions $[0,1]^2$ such that the aspect ratio of all rectangles is less than $3\lambda^2$ and their volume according to v is $1/n$. In particular, for every Q_i,*

$$(2.12) \qquad\qquad \mathrm{diam}(Q_i) \le \frac{C(\lambda)}{\sqrt{n}},$$

where $C(\lambda)$ is a constant only depending on λ.

The task now is to show that with high probability we can indeed find a matching between the points $X_1, \ldots, X_n$ and the rectangles $Q_1, \ldots, Q_n$, in such a way that every point is close to its matched rectangle. When $\rho \equiv 1$, the previous statement is directly related to the result of Leighton and Shor [11]. The proof of Leighton and Shor depends on discrepancy estimates over all regions R formed by squares from a suitable regular grid G' defined on D. By discrepancy we mean the difference between $v(R)$ and $v_n(R)$ for a given region R. Obtaining a uniform bound on the discrepancy over all regions R can be interpreted as obtaining probabilistic estimates on the supremum of a stochastic process indexed by the mentioned class of regions R. A conceptually clear and efficient proof of this matching result, based on obtaining upper bounds of stochastic processes, was presented by Talagrand [13,14]. In order to prove Theorem 2.5 we follow the framework of Talagrand and start by stating a general result on obtaining bounds on the supremum of more general stochastic processes ([13, Section 1]).

Let (Y, d) be an arbitrary metric space. For $n \in \mathbb{N}$ define,

$$e_n(Y, d) = \inf \sup_{y \in Y} d(y, Y_n),$$

where the infimum is taken over all subsets Y_n of Y with cardinality less than 2^{2^n}. Let $\{A_n\}_{n \in \mathbb{N}}$ be a sequence of partitions of Y. This sequence of partitions is called *admissible* if it is increasing (in the sense that for every n, A_{n+1} is a refinement of A_n) and it is such that the cardinality of A_n is no bigger than 2^{2^n}. For a given $y \in Y$ and $\{A_n\}_{n \in \mathbb{N}}$ admissible, $A_n(y)$ represents the unique set in A_n containing y. For an $\alpha > 0$, consider

$$\gamma_\alpha(Y, d) = \inf \sup_{y \in Y} \sum_{n \ge 0} 2^{n/\alpha} \, \mathrm{diam}(A_n(y)),$$

where $\mathrm{diam}(A_n(y))$ represents the diameter of the set $A_n(y)$ using the distance function d and where the infimum is taken over all $\{A_n\}_{n \in \mathbb{N}}$ admissible sequences of partitions of Y. With these definitions we can now state [13, Theorem 1.2.9].

Lemma 2.7 *Let Y be a set and let d_1, d_2 be two distance functions defined on Y. Let $\{Z_y\}_{y \in Y}$ be a stochastic process satisfying: for all $y, y' \in Y$ and all $u > 0$,*

$$(2.13) \qquad \mathbb{P}\Big(|Z_y - Z_{y'}| \geq u\Big) \leq 2 \exp\Big(-\min\Big(\frac{u^2}{d_2(y, y')^2}, \frac{u}{d_1(y, y')}\Big)\Big),$$

and also $\mathbb{E}[Z_y] = 0$ for all $y \in Y$. Then there is a constant $L > 0$ large enough such that for all $u_1, u_2 > 0$

$$(2.14) \qquad \mathbb{P}\big(\sup_{y \in Y} |Z_y - Z_{y_0}| \geq L(\gamma_1(Y, d_1) + \gamma_2(Y, d_2)) + u_1 D_1 + u_2 D_2\big)$$

$$\leq L \exp\big(-\min\{u_2^2, u_1\}\big),$$

where $D_1 = 2 \sum_{n \geq 0} e_n(Y, d_1)$ and $D_2 = 2 \sum_{n \geq 0} e_n(Y, d_2)$.

One of the consequences of the previous lemma is the following. In order to prove a tail estimate of the supremum of the stochastic process $\{Z_y\}_{y \in Y}$, like the one in (2.14), one needs to do two things. First, estimate the quantities $\gamma_1(Y, d_1)$, $\gamma_2(Y, d_2)$, D_1 and D_2. Note that these quantities depend only on the distances d_1, d_2 and hence are not a priori related to the process $\{Z_y\}_{y \in Y}$. Secondly, relate the stochastic process $\{Z_y\}_{y \in Y}$ with the distances d_1, d_2 by establishing condition (2.13).

We are now ready to prove Theorem 2.5. As mentioned earlier, this result is an adaptation of the proof by Talagrand of Leighton and Shor's theorem. We sketch some of the main steps in the proof by Talagrand and give the details on how to generalize it to non-constant densities.

Proof of Theorem 2.5 In what follows $L > 0$ is a constant that may increase from line to line.

Discrepancy estimates. Let l_1 be the largest integer such that $2^{-l_1} \geq (\ln n)^{3/4}/\sqrt{n}$. Consider G to be the regular grid of mesh 2^{-l_1} given by

$$(2.15) \qquad G = \big\{(x_1, x_2) \in [0, 1]^2; 2^{l_1} x_1 \in \mathbb{N} \text{ or } 2^{l_1} x_2 \in \mathbb{N}\big\}$$

A *vertex* of the grid G is a point (x_1, x_2) in $[0, 1]^2$ such that $2^{l_1} x_1 \in \mathbb{N}$ and $2^{l_1} x_2 \in \mathbb{N}$. A *square* of the grid G is a square of side length equal to 2^{-l_1} and whose edges belong to G. The edges are included in the squares.

For a given vertex w of G and a given integer k, consider $\mathcal{C}(w, k)$ the set of simple closed curves that lie on G that contain the vertex w and have length $l(C) \leq 2^k$. Note that every closed simple curve C in $\mathbb{R}^2$ divides the space into two regions, one of which is bounded; this latter one is called the *interior* of the curve C and is denoted by C°. For $C, C' \in \mathcal{C}(w, k)$ set $d_1(C, C') = 1$ if $C \neq C'$ and $d_1(C, C') = 0$ if $C = C'$. Also set $d_2(C, C') = \sqrt{n} \|\chi_{C^\circ} - \chi_{C'^\circ}\|_{L^2(D)}$.

Claim 1: For a given vertex w of G and a given integer k with $k \leq l_1 + 2$, there exists $L > 0$ large enough such that with probability at least $1 - L \exp(-(\ln n)^{3/2}/L)$,

$$\sup_{C \in \mathcal{C}(w, k)} \big| \sum_{i \leq n} \big(\chi_{C^\circ}(X_i) - \nu(C^\circ)\big)\big| \leq L 2^k \sqrt{n}(\ln n)^{3/4}.$$

To prove the claim, the idea is to study the supremum of the stochastic process $\{Z_C\}_{C \in \mathcal{C}(w,k)}$, where

$$Z_C := \frac{1}{L} \sum_{i \leq n} \left(\chi_{C^\circ}(X_i) - \nu(C^\circ) \right).$$

For fixed $C, C' \in \mathcal{C}(w,k)$ one can write the difference $Z_C - Z_{C'}$ as

$$Z_C - Z_{C'} = \sum_{i \leq n} Z_i,$$

where $Z_i = \frac{1}{L}\left(\chi_{C^\circ}(X_i) - \chi_{C'^\circ}(X_i) - \nu(C^\circ) + \nu(C'^\circ) \right)$. The random variables $\{Z_i\}_{i \leq n}$ are independent and identically distributed with mean zero; they satisfy $|Z_i| \leq \frac{2}{L}$, and furthermore, their variance σ^2 is bounded by

$$\sigma^2 \leq \frac{1}{L^2} \mathbb{E}\left[|\chi_{C^\circ}(X_i) - \chi_{C'^\circ}(X_i)|^2 \right] \leq \frac{\lambda}{L^2} \|\chi_{C^\circ} - \chi_{C'^\circ}\|_{L^2(D)}^2.$$

Using Bernstein's inequality and choosing $L > 0$ to be large enough, we obtain

$$\mathbb{P}\left(|Z_C - Z_{C'}| \geq u \right) \leq 2\exp\left(-\frac{u^2}{n\|\chi_{C^\circ} - \chi_{C'^\circ}\|_{L^2(D)}^2 + u} \right)$$

$$= 2\exp\left(-\min\left(\frac{u^2}{d_2(C,C')^2}, \frac{u}{d_1(C,C')} \right) \right).$$

In the proof of [13, Proposition 3.4.3] in Talagrand, the estimates $\gamma_1(\mathcal{C}(w,k), d_1) \leq L2^k\sqrt{n}$, $\gamma_2(\mathcal{C}(w,k), d_2) \leq L2^k\sqrt{n}(\ln n)^{3/4}$, $D_1 \leq 2(k + l_1 + 1)$, and $D_2 \leq L2^{k+1}\sqrt{n}$ are established. Setting $u_1 = (\ln n)^{3/2}$ and $u_2 = (\ln n)^{3/4}$ one can use Lemma 2.7 (with $Y = \mathcal{C}(w,k)$, d_1, d_2 as above and $y_0 = \{w\}$) to prove the claim.

Considering all possible vertices w of G and all possible integers k with $-l_1 \leq k \leq l_1 + 2$. It is a direct consequence of Claim 1 above that with probability at least $1 - L\exp(-(\ln n)^{3/2}/L)$,

$$(2.16) \qquad \sup_C |\sum_{i \leq n} \left(\chi_{C^\circ}(X_i) - \nu(C^\circ) \right)| \leq Ll(C)\sqrt{n}(\ln n)^{3/4},$$

where the supremum is taken over all C closed, simple curves on G. See the proof of [13, Theorem 3.4.2]. We denote by Ω_n the event for which (2.16) holds.

Enlarging Regions. Consider an integer l_2 with $l_2 < l_1$. We consider G' the grid defined as in (2.15) but with mesh size 2^{-l_2}. Note that in particular $G' \subseteq G$. Let R be a union of squares of the grid G'. One can define R' to be the region formed by taking the union of all the squares in G' with at least one side contained in R. With no change in the proof of [13, Theorem 3.4.1], it follows from the discrepancy estimates obtained previously that in the event Ω_n one has

$$(2.17) \qquad \nu(R') \geq \nu_n(R)$$

for all regions R formed with squares from G', provided that

$$2^{-l_2} \geq \frac{2^6 L}{\sqrt{n}}(\ln n)^{3/4}.$$

What this is saying is that given the discrepancy estimates obtained previously, in the event Ω_n, for any region R formed by taking the union of squares in G', one can

enlarge R a bit to obtain a region R' in such a way that the area of the enlarged region R' according to ν is greater than the area of the original region R according to ν_n. It is worth remarking that the restriction to the number 2^{-l_2} (the mesh size of G'), for this to be possible, coincides with the scaling for the transportation cost we are after.

Matching between rectangles and data points. We choose l_2 to be the largest integer satisfying

$$2^{-l_2} \geq \frac{2^6 L}{\sqrt{n}} (\ln n)^{3/4}.$$

Consider $\{Q_1, \ldots, Q_n\}$, the rectangles constructed from Lemma 2.6. Let $B_i = \{j \leq n : \operatorname{dist}(X_i, Q_j) \leq 2\sqrt{2} \cdot 2^{-l_2}\}$ for $i \in \{1, \ldots, n\}$

Claim 2: In the event Ω_n, there is a bijection $\pi \colon \{1, \ldots, n\} \to \{1, \ldots, n\}$ with $\pi(i) \in B_i$ for all i.

By the Hall marriage lemma, to prove this claim it is enough to prove that for every $I \subseteq \{X_1, \ldots, X_n\}$, the cardinality of $\bigcup_{i \in I} B_i$ is greater than the cardinality of I. Fix $I \subseteq \{1, \ldots, n\}$ and denote by R_I the region formed with the squares of G' that contain at least one of the points X_i with $i \in I$. Now, take

$$J = \{j \leq n : Q_j \cap (R_I)' \neq \varnothing\};$$

then $J \subseteq \bigcup_{i \in I} B_i$. From the properties of the boxes Q_i and from (2.17) it follows that $\# \bigcup_{i \in I} B_i \geq \#J = n\nu(\bigcup_{j \in J} Q_j) \geq n\nu((R_I)') \geq \#I$. This proves the claim.

Finally, we construct a transportation map T_n between ν and ν_n. Indeed, for x in Q_i, set $T_n(x) = X_{\pi^{-1}(i)}$. From the properties of the boxes Q_i, it is straightforward to check that $T_{n \sharp} \nu = \nu_n$ and that $\|T_n - \operatorname{Id}\|_{L^\infty(D)} \leq L \frac{(\ln n)^{3/4}}{\sqrt{n}}$ due to the estimate on the diameter of the rectangles Q_i in (2.12). ∎

3   The Matching Results for General D

The goal of this section is to prove the optimal bounds on matching for all open, connected, bounded domains D with Lipschitz boundary. In order to achieve this, we first prove Theorem 1.2 for general domains D. It is useful to consider first a class of domains D that are well partitioned.

Definition 3.1   Let $D \subseteq \mathbb{R}^d$. We say that D satisfies the (WP) property with k polytopes if D is an open, bounded and connected set and is such that there exists a finite family of closed convex polytopes $\{A_i\}_{i=1}^k$ covering D and satisfying the following. For all $i, j = 1, \ldots, k$:

(i) $\operatorname{int}(A_i) \cap D \neq \varnothing$;

(ii)   if $i \neq j$ then $\operatorname{int}(A_i) \cap \operatorname{int}(A_j) = \varnothing$;

(iii) $A_i \cap \overline{D}$ is bi-Lipschitz homeomorphic to a closed cube.

The class of domains satisfying the (WP) property is convenient for our purposes for two reasons. The first one because as we see below, in order to prove the matching results for sets with the (WP) property, we can use induction on the number of polytopes. The second reason, has to do with the fact that the class of sets that are well

partitioned contains the class of open, bounded, connected domains with smooth boundary. This is the content of the next proposition whose proof is presented in the Appendix.

Proposition 3.2   *Let $D \subseteq \mathbb{R}^d$ be an open, bounded, and connected domain with smooth boundary. Then, D satisfies the (WP) property with k polytopes for some $k \in \mathbb{N}$.*

We now prove a lemma that prepares the ground for an inductive argument to be used in the proof of the matching results for domains with the (WP) property.

Lemma 3.3   *Suppose that D is a domain that satisfies hypothesis (WP) with k polytopes ($k > 1$). Let $\{A_i\}_{i=1}^k$ be associated polytopes. Then there exists j such that $D' := D \smallsetminus A_j$ is connected.*

Proof   We say that $A_l \sim A_m$ if $\text{resint}(\partial A_m) \cap \text{resint}(\partial A_l) \cap D \neq \varnothing$, where $\text{resint}(\partial A_i)$ is the union of the relative interiors of the facets of A_i ($(d-1)$-dimensional faces). This relation induces a graph $G = (V, E)$ where the set of nodes V is the set of polytopes A_i and where an edge between A_m and A_l ($m \neq l$) belongs to the graph if and only if $A_m \sim A_l$. We claim that G is a connected graph.

Indeed, consider $m \neq l$. We want to show that there exists a path in the graph G connecting A_m with A_l. For this purpose consider $x \in \text{int}(A_m) \cap D$ and $y \in \text{int}(A_l) \cap D$. Denote by C the union of all the ridges ($(d-2)$-dimensional faces) of all the polytopes A_i. Given that C is the union of finitely many $(d-2)$-dimensional objects in $\mathbb{R}^d$, we conclude that $D \smallsetminus C$ is a connected open set and as such it is path connected. Since $x \in \text{int}(A_m) \cap D$ and $y \in \text{int}(A_l) \cap D$, in particular $x, y \in D \smallsetminus C$ and so there exists a continuous function $\gamma \colon [0,1] \to D \smallsetminus C$ such that $\gamma(0) = x$ and $\gamma(1) = y$. Let $A_{i_0}, A_{i_1}, \ldots, A_{i_N}$ be the polytopes visited by the path γ in order of appearance; this list satisfies $A_{i_s} \neq A_{i_{s+1}}$ for all s, $A_{i_0} = A_m$ and $A_{i_N} = A_l$. Now, note that for any given s, the path γ intersects $\partial A_{i_s} \cap \partial A_{i_{s+1}}$ at a point that belongs to the relative interior of a facet ($d-1$ dimensional face) of A_{i_s} and of $A_{i_{s+1}}$; this because γ lies in $D \smallsetminus C$. From this fact we conclude that $A_{i_s} \sim A_{i_{s+1}}$, and hence there is a path in G connecting A_m and A_l. This proves that G is connected.

From the fact that G is connected, we deduce that it has a spanning tree G'. That is, there exists a subgraph G' of G that is a tree and includes all of the vertices of G. Let A_j be a leave of the spanning tree G'. It is now straightforward to show that A_j is the desired polytope from the statement.   ∎

Remark 3.4   Consider D and A_j as in the statement of Lemma 3.3. Then $D' := D \smallsetminus A_j$ satisfies the property (WP) with $(k-1)$ polytopes, and $D'' := D \cap A_j$ satisfies the property (WP) with one polytope.

Let A_j be the polytope as in Lemma 3.3. Note that there exists $i \neq j$ such that $\text{resint}(\partial A_i) \cap \text{resint}(\partial A_j) \cap D \neq \varnothing$; we denote this polytope by $\widetilde{A}_j$. Let $\widetilde{x} \in \text{resint}(\partial \widetilde{A}_j) \cap \text{resint}(\partial A_j) \cap D$. Note that, necessarily, $F := \text{resint}(\partial \widetilde{A}_j) \cap \text{resint}(\partial A_j)$ is contained in a hyperplane, and hence we can consider e a unit vector that is orthogonal to F. Take $r > 0$ such that $B(\widetilde{x}, r) \subseteq \text{int}((\widetilde{A}_j \cup A_j) \cap D)$. Let $z_1 := \widetilde{x} + re$ and let

$z_{-1} := \widetilde{x} - re$. Without the loss of generality we can assume that $z_1 \in \text{int}(\widetilde{A}_j)$. Denote by C_1 the set of points of the form $tz_1 + (1 - t)y$, where $t \in [0, 1]$ and $y \in B(\widetilde{x}, r) \cap F$. Similarly, denote by C_{-1} the set of points of the form $tz_{-1} + (1 - t)y$ where $t \in [0, 1]$ and where $y \in B(\widetilde{x}, r) \cap F$. Let $z_{-1/2} := \widetilde{x} - \frac{r}{2}e$ and consider the set $C_{-1/2}$ defined analogously to the way C_1 and C_{-1} are defined. We can think of C_1 and C_{-1} as *gates* connecting the sets $D' = D \smallsetminus A_j$ and $D'' = D \cap A_j$. We illustrate the construction in Figures 1 and 2.
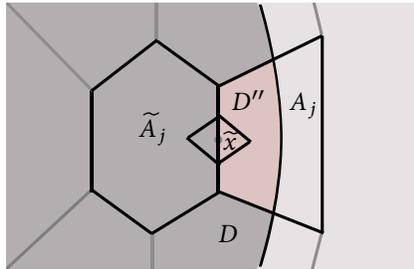
Figure 1. Polytope A_j with neighbor $\widetilde{A}_j$.
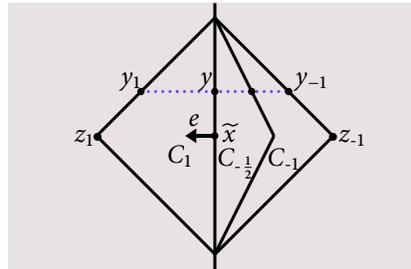
Figure 2. Gate, enlarged.

We claim that there is a function $\psi: D'' \cup C_1 \to D''$ that is a bi-Lipschitz homeomorphism. In fact, for a given point $y \in F \cap B(\widetilde{x}, r)$ consider the line with direction e passing through the point y. This line intersects ∂C_1 at the points y and y_1; it intersects ∂C_{-1} at the points y and y_{-1}, and finally it intersects $\partial C_{-1/2}$ at the points y and $y_{-1/2}$. We set $\psi(y_1) := y$, $\psi(y) := y_{-1/2}$ and $\psi(y_{-1}) := y_{-1}$. On the segments $[y_{-1}, y]$, $[y, y_1]$ we define ψ to be continuous and piecewise linear. In this way we define ψ for all points in $C_1 \cup C_{-1}$. Finally, set ψ to be the identity on $D'' \smallsetminus C_{-1}$. It is straightforward to check that ψ constructed in this way is a bi-Lipschitz homeomorphism.

Now we are ready to prove Theorem 1.2 for general domains.

Proof of Theorem 1.2

Step 1: Instead of proving the result for domains as in the statement, we first prove the result for domains D satisfying the (WP) property. The proof is by induction on the number of polytopes k.

We remark that the constant $D(d, \lambda)$ may change (increase) from line to line in the proof.

Base case. Suppose $k = 1$. In this case there exists $\psi: \overline{D} \to [0, 1]^d$ a bi-Lipschitz homeomorphism between $\overline{D}$ and the unit box. We use the map ψ to obtain measures $\widetilde{\nu}_1, \widetilde{\nu}_2$ on $(0, 1)^d$ by setting $\widetilde{\nu}_i := \psi_\sharp \nu_i$ for $i = 1, 2$. Using the fact that ψ is bi-Lipschitz, we can use the change of variables formula to deduce that $\widetilde{\nu}_1$ and $\widetilde{\nu}_2$ are absolutely continuous with respect to the Lebesgue measure with densities

$$\widetilde{\rho}_i(y) = \rho_i(\psi^{-1}(y))|\det(J\psi^{-1}(y))| \quad \text{for } i = 1, 2.$$

Here, $J\psi^{-1}$ represents the Jacobian matrix of ψ^{-1}.

Using the fact that ψ is bi-Lipschitz, we deduce that $1/\widetilde{\lambda} \le \widetilde{\rho}_1, \widetilde{\rho}_2 \le \widetilde{\lambda}$, where $\widetilde{\lambda} = \max\{\mathrm{Lip}(\psi)^d, \mathrm{Lip}(\psi^{-1})^d\}$. By Theorem 1.2 applied to the unit cube,

$$d_\infty(\widetilde{\nu}_1, \widetilde{\nu}_2) \le C(\widetilde{\lambda}, d)\|\widetilde{\rho}_1 - \widetilde{\rho}_2\|_{L^\infty((0,1)^d)}.$$

Consequently,

$$d_\infty(\nu_1, \nu_2) \le \mathrm{Lip}(\psi^{-1})d_\infty(\widetilde{\nu}_1, \widetilde{\nu}_2) \le C\|\widetilde{\rho}_1 - \widetilde{\rho}_2\|_{L^\infty((0,1)^d)} \le C\|\rho_1 - \rho_2\|_{L^\infty(D)}.$$

for some constant C depending on λ and D only.

Inductive Step. Suppose that for any domain in $\mathbb{R}^d$ satisfying the (WP) property with $(k-1)$ polytopes the proposition is true. Let D be a domain satisfying the (WP) property with k polytopes and let $\rho_1, \rho_2 \colon D \to (0, \infty)$ be functions as in the statement. By relabeling the functions if necessary, we can assume without loss of generality that $\int_{D'} \rho_1(x)dx - \int_{D'} \rho_2(x)dx \ge 0$, where D' is as in Remark 3.4. Since there is more mass in D' according to ν_1 than according to ν_2, we transfer this excess of mass from the set D' to the set D''. To achieve this, we first move the excess of mass on D' to the gate C_1 so that we can subsequently move it to the set D''. In mathematical terms, we consider an intermediate distribution $d\widetilde{\nu}_1 = \widetilde{\rho}_1 dx$, where

$$\widetilde{\rho}_1(x) := \begin{cases} \rho_2(x), & \text{if } x \in D' \smallsetminus C_1, \\ \beta\rho_1(x), & \text{if } x \in C_1, \\ \rho_1(x), & \text{if } x \in D'', \end{cases}$$

and where

$$\beta = \frac{\int_{D'}(\rho_1(x) - \rho_2(x))dx + \int_{C_1} \rho_2(x)dx}{\int_{C_1} \rho_1(x)dx}.$$

The idea is to compare ν_1 with $\widetilde{\nu}_1$ and then compare $\widetilde{\nu}_1$ with ν_2.

First, note that there is a $\lambda' > 1$ depending only on λ and D such that

$$\frac{1}{\lambda'} \le \rho_1, \widetilde{\rho}_1 \le \lambda'.$$

Since by construction $\nu_1(D') = \widetilde{\nu}_1(D')$, we use Remark 3.4 and the induction hypothesis to conclude that

$$d_\infty(\nu_{1 \llcorner D'}, \widetilde{\nu}_{1 \llcorner D'}) \le C(\lambda', D')\|\rho_1 - \widetilde{\rho}_1\|_{L^\infty(D')} = C(\lambda, D)\|\rho_1 - \widetilde{\rho}_1\|_{L^\infty(D')},$$

where $\nu_{1 \llcorner D'}$ denotes the measure ν_1 restricted to D' and $\widetilde{\nu}_1|_{D'}$ the measure $\widetilde{\nu}_1$ restricted to D'. Notice that we can write $C(\lambda', D') = C(\lambda, D)$, because λ' depends on λ and D only. An immediate consequence of the previous estimate is that

$$(3.1) \qquad\qquad d_\infty(\nu_1, \widetilde{\nu}_1) \le C(\lambda, D)\|\rho_1 - \widetilde{\rho}_1\|_{L^\infty(D)}.$$

Given the definition of β, it is straightforward to show that

$$\|\rho_1 - \widetilde{\rho}_1\|_{L^\infty(D)} \le C(\lambda, D)\|\rho_1 - \rho_2\|_{L^\infty(D)}$$

for some constant $C(\lambda, D)$ depending only on D and λ. The previous inequality combined with (3.1) gives

$$d_\infty(\nu_1, \widetilde{\nu}_1) \le C(\lambda, D)\|\rho_1 - \rho_2\|_{L^\infty(D)}.$$

Now we compare $\widetilde{\nu}_1$ with ν_2. First of all note that $\widetilde{\nu}_1(D_1'') = \nu_2(D_1'')$, where $D_1'' := D'' \cup C_1$. From the discussion proceeding Remark 3.4 we know that D_1'' is bi-Lipschitz

homeomorphic to the set D'', which in turn is bi-Lipschitz homeomorphic to the unit box. Thus, D_1'' is bi-Lipschitz homeomorphic to the unit box and hence proceeding as in the base case, we conclude that

$$d_\infty(\widetilde{v}_{1 \llcorner D_1''}, v_{2 \llcorner D_1''}) \le C(\lambda, D)\|\widetilde{\rho}_1 - \rho_2\|_{L^\infty(D_1'')}$$

and consequently

$$d_\infty(\widetilde{v}_1, v_2) \le C(\lambda, D)\|\widetilde{\rho}_1 - \rho_2\|_{L^\infty(D)}.$$

A straightforward computation shows that

$$\|\widetilde{\rho}_1 - \rho_2\|_{L^\infty(D)} \le C(\lambda, D)\|\rho_1 - \rho_2\|_{L^\infty(D)}$$

and thus

$$d_\infty(\widetilde{v}_1, v_2) \le C(\lambda, D)\|\rho_1 - \rho_2\|_{L^\infty(D)}.$$

Using the previous inequality, (3.1), and the triangle inequality we obtain the desired result.

Step 2: Now consider an open, connected, bounded domain D with Lipschitz boundary. By [3, Remark 5.3] there exists an open set $\widetilde{D}$ with smooth boundary that is bi-Lipschitz homeomorphic to D. In particular $\widetilde{D}$ is bounded and connected. By Propositions 3.2 and Step 1, the result holds for $\widetilde{D}$. Proceeding as in the base case in Step 1 and using the fact that D and $\widetilde{D}$ are bi-Lipschitz homeomorphic we obtain the desired result. ∎

We are now ready to prove Theorem 1.1.

Proof of Theorem 1.1 Let us consider the function $\phi: \mathbb{N} \to (0, \infty)$, which is given by

$$\phi(n) = \begin{cases} \dfrac{(\ln n)^{1/d}}{n^{1/d}}, & \text{if } d \ge 3, \\[2mm] \dfrac{(\ln n)^{3/4}}{n^{1/2}}, & \text{if } d = 2. \end{cases}$$

Step 1. We first prove the result for domains D satisfying the (WP) property. The proof is by induction on k, the number of polytopes used in the definition of the property (WP). In what follows C may change from line to line, but always represents a constant that depends only on λ and D. Furthermore, since the probability that a sample point belongs to a boundary of one of the k polytopes is zero, we assume without loss of generality that no sample point belongs to the boundary of any of the polytopes considered.

Base Case. Suppose that D is a domain satisfying the (WP) property with one polytope. Then $\overline{D}$ is bi-Lipschitz homeomorphic to the unit box. That is, there exists a bi-Lipschitz mapping $\psi: \overline{D} \to [0,1]^d$. Given a density $\rho: D \to (0, \infty)$ satisfying (1.2), we define measure $\widetilde{v}$ on $(0,1)^d$ to be the push-forward of v by ψ, *i.e.*, $\widetilde{v} := \psi_\sharp v$. Given the i.i.d. random points $X_1, \dots, X_n$ on D distributed according to v, we note that

$$\widetilde{X}_i = \psi(X_i) \quad \text{for } i = 1, \dots, n$$

are i.i.d. random points on $(0,1)^d$ distributed according to $\widetilde{v}$.

As in the proof of Theorem 1.2 we use the fact that ψ is bi-Lipschitz to deduce that $\widetilde{\nu}$ has a density $\widetilde{\rho}$ satisfying

$$\frac{1}{\widetilde{\lambda}} \le \widetilde{\rho} \le \widetilde{\lambda},$$

where $\widetilde{\lambda} = \lambda \max\{\mathrm{Lip}(\psi)^d, \mathrm{Lip}(\psi^{-1})^d\}$. From Theorem 1.1 applied to the unit cube, we know that for $\alpha > 2$, except on a set with probability $O(n^{-\alpha/2})$,

$$d_\infty(\widetilde{\nu}, \widetilde{\nu}_n) \le C\phi(n),$$

which implies

$$d_\infty(\nu, \nu_n) \le \mathrm{Lip}(\psi^{-1}) d_\infty(\widetilde{\nu}, \widetilde{\nu}_n) \le C\phi(n).$$

where C only depends on λ, D and α.

Inductive Step. Suppose that the theorem is true for any domain in $\mathbb{R}^d$ satisfying the (WP) property with $k-1$ polytopes. Let D be a domain satisfying the (WP) property with k polytopes and let $\rho \colon D \to (0, \infty)$ be a density function satisfying (1.2). Consider $\widetilde{\rho}_n \colon D \to D$ the density function given by

$$\widetilde{\rho}_n(x) = \begin{cases} \dfrac{\nu_n(D')}{\nu(D')}\rho(x), & \text{if } x \in D', \\ \dfrac{\nu_n(D'')}{\nu(D'')}\rho(x), & \text{if } x \in D'', \end{cases}$$

where D' and D'' are as in Remark 3.4. Let $\widetilde{\nu}_n$ be the measure $d\widetilde{\nu}_n = \widetilde{\rho}_n dx$ and note that $\nu_n(D') = \widetilde{\nu}_n(D')$ and $\nu_n(D'') = \widetilde{\nu}(D'')$. Also, notice that

$$(3.2) \qquad \|\rho - \widetilde{\rho}_n\|_{L^\infty(D)} \le C|\nu_n(D') - \nu(D')|$$

for some constant C that depends only on λ and D.

To give some probabilistic estimates on $|\nu_n(D') - \nu(D')|$, we use Chernoff's inequality (2.4) to conclude that

$$(3.3) \qquad P\Big(|\nu_n(D') - \nu(D')| > \sqrt{\frac{\alpha \ln n}{n}}\Big) \le 2n^{-2\alpha}.$$

Denote by Ω_n the event in which $|\nu_n(D') - \nu(D)| \le \sqrt{\frac{\alpha \ln n}{n}}$. By (3.2) and Theorem 1.2 (from its proof, it holds for well partitioned domains), given Ω_n we have

$$(3.4) \qquad d_\infty(\nu, \widetilde{\nu}_n) \le C\frac{(\ln n)^{1/2}}{n^{1/2}}.$$

We use the fact that $\nu_n(D') = \widetilde{\nu}_n(D')$ and $\nu_n(D'') = \widetilde{\nu}_n(D'')$ to estimate $d_\infty(\widetilde{\nu}_n, \nu_n)$. Indeed, by the induction hypothesis, given the event Ω_n, with probability at least $1 - cn^{-\alpha/2}$,

$$d_\infty(\widetilde{\nu}_{n \llcorner D'}, \nu_{n \llcorner D'}) \le C\phi(n) \ \text{ and } \ d_\infty(\widetilde{\nu}_{n \llcorner D''}, \nu_{n \llcorner D''}) \le C\phi(n).$$

In case the previous inequalities hold we conclude that

$$d_\infty(\widetilde{\nu}_n, \nu_n) \le \max\big\{ d_\infty(\widetilde{\nu}_{n \llcorner D'}, \nu_{n \llcorner D'}), d_\infty(\widetilde{\nu}_{n \llcorner D''}, \nu_{n \llcorner D''})\big\} \le C\phi(n).$$

Thus, given Ω_n, with probability at least $1 - cn^{-\alpha/2}$,

$$d_\infty(\widetilde{\nu}_n, \nu_n) \le C\phi(n).$$

From the previous discussion, (3.3), and (3.4) we conclude that with probability at least $1 - cn^{-\alpha/2}$,

$$d_\infty(v, v_n) \le C\phi(n) + C\frac{(\ln n)^{1/2}}{n^{1/2}} \le C\phi(n).$$

Step 2. To prove the theorem for an arbitrary open, connected, bounded domain D with Lipschitz boundary it is enough to notice that by [3, Remark 5.3] there exists an open set $\widetilde{D}$ with smooth boundary that is bi-Lipschitz homeomorphic to D. In particular, $\widetilde{D}$ is bounded and connected. By Proposition 3.2 the result holds for $\widetilde{D}$ by Step 1. Proceeding as in the base case in Step 1 and using the fact that D and $\widetilde{D}$ are bi-Lipschitz homeomorphic we obtain the desired result. ∎

Appendix A   Proof of Proposition 3.2

Consider D to be a bounded open set with smooth boundary. For $\varepsilon > 0$ we denote by $\partial_\varepsilon D$ the set of points $x \in \mathbb{R}^d$ with $d(x, \partial D) \le \varepsilon$. The fact that ∂D is a smooth compact manifold implies that there exists $0 < \varepsilon_0 < 1$ such that for every $x \in \partial_{\varepsilon_0} D$ there is a unique point $P(x)$ on ∂D closest to x. Furthermore the function $P: x \in \partial_{2\varepsilon_0} D \mapsto P(x)$ is smooth.

For a given $z \in \partial D$ we let $\vec{n}_z$ be the unit outer normal vector to ∂D at the point z. The fact that ∂D is a smooth manifold in $\mathbb{R}^d$ also implies that the outer unit normal vector field changes smoothly over ∂D.

We consider the signed distance function to ∂D, $g: \partial_{2\varepsilon_0} D \to \mathbb{R}$,

$$g(y) := \begin{cases} \operatorname{dist}(y, \partial D), & \text{if } y \in D^c, \\ -\operatorname{dist}(y, \partial D), & \text{if } y \in D. \end{cases}$$

This function is smooth, and its gradient is given by

$$(A.1) \qquad\qquad \nabla g(y) = \vec{n}_{P(y)}.$$

We remark that for every $y \in \partial_{\varepsilon_0} D$, $g(y) = |y - P(y)|$ if $y \notin D$ and $g(y) = -|y - P(y)|$ if $y \in D$.

For a fixed $0 < \varepsilon < \varepsilon_0$ consider the family of open balls $\{B(x, \varepsilon^2)\}_{x \in \partial D}$. This is an open cover of the set ∂D which is compact. Hence, there exists a finite subcover $\{B(x_1, \varepsilon^2), \ldots, B(x_N, \varepsilon^2)\}$ of ∂D. To fix some notation, we let $\vec{n}_i$ be the vector $\vec{n}_{x_i}$ and we let T_i be the tangent plane to ∂D at the point x_i. Let $V_1, \ldots V_N$ be the Voronoi cells induced by the points $x_1, \ldots, x_N$; that is , we let V_i be the set

$$V_i := \left\{ y \in \mathbb{R}^d : |x_i - y| \le |x_j - y|, \forall j \ne i \right\}.$$

Note that for every $t \in [-\varepsilon, \varepsilon]$ we have $P(x_i + t\vec{n}_i) = x_i$. In particular,

$$(A.2) \qquad\qquad |x_i + t\vec{n}_i - x_i| < |x_i + t\vec{n}_i - x_j|$$

for every $j \ne i$. Consider $\widetilde{x}_i$ to be the point $\widetilde{x}_i := -\frac{\varepsilon}{2}\vec{n}_i + x_i$ and let $T_i^+ := \varepsilon\vec{n}_i + T_i$, $T_i^- := \varepsilon\vec{n}_i + T_i$ be the planes parallel to T_i passing though the points $\varepsilon\vec{n}_i + x_i$ and $-\varepsilon\vec{n}_i + x_i$, respectively. We denote by S_i the closed strip delimited by the planes T_i^+ and T_i^- and let $A_i := V_i \cap S_i$. See Figure 3.

We first want to show that the region A_i is contained in a circular cylinder whose axis is the line passing through the point x_i with direction $\vec{n}_i$ and whose radius is small compared to ε. To achieve this, for a point $y \in \mathbb{R}^d$, denote by y_i the projection of y along the line passing through x_i with direction $\vec{n}_i$.

Claim 1: For all $0 < \varepsilon < \frac{\varepsilon_0}{2}$ small enough, $y \in A_i$ implies that $|y - y_i\| \le 4\varepsilon^{3/2}$.

To prove the claim suppose for the sake of contradiction that there is $y \in A_i$ with $|y - y_i| \ge 4\varepsilon^{3/2}$. Since $y \in S_i$, in particular $|y_i - x_i| = \mathrm{dist}(y_i, \partial D) \le \varepsilon$. Consider a point $\widetilde{y}$ in the segment $[y, y_i]$ such that $4\varepsilon^{3/2} \ge |\widetilde{y} - y_i| \ge 3\varepsilon^{3/2}$. Then $|\widetilde{y} - x_i| \le |\widetilde{y} - y_i| + |y_i - x_i| < 4\varepsilon^{3/2} + \varepsilon < 2\varepsilon$ if ε is small enough. Thus $\widetilde{y} - P(\widetilde{y})| < 2\varepsilon$. Note also that $y \in A_i$ and $y_i \in A_i$ (from (A.2)). Since the set A_i is convex, we conclude that $\widetilde{y} \in A_i$. To get to a contradiction we want to show that $|\widetilde{y} - x_k| < |\widetilde{y} - x_i|$ for some k; this would imply that $\widetilde{y} \notin V_i$, which indeed would be a contradiction given that $\widetilde{y} \in A_i$.

Note that $P(\widetilde{y}) \in B(x_k, \varepsilon^2)$ for some k. Thus,

$$(A.3) \qquad |\widetilde{y} - x_k|^2 \le \left(|\widetilde{y} - P(\widetilde{y})| + |P(\widetilde{y}) - x_k| \right)^2$$
$$= |\widetilde{y} - P(\widetilde{y})|^2 + 2|\widetilde{y} - P(\widetilde{y})| \cdot |P(\widetilde{y}) - x_k| + |P(\widetilde{y}) - x_k|^2$$
$$\le |\widetilde{y} - P(\widetilde{y})|^2 + 4\varepsilon^3 + \varepsilon^4.$$

Furthermore, note that

$$(A.4) \qquad |\widetilde{y} - x_i|^2 = |y_i - x_i|^2 + |\widetilde{y} - y_i|^2$$
$$= g(y_i)^2 + |\widetilde{y} - y_i|^2$$
$$= g(\widetilde{y})^2 + g(y_i)^2 - g(\widetilde{y})^2 + |\widetilde{y} - y_i|^2$$
$$\ge |\widetilde{y} - P(\widetilde{y})|^2 - |g(y_i)^2 - g(\widetilde{y})^2| + |\widetilde{y} - y_i|^2.$$

Since g is smooth in $\overline{\partial_{\varepsilon_0} D}$, there exists M such that $M \ge \|D^2 g(x)\|$ for all $x \in \overline{\partial_{\varepsilon_0} D}$. By (A.1), the gradient of the signed distance function g at the point y_i is equal to $\vec{n}_i$. Since $\widetilde{y} - y_i$ is orthogonal to $\vec{n}_i$, by Taylor expansion

$$|g(\widetilde{y}) - g(y_i)| = |g(\widetilde{y}) - g(y_i) - Dg(y_i) \cdot (\widetilde{y} - y_i)| \le M|\widetilde{y} - y_i|^2.$$

Thus,

$$|g(\widetilde{y})^2 - g(y_i)^2| = |g(\widetilde{y}) - g(y_i)| \cdot |g(\widetilde{y}) + g(y_i)| \le 3M\varepsilon|\widetilde{y} - y_i|^2.$$

Using (A.4) we deduce that

$$|\widetilde{y} - x_i|^2 \ge |\widetilde{y} - P(\widetilde{y})|^2 + (1 - 3M\varepsilon)|\widetilde{y} - y_i|^2,$$

Therefore, for small enough $\varepsilon > 0$

$$|\widetilde{y} - x_i|^2 \ge |\widetilde{y} - P(\widetilde{y})|^2 + 5\varepsilon^3.$$

Combining the previous inequality with (A.3) we deduce that $|\widetilde{y} - x_i| > |\widetilde{y} - x_k|$. This proves the claim.

Consider the circular cylinder whose axis is the line passing through the point x_i with direction $\vec{n}_i$ and whose radius is $4\varepsilon^{3/2}$. We let C_i^+ be the portion of the cylinder contained in S_i.

By (A.2) we can find a circular cylinder of smaller radius, whose axis is the same as that of C_i^+, but such that the portion of it contained in S_i, denoted by C_i^-, satisfies

$$C_i^- \subseteq A_i \subseteq C_i^+.$$

See Figure 3.
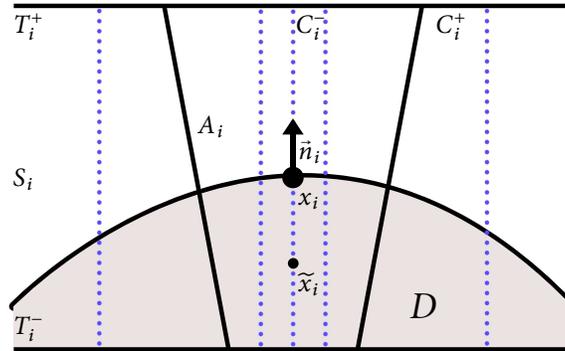

Figure 3

Claim 2. Let $0 < \varepsilon < \frac{\varepsilon_0}{2}$ be small enough. Then there exists a map $\Phi_i \colon A_i \cap \overline{D} \to A_i$ which is a bi-Lipschitz homeomorphism. In particular, since A_i is a closed convex body with nonempty interior, we conclude that $A_i \cap \overline{D}$ is bi-Lipschitz homeomorphic to the unit cube.

To prove the claim, fix $0 < \varepsilon$ so that the conclusions from Claim 1 hold. From the bound on the second derivative of g and since the radius of C_i^+ is $4\varepsilon^{3/2}$, we deduce that there exists a universal constant $L > 0$ such that

(A.5) $$|\vec{n}_z - \vec{n}_i| \le L\varepsilon^{3/2}, \quad \forall z \in \partial D \cap A_i,$$

due to the fact that $A_i \subseteq C_i^+$.

We now turn to constructing the bi-Lipschitz mapping between $\overline{D} \cap A_i$ and A_i. We do that by linear mappings along rays emanating from $\widetilde{x}_i$. Consider $\mathcal{S}^{d-1}$ the set of all unit vectors in $\mathbb{R}^d$. For $\vec{n} \in \mathcal{S}^{d-1}$ define $s_{\vec{n}}$ and $t_{\vec{n}}$ by

$$s_{\vec{n}} := \sup\left\{ s > 0 : \widetilde{x}_i + s\vec{n} \in \overline{D} \cap A_i \right\}, \quad t_{\vec{n}} := \sup\left\{ t > 0 : \widetilde{x}_i + t\vec{n} \in A_i \right\}.$$

Since $C_i^- \subseteq A_i \subseteq C_i^+$, we deduce that both functions $\vec{n} \in \mathcal{S}^{d-1} \mapsto s_{\vec{n}}$ and $\vec{n} \in \mathcal{S}^{d-1} \mapsto t_{\vec{n}}$ are bounded above and below by positive constants.

Now, note that for every $\vec{n} \in \mathcal{S}^{d-1}$, we have $s_{\vec{n}} \le t_{\vec{n}}$. Moreover, by (A.5) and the fact that $A_i \subseteq C_i^+$, we deduce that if $s_{\vec{n}} < t_{\vec{n}}$, then $|\vec{n}_i - \vec{n}| \le L\varepsilon^{3/2}$, where L is a universal constant, which is not necessarily the same as in (A.5). In particular, by choosing ε to be small enough we can assume that if $s_{\vec{n}} < t_{\vec{n}}$ then, the ray starting at $\widetilde{x}_i$ with direction $\vec{n}$ only intersects $\partial D \cap A_i$ at one point. This fact, together with the smoothness of the outer normal vector field implies that the map $\vec{n} \in \mathcal{S}^{d-1} \mapsto s_{\vec{n}}$ is Lipschitz. On the other hand, since the set A_i is a convex set with piecewise smooth

boundary (a convex polytope), we deduce that the function $\vec{n} \in \mathbb{S}^{d-1} \mapsto t_{\vec{n}}$ is Lipschitz as well.

Consider the map $\Phi_i \colon \overline{D} \cap A_i \to A_i$ defined as follows. Set $\Phi_i(\widetilde{x}_i) = \widetilde{x}_i$. For $x \in \overline{D} \cap A_i$, $x \neq \widetilde{x}_i$ we can write $x = \widetilde{x}_i + s\vec{n}$, for some $\vec{n} \in \mathbb{S}^{d-1}$ and for some $0 < s \le s_{\vec{n}}$; we let $\Phi_i(x)$ be

$$\Phi_i(x) := \widetilde{x}_i + \frac{st_{\vec{n}}}{s_{\vec{n}}}\vec{n}.$$

Since both functions $\vec{n} \in \mathbb{S}^{d-1} \mapsto s_{\vec{n}}$ and $\vec{n} \in \mathbb{S}^{d-1} \mapsto t_{\vec{n}}$ are bounded above and below by positive constants and are Lipschitz, we deduce that the map Φ_i is a bi-Lipschitz homeomorphism between $\overline{D} \cap A_i$ and A_i. This proves the claim.

Claim 3. For any $\varepsilon < 1$ it holds that $\partial D \cap (V_i \setminus S_i) = \varnothing$. To prove this claim, assume for the sake of contradiction that there exists $x \in \partial D \cap (V_i \setminus S_i)$. Since $x \notin S_i$, it follows that $|x - x_i| \ge \varepsilon$. On the other hand, given that $x \in \partial D$, we know there exists k such that $x \in B(x_k, \varepsilon^2)$. Since $\varepsilon < 1$, we deduce that $|x - x_k| < |x - x_i|$ and thus $x \notin V_i$. This is a contradiction.

Now we have all the ingredients needed to prove Proposition 3.2. Indeed, take $\varepsilon > 0$ small enough so that all of the conclusions of all the previous claims hold. From Claim 3, we deduce that every V_i can be partitioned into three convex polytopes; one that intersects ∂D, namely $A_i = V_i \cap S_i$ and other two polytopes, one which is contained in $\text{int}(D^c)$ and another one contained in D. We denote the later one by $\widehat{A}_i$. We consider the family $\{A_1, \widehat{A}_1, \ldots, A_N, \widehat{A}_N\}$ of convex polytopes. This family covers D and is such that properties (i) and (ii) from Definition 3.1 are satisfied. Moreover, given that $\widehat{A}_i \subseteq D$ and given that $\widehat{A}_i$ is convex, we deduce that $\widehat{A}_i$ satisfies property (iii) automatically, since all closed convex bodies with piecewise smooth boundary are bi-Lipschitz homeomorphic. Finally, Claim 2 implies that property (iii) holds for each of the A_i. All together this implies that D satisfies the (WP) property.

References

[1] M. Ajtai, J. Komlós, and G. Tusnády, *On optimal matchings*. Combinatorica **4**(1984), no. 4, 259–264. http://dx.doi.org/10.1007/BF02579135

[2] S. Arora, S. Rao, and U. Vazirani, *Expander flows, geometric embeddings and graph partitioning*. J. ACM (JACM) **56**(2009), no. 2, Art. 5, 1–37.

[3] J. M. Ball and A. Zarnescu, *Partial regularity and smooth topology-preserving approximations of rough domains*. arxiv:1312.5156

[4] S. Berstein, *On a modification of Chebyshev's inequality and of the error formula of Laplace*. Ann. Sci. Inst. Sav. Ukraine, Sect. Math. **1**(1924), no. 4, 38–49.

[5] E. Boissard, *Simple bounds for convergence of empirical and occupation measures in 1-Wasserstein distance*. Electron. J. Probab. **16**(2011), no. 83, 2296–2333. http://dx.doi.org/10.1214/EJP.v16-958

[6] F. Bolley, A. Guillin, and C. Villani, *Quantitative concentration inequalities for empirical measures on non-compact spaces*. Probab. Theory Related Fields **137**(2007), no. 3–4, 541–593. http://dx.doi.org/10.1007/s00440-006-0004-7

[7] T. Champion, Thierry, L. De Pascale, and P. Juutinen, *The ∞-Wasserstein distance: local solutions and existence of optimal transport maps*. SIAM J. Math. Anal. **40**(2008), no. 1, 1–20.
http://dx.doi.org/10.1137/07069938X

[8] H. Chernoff, *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations*. Ann. Math. Statistics **23**(1952), 493–507. http://dx.doi.org/10.1214/aoms/1177729330

[9] V. Dobrić and J. E. Yukich, *Asymptotics for transportation cost in high dimensions*. J. Theoret. Probab. **8**(1995), no. 1, 97–118. http://dx.doi.org/10.1007/BF02213456

[10] R. M. Dudley, *The speed of mean Glivenko-Cantelli convergence*. Ann. Math. Statist. **40**(1968), 40–50. http://dx.doi.org/10.1214/aoms/1177697802

[11] T. Leighton and P. Shor, *Tight bounds for minimax grid matching with applications to the average case analysis of algorithms*. Combinatorica **9**(1989), no. 2, 161–187.
http://dx.doi.org/10.1007/BF02124678

[12] P. W. Shor and J. E. Yukich, *Minimax grid matching and empirical measures*. Ann. Probab. **19**(1991), no. 3, 1338–1348. http://dx.doi.org/10.1214/aop/1176990347

[13] M. Talagrand, *The generic chaining*. Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2005.

[14] _____, *Upper and lower bounds of stochastic processes*. A Series of Modern Surveys in Mathematics, 60, Springer-Verlag, Berlin Heidelberg, 2014

[15] _____, *The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3*. Ann. Probab. **22**(1994), no. 2, 919–959. http://dx.doi.org/10.1214/aop/1176988735

[16] M. Talagrand and J. E. Yukich, *The integrability of the square exponential transportation cost*. Ann. Appl. Probab. **3**(1993), no. 4, 1100–1111. http://dx.doi.org/10.1214/aoap/1177005274

Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA, 15213, USA
e-mail: ngarciat@andrew.cmu.edu slepcev@math.cmu.edu