



HAL
open science

Optimal transport methods for combinatorial optimization over two random point sets

Michael Goldman, Dario Trevisan

► **To cite this version:**

Michael Goldman, Dario Trevisan. Optimal transport methods for combinatorial optimization over two random point sets. 2022. hal-03792981v1

HAL Id: hal-03792981

<https://hal.science/hal-03792981v1>

Preprint submitted on 30 Sep 2022 (v1), last revised 28 Nov 2023 (v2)

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

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

OPTIMAL TRANSPORT METHODS FOR COMBINATORIAL OPTIMIZATION OVER TWO RANDOM POINT SETS

MICHAEL GOLDMAN AND DARIO TREVISAN

ABSTRACT. We investigate the minimum cost of a wide class of combinatorial optimization problems over random bipartite geometric graphs in \mathbb{R}^d where the edge cost between two points is given by a p -th power of their Euclidean distance. This includes e.g. the travelling salesperson problem and the bounded degree minimum spanning tree. We establish in particular almost sure convergence, as n grows, of a suitable renormalization of the random minimum cost, if the points are uniformly distributed and $d \geq 3$, $1 \leq p < d$. Previous results were limited to the range $p < d/2$.

Our proofs are based on subadditivity methods and build upon new bounds for random instances of the Euclidean bipartite matching problem, obtained through its optimal transport relaxation and functional analytic techniques.

CONTENTS

1. Introduction	1
2. Notation and preliminary results	6
3. Combinatorial optimization problems over bipartite graphs	14
4. Convergence results for Poisson point processes	22
5. Proof of main result	32
6. Bounds for the Euclidean assignment problem	37
References	48

1. INTRODUCTION

Combinatorial optimization problems on graphs are widespread in operation research, with applications in planning and logistics. Their study is strongly related to algorithm theory and computational complexity theory. The most representative example of such discrete variational problems is the travelling salesperson problem (TSP) [41]: given a set of cities and distances between each pair of them, one asks for the shortest route that visits each city exactly once and returns to the origin city (i.e. a tour). Like many related combinatorial problems and despite its straightforward formulation, the TSP belongs to the class of NP-hard problems. In practical terms, computing an exact solution becomes computationally intractable as known algorithms perform exponentially many steps in the number of cities.

In real-world situations, there is quite often the need to solve many similar instances of a given combinatorial optimization problem. In that case, additional structure, including geometry and randomness, can be exploited. The Euclidean formulation of the TSP, i.e., when cities are points in \mathbb{R}^d and distances are given by the Euclidean distance, is still NP-hard [35], but Karp [31] observed that solutions to random instances, i.e., when cities are sampled independently and uniformly, can be efficiently approximated via a partitioning

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

Key words and phrases. Travelling Salesperson Problem, matching problem, optimal transport, geometric probability.

D.T. was partially supported by the INdAM-GNAMPA project 2022 “Temi di Analisi Armonica Subellittica”.

scheme. His proof relies upon the seminal work by Beardwood, Halton and Hammersely [9], where precise asymptotics for optimal costs of a random instance of the problem were first established: given i.i.d. points $(X_i)_{i=1}^n$ distributed according to a probability density ρ on \mathbb{R}^d , denoting the length $\mathcal{C}_{\text{TSP}}((X_i)_{i=1}^n)$ of the (random) solution to the TSP cycling through such points satisfies the \mathbb{P} -a.s. limit

$$\lim_{n \rightarrow \infty} n^{\frac{1}{d}-1} \mathcal{C}_{\text{TSP}}((X_i)_{i=1}^n) = \beta_{\text{BHH}} \int_{\mathbb{R}^d} \rho^{1-\frac{1}{d}}, \quad (1.1)$$

where $\beta_{\text{BHH}} = \beta_{\text{BHH}}(d) \in (0, \infty)$ is a constant depending on the dimension d only. The scaling $n^{1-1/d}$ is intuitively explained by the fact that the n cities are connected through paths of typical length $n^{-1/d}$ (as if they were on a regular grid).

Building upon these ideas, several authors [36, 42, 43, 48] contributed towards establishing a general theory to obtain limit results of BHH-type, i.e., as in (1.1), for a wide class of random Euclidean combinatorial optimization problems. The theory allows also for more general weights than the Euclidean length, including p -th powers of the Euclidean distance, a variant often motivated by modelling needs. If $0 < p < d$, with a minimal modification of the techniques one obtains BHH-type results as in (1.1), with the scaling replaced by $n^{1-p/d}$, the constant β_{BHH} now depending on p , d and the specific combinatorial optimization problem, and the integrand $\rho^{1-1/d}$ replaced by $\rho^{1-p/d}$. For $p \geq d$, the situation becomes subtler and (1.1) is known for the TSP only if $p = d$, see [49] and [48, Section 4.3].

Despite the wide applicability of this theory, several classical problems such as those formulated over two random sets of points, are not covered and require different mathematical tools. The Euclidean assignment problem, also called bipartite matching, is certainly the most representative among these: given two sets of n points $(x_i)_{i=1}^n, (y_j)_{j=1}^n \subseteq \mathbb{R}^d$, one defines the matching cost functional as

$$\text{MP}((x_i)_{i=1}^n, (y_j)_{j=1}^n) = \min_{\sigma} \sum_{i=1}^n |x_i - y_{\sigma(i)}|^p,$$

where the minimum is taken among all the permutations σ over n elements. This is often interpreted in terms of optimal planning for the execution of a set of jobs at positions y_j 's to be assigned to a set of workers at the positions x_i 's. Although the assignment problem belongs to the P complexity class, i.e., an optimal σ can be found in a polynomial number of steps (with respect to n) the analysis of random instances shows some interesting behavior in low dimensions. Indeed, if $(X_i)_{i=1}^n, (Y_j)_{j=1}^n$ are i.i.d. and uniformly distributed on the cube $(0, 1)^d$, it is known [20, 3, 44, 21] that ¹

$$\mathbb{E} [\text{M}^1((X_i)_{i=1}^n, (Y_j)_{j=1}^n)] \sim \begin{cases} \sqrt{n} & \text{for } d = 1 \\ \sqrt{n \log n} & \text{for } d = 2 \\ n^{1-\frac{1}{d}} & \text{for } d \geq 3. \end{cases}$$

In particular, for $d \in \{1, 2\}$ the cost is asymptotically larger than the heuristically motivated $n^{1-1/d}$. This exceptional scaling is intuitively due to local fluctuations of the distributions of the two families of points.

Inspired by the combinatorial approach in [11] for the random Euclidean bipartite matching problem in dimension $d \geq 3$, Barthe and Bordenave [6] first proposed a general theory to establish results of BHH-type (1.1) for a wide class of random Euclidean combinatorial optimization problems over two sets of n points. Let us point out that the equality in (1.1) is actually only proven for uniform measures while in general only upper

¹The notation $A \lesssim B$ means that there exists a constant $C > 0$, such that $A \leq CB$, where C depends on the dimension d , p and possibly other quantities tacitly considered as fixed, e.g. a domain $\Omega \subseteq \mathbb{R}^d$ or a probability density ρ . We use the notation \lesssim_q to indicate the dependence on the parameter q . We write $A \sim B$ if both $A \lesssim B$ and $B \lesssim A$.

and lower bounds (which are conjectured to coincide) are known. In case of p -th power weighted distances, the theory developed in [6] applies in the range $0 < p < d/2$, which appears quite naturally in their arguments. The difficulty to go beyond the threshold $p = d/2$ is that (1.1) cannot hold without additional hypothesis on the density ρ . For example, because of fluctuations a necessary condition is connectedness of the support of ρ . Nevertheless, in the case of the Euclidean bipartite matching problem, it was recently proved [24] that if ρ is the uniform measure on the unit cube with $d \geq 3$ and $p \geq 1$, then

$$\lim_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathbf{M}^p((X_i)_{i=1}^n, (Y_j)_{j=1}^n)] = \beta_{\mathbf{M}}. \quad (1.2)$$

Here $\beta_{\mathbf{M}} \in (0, \infty)$ depends on d and p only. The proof is a combination of classical subadditivity arguments – that originate from [9] – and tools from the theory of optimal transport. In particular, the defect in subadditivity is estimated using the connection between Wasserstein distances and negative Sobolev norms. In this context, the use of this type of estimates can be traced back to a recent PDE ansatz proposed in statistical physics [14]. Since then, it has been successfully used in the mathematical literature [5, 33, 27, 23, 10, 34, 22, 25, 16], even beyond the case of i.i.d. points [47, 29, 26, 12]. We refer to [17, 7, 8] for further statistical physics literature. In fact, the technique in [24] is quite robust and coarser estimates can be used, avoiding the use of PDEs. Still, the results apply only for the Euclidean bipartite matching problem thanks to its connection with optimal transport. The main purpose of this paper is to show that for a quite general class of bipartite combinatorial problems it is actually possible to rely on the good bounds for the matching problem to obtain the analog of (1.1) provided $p < d$. This is inspired by [13] where a similar idea is used for the TSP and the 2-factor problem when $p = d = 2$.

As alluded to, an important open question left from the theory developed in [6] (see also [19]) is the existence of a limit in (1.1) for general densities. The only result in this direction is [2], which established for $p = d = 2$ that the limit of the expected cost (suitably renormalized) exists if Ω is a bounded connected open set, with Lipschitz boundary and ρ is Hölder continuous and uniformly strictly positive and bounded from above on Ω . This settled a conjecture from [8] and, more importantly for our purposes, combined subadditivity and PDE arguments with a Whitney-type decomposition to take into account the structure of Ω and its boundary. While we do not address this question here, some of the ideas from [2] are further developed in this work.

1.1. Main result. Our aim is to establish limit results for the cost of a wide class of Euclidean combinatorial optimization problems of two random point sets, in the range $d/2 \leq p < d$ for any dimension $d \geq 3$. This overcomes the limitations of [6], showing that in higher dimensions bipartite problems behave much more similarly to non-bipartite ones. Our general theorem can be stated as follows (a precise description of all the assumptions and notation is given in Section 2).

Theorem 1.1. *Let $d \geq 3$, $p \in [1, d)$ and let $\mathbf{P} = (\mathcal{F}_{n,n})_{n \in \mathbb{N}}$ be a combinatorial optimization problem over complete bipartite graphs such that assumptions **A1**, **A2**, **A3**, **A4** and **A5** hold and write $\mathcal{C}_{\mathbf{P}}^p((x_i)_{i=1}^n, (y_j)_{j=1}^n)$ for the optimal cost of the problem over the two sets of n points $(x_i)_{i=1}^n, (y_j)_{j=1}^n \subseteq \mathbb{R}^d$, with respect to the Euclidean distance raised to the power p . Then, there exists $\beta_{\mathbf{P}} \in (0, \infty)$ depending on p , d and \mathbf{P} only such that the following hold.*

Let $\Omega \subseteq \mathbb{R}^d$ be a bounded open set and assume that it is either convex or has C^2 boundary. Let ρ be a Hölder continuous probability density on Ω , uniformly strictly positive and bounded from above. Given i.i.d. random variables $(X_i)_{i=1}^{\infty}, (Y_j)_{j=1}^{\infty}$ with common law ρ we have \mathbb{P} -a.s. that

$$\limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathcal{C}_{\mathbf{P}}^p((X_i)_{i=1}^n, (Y_j)_{j=1}^n) \leq \beta_{\mathbf{P}} \int_{\Omega} \rho^{1-\frac{p}{d}}. \quad (1.3)$$

Moreover, if ρ is the uniform density and Ω is either a cube or has C^2 boundary, then the above is a \mathbb{P} -a.s. limit and equality holds.

Our assumptions **A1**, **A2**, **A3**, **A4** and in particular **A5** are slightly stronger than those introduced in [6, Section 5.3], but it is not difficult to show that all the specific examples discussed in [6] satisfy them. In particular, our result apply to the TSP, the minimum weight connected k -factor problem and the k -bounded degree minimum spanning tree. It is thus fair to say that for compactly supported densities, Theorem 1.1 extends the main results in [6].

Remark 1.2. Let us point out that (1.3) also holds in expectation (see Proposition 5.1).

Remark 1.3. Arguing as in [6] (see also [2]) and considering a “boundary” variant of \mathbb{P} it should be possible to adapt the proof of Theorem 1.1 to show that there exists $\beta_{\mathbb{P}}^b > 0$ such that

$$\beta_{\mathbb{P}}^b \int_{\Omega} \rho^{1-\frac{p}{d}} \leq \liminf_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathcal{C}_{\mathbb{P}}^p((X_i)_{i=1}^n, (Y_j)_{j=1}^n).$$

However since we are currently not able to prove that $\beta_{\mathbb{P}}^b = \beta_{\mathbb{P}}$ we decided to leave it aside.

Remark 1.4. In fact our result applies, at least in expectation, to any p -homogeneous bi-partite functional \mathcal{C} satisfying the subadditivity inequality (5.2) (which is similar to the condition (\mathcal{S}_p) from [6]) and the growth condition (5.3) (somewhat reminiscent of condition (\mathcal{R}_p) from [6]). See Remark 5.2.

Of course, our result applies in particular for the Euclidean assignment problem.

Corollary 1.5. For $d \geq 3$, $p \in [1, d)$, let $\Omega \subseteq \mathbb{R}^d$ be a cube or a bounded connected open set with C^2 boundary and let ρ be a Hölder continuous probability density on Ω , uniformly strictly positive and bounded from above. Then, given i.i.d. $(X_i)_{i=1}^{\infty}, (Y_j)_{j=1}^{\infty}$ with common law ρ , we have \mathbb{P} -a.s. that

$$\limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathcal{M}^p((X_i)_{i=1}^n, (Y_j)_{j=1}^n) \leq \beta_{\mathbb{M}} \int_{\Omega} \rho^{1-\frac{p}{d}},$$

with $\beta_{\mathbb{M}}$ as in (1.2). Moreover, if ρ is the uniform density and Ω has C^2 boundary, then the above is a \mathbb{P} -a.s. limit and equality holds.

Remark 1.6. In the case of the matching problem, combining ideas from this paper and [24] the conclusion of Corollary 1.5 could be extended to every $p \geq 1$ (at least in expectation).

1.2. Comments on the proof technique. Our proof leverages on the techniques developed for the bipartite matching problem, in particular [24, 2] to carefully estimate the defects in a geometric subadditivity argument. Comparing the approach in [6], which works if $p < d/2$, with that in [24], which holds instead for any p , a crucial difference is that the errors due to local oscillations in the two distributions of points are mitigated in the latter by spreading them evenly across all the points. This is possible since the optimal transport relaxation allows for general couplings as well as continuous densities, rather than discrete matchings only.

The overall strategy is thus to find a suitable replacement for such operation in the purely combinatorial setting. The starting point is Proposition 3.7 where we prove a subadditivity inequality. The problem is then to estimate the defect in subadditivity. This is achieved by combining the following three key observations.

The first one is to bound from above the cost of the problem over any two point sets $(x_i)_{i=1}^n, (y_j)_{j=1}^n$ by the sum of a term of order $n^{1-p/d}$ plus the bipartite matching cost between the two point sets. This is stated as an assumption (**A5**), but can be easily checked on many specific problems (Lemma 3.10): being an upper bound, it usually suffices to combine an optimal matching with the solution to an additional non-bipartite

combinatorial optimization problem, such as the TSP, to build a feasible solution. This approach was first successfully used in [13] (see also [2]) for the random bipartite TSP in the case $p = d = 2$, where one can simply argue that the main contribution comes from the logarithmic corrections in the matching cost.

The second key observation is that for point sets mostly made of i.i.d. points (while much less is assumed on the remaining ones), it is still possible to obtain good bounds for the matching cost. We refer to Section 6 for the precise statements, but the underlying idea is strongly related to bounds for the optimal transport cost in terms of the negative Sobolev norms – thus relying again on the PDE ansatz originally introduced in the statistical physics literature.

The third observation is that, in order to ensure that a small fraction of i.i.d. uniformly distributed points can indeed be found in the subadditivity defect terms, it is enough to keep them out of the optimization procedure on the smaller scales. As usual with those arguments, the proof of existence of the limit is performed first on the Poisson version of the random problem, so to retain a fraction of points we perform a thinning procedure.

Besides these main ideas, plenty of technical modifications with respect to the arguments in [6] and [24, 2] are required, e.g. in order to establish improved subadditivity inequalities (Proposition 3.7) and to extend the Whitney-type decomposition argument from [2] to $p \neq 2$.

1.3. Further questions and conjectures. Our results raise several questions about costs and properties of solutions to Euclidean random combinatorial optimization problems over two point sets. We list here a few which we believe are worth exploring.

1. Existence of a limit in (1.3) for non-uniform densities is rather easy to conjecture, but so far our techniques do not improve upon [6], hence the problem remains largely open.
2. Our techniques break down if $p \geq d$, but it is natural to conjecture that Theorem 1.1 should hold also in that range. In fact, the correct rate $n^{1-p/d}$ could follow directly from (A5) combined with the corresponding result for the matching problem.
3. In this work we considered only the case of compactly supported densities ρ . It would be interesting to investigate the case where the support is \mathbb{R}^d . To the best of our knowledge, the only results available so far in this direction are [33, 34] where the correct rates are established for the Gaussian density in the case of the matching problem.
4. The assumptions in [6] are slightly different than ours, although the specific problems considered therein satisfy both. It would be interesting to find examples which satisfy only one set of these, or possibly simplify even more our assumptions.
5. Many problems, such as the bounded degree minimum spanning tree, but also the bipartite matching problem itself, can be naturally formulated also for two families of points with different number of elements: it could be of interest to investigate limit results also in those cases.
6. The cases $d \in \{1, 2\}$ are necessarily excluded by our analysis, since subadditivity arguments do not apply already for the random bipartite matching problem. It is however already an open question, whether the additional logarithmic correction indeed appears in the asymptotic rates for many other problems. As an example, we mention that for the Euclidean minimum spanning tree over two random point sets (without any uniform bound on the degree) no logarithmic corrections appear [18], but the maximum degree is unbounded, hence it is not covered by our results.
7. In the deterministic literature, for the TSP and other NP-hard Euclidean combinatorial optimization problems, polynomial time approximation schemes are known [4] for any (fixed) dimension d , as the number of points grows. Can our approach

lead to similar schemes for problems on two families of points, possibly under some mild regularity assumption on their spatial distributions?

1.4. Structure of the paper. In Section 2 we first introduce some general notation. We then discuss Whitney-type decompositions, Sobolev spaces as well as recall useful known facts on the Optimal Transport problem, and possibly some novel ones (Proposition 2.9). We close the section with a variant of the standard subadditivity (Fekete-type) arguments, suited for our purposes together with some simple concentration inequalities. Section 3 is devoted to the combinatorial optimization problems we consider, discussing in particular the main assumptions that we require and some useful consequences. In Section 4 we establish a variant our main result in the case of Poisson point processes and in Section 5 we use it to deduce Theorem 1.1. These two sections in fact rely upon the novel bounds for the Euclidean assignment problem that we finally establish in Section 6.

2. NOTATION AND PRELIMINARY RESULTS

2.1. General notation. Given $n \in \mathbb{N}$, we write $[n] = \{1, \dots, n\}$ and $[n]_1 = \{(1, i)\}_{i=1}^n$, $[n]_2 = \{(2, i)\}_{i=1}^n$, which easily allows to define two disjoint copies of $[n]$. Given a finite set A , we write $|A|$ for the number of its elements, while, if $A \subseteq \mathbb{R}^d$ is infinite, $|A|$ denotes its Lebesgue measure.

Given a metric space (Ω, d) , $x \in \Omega$, $A \subseteq \Omega$, we write $d(x, A) = \min_{y \in A} \{d(x, y)\}$ and $\text{diam}(A) = \sup_{x, y \in A} d(x, y)$. We endow every set $\Omega \subseteq \mathbb{R}^d$ with the Euclidean distance. A partition $\{\Omega_k\}_{k=1}^K$ of a set Ω is always intended up to a set of Lebesgue measure zero. A rectangle $R \subseteq \mathbb{R}^d$ is a subset of the form $R = \prod_{i=1}^d (x_i, x_i + L_i)$, and is said to be of moderate aspect ratio if for every i, j , $L_i/L_j \leq 2$. If $L_i = L$ for every i , then $R = Q$ is a cube of side length L . We write $Q_L = (0, L)^d$. We write I_Ω for the indicator function of a set Ω .

2.2. Families of points. Given a set Ω , we consider finite ordered families of points $\mathbf{x} = (x_i)_{i=1}^n \subseteq \Omega$, with $n \in \mathbb{N}$, letting $\mathbf{x} = \emptyset$ if $n = 0$. For many purposes the order will not be relevant, but we thus may allow e.g. for repetitions (which will be probabilistically negligible anyway). Given a family $\mathbf{x} \subseteq \mathbb{R}^d$, we write $\mu^{\mathbf{x}} = \sum_{i=1}^n \delta_{x_i}$ for the associated empirical measure and, for every (Borel) $\Omega \subseteq \mathbb{R}^d$, we let $\mathbf{x}(\Omega) = \mu^{\mathbf{x}}(\Omega)$. In the special case $\Omega = \mathbb{R}^d$, we simply write $|\mathbf{x}| = \mathbf{x}(\mathbb{R}^d) = \mu^{\mathbf{x}}(\mathbb{R}^d)$ for the total number of points (counted with multiplicity). We also write \mathbf{x}_Ω for its restriction to Ω , i.e., the family of all points $x_i \in \Omega$, so that $\mathbf{x} = \mathbf{x}_\Omega$ if $\mathbf{x} \subseteq \Omega$ (conventionally, we naturally re-index it over $i = 1, \dots, \mathbf{x}(\Omega)$ with the order inherited from that in \mathbf{x}). Given $\mathbf{x} = (x_i)_{i=1}^n$, $\mathbf{y} = (y_j)_{j=1}^m \subseteq \mathbb{R}^d$, their union is $\mathbf{x} \cup \mathbf{y} = (x_1, \dots, x_n, y_1, \dots, y_m)$. Strictly speaking, the union should be called concatenation, since the operation is not commutative, in general.

2.3. Whitney partitions. We recall following partitioning result [2, Lemma 5.1].

Lemma 2.1. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary and let $\mathcal{Q} = \{Q_i\}_i$ be a Whitney partition of Ω . Then, for every $\delta > 0$ sufficiently small, letting $\mathcal{Q}_\delta = \{Q_i : \text{diam}(Q_i) \geq \delta\}$, there exists a finite family $\mathcal{R}_\delta = \{\Omega_j\}_j$ of disjoint open sets such that:*

- (i) $(\Omega_k)_{k=1}^K = \mathcal{Q}_\delta \cup \mathcal{R}_\delta$ is a partition of Ω ,
- (ii) $|\Omega_k| \sim \text{diam}(\Omega_k)^d$ for every $k = 1, \dots, K$,
- (iii) if $\Omega_k \in \mathcal{Q}_\delta$, then $\text{diam}(\Omega_k) \sim d(x, \Omega^c)$ for every $x \in \Omega_k$,
- (iv) if $\Omega_k \in \mathcal{R}_\delta$, then $\text{diam}(\Omega_k) \sim \delta$ and $d(x, \Omega^c) \lesssim \delta$, for every $x \in \Omega_k$.

Here all the implicit constants depend only on the initial partition \mathcal{Q} (and not on δ).

For later use, we collect some useful bounds related to these partitions.

Lemma 2.2. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary and let $\mathcal{Q} = \{Q_i\}_i$ be a Whitney partition of Ω . Then, for every $\delta > 0$ sufficiently small, letting $(\Omega_k)_{k=1}^K = \mathcal{Q}_\delta \cup \mathcal{R}_\delta$ as in Lemma 2.1, one has that $|\mathcal{R}_\delta| \lesssim \delta^{1-d}$ and the following holds:*

(1) For every $\alpha \in \mathbb{R}$,

$$\sum_{k=1}^K \text{diam}(\Omega_k)^\alpha \lesssim_\alpha \begin{cases} 1 & \text{if } \alpha > d-1, \\ |\log \delta| & \text{if } \alpha = d-1, \\ \delta^{1-d-\alpha} & \text{if } \alpha < d-1. \end{cases} \quad (2.1)$$

(2) For every $k = 1, \dots, K$, and $x \in \Omega_k$,

$$\sum_{j: \text{d}(x, \Omega_j) \leq \text{diam}(\Omega_k)} |\Omega_j|^{-\frac{1}{2}} \lesssim \delta^{1-\frac{d}{2}} \text{diam}(\Omega_k)^{d-1}, \quad (2.2)$$

and

$$\sum_{j: \text{d}(x, \Omega_j) > \text{diam}(\Omega_k)} \frac{|\Omega_j|^{\frac{1}{2}}}{\text{d}(x, \Omega_j)^{d-1}} \lesssim \delta^{1-\frac{d}{2}} |\log(\text{diam}(\Omega_k))|, \quad (2.3)$$

In all the inequalities the implicit constants depend upon \mathcal{Q} (and α in (2.1)) only.

By property (ii), inequality (2.1) also holds for the sum $\sum_{k=1}^K |\Omega_k|^\alpha$, with αd instead of α .

Proof. Since $\partial\Omega$ is Lipschitz, it follows from properties (i), (ii) and (iv) that, for every $x \in \Omega$ and $r \geq s \geq \delta$,

$$|\{k : \Omega_k \subseteq B(x, r), \text{diam}(\Omega_k) \in [s, 2s]\}| \lesssim (r/s)^{d-1}, \quad (2.4)$$

with the implicit constant depending on \mathcal{Q} only. It follows that $|\mathcal{R}_\delta| \lesssim \delta^{1-d}$ and, for every $\ell \leq |\log_2 \delta|$, the number of cubes $\Omega_k \in \mathcal{Q}_\delta$ with $\text{diam}(\Omega_k) \in [2^{-\ell}, 2^{-\ell+1})$ is estimated by $2^{\ell(d-1)}$. Therefore, for $\alpha \in \mathbb{R}$,

$$\begin{aligned} \sum_{k=1}^K \text{diam}(\Omega_k)^\alpha &\lesssim \sum_{\Omega_k \in \mathcal{Q}_\delta} \text{diam}(\Omega_k)^\alpha + \sum_{\Omega_k \in \mathcal{R}_\delta} \text{diam}(\Omega_k)^\alpha \\ &\lesssim \sum_{\ell \leq |\log_2 \delta|} \left| \left\{ \Omega_k \in \mathcal{Q}_\delta : \text{diam}(\Omega_k) \in [2^{-\ell}, 2^{-\ell+1}) \right\} \right| 2^{-\ell\alpha} + |\mathcal{R}_\delta| \cdot \delta^\alpha \\ &\lesssim \sum_{\ell \leq |\log_2 \delta|} 2^{\ell(d-1)} \cdot 2^{-\ell\alpha} + \delta^{1-d} \cdot \delta^\alpha. \end{aligned}$$

Since ℓ is also bounded from below in the summation (e.g. by $-|\log_2 \text{diam}(\Omega)|$), we obtain (2.1).

In order to prove (2.2) and (2.3) we further notice that, given Ω_k, Ω_j and $x \in \Omega_k$, we have that, for some constant $C = C(\mathcal{Q})$,

$$\Omega_j \subseteq B(x, C \max\{\text{d}(x, \Omega_j), \text{diam}(\Omega_k)\}). \quad (2.5)$$

Indeed, if $\Omega_j \in \mathcal{R}_\delta$, then $\text{diam}(\Omega_j) \lesssim \delta \lesssim \text{diam}(\Omega_k)$, hence (2.5) holds. If instead $\Omega_j \in \mathcal{Q}_\delta$, then we can find $y \in \Omega_j$ with $|x - y| \leq 2\text{d}(x, \Omega_j)$, so that, by the triangle inequality,

$$\text{d}(y, \Omega^c) \leq |x - y| + \text{d}(x, \Omega^c) \lesssim \max\{\text{d}(x, \Omega_j), \text{diam}(\Omega_k)\}$$

and by property (iii) we obtain that $\text{diam}(\Omega_j) \lesssim \max\{\text{d}(x, \Omega_j), \text{diam}(\Omega_k)\}$, yielding again the desired inclusion.

We now prove (2.2). Let $\ell_k \leq |\log_2 \delta|$ be such that $\text{diam}(\Omega_k) \in [2^{-\ell_k}, 2^{-\ell_k+1})$. Combining (2.5) and (2.4), we see that, for every $\ell \leq |\log_2 \delta|$, there are at most $2^{(\ell-\ell_k)(d-1)}$ sets

Ω_j such that $d(x, \Omega_j) \leq \text{diam}(\Omega_k)$ and $\text{diam}(\Omega_j) \in [2^{-\ell}, 2^{-\ell+1})$. Therefore,

$$\begin{aligned} \sum_{j: d(x, \Omega_j) \leq \text{diam}(\Omega_k)} |\Omega_j|^{-\frac{1}{2}} &\lesssim \sum_{\ell \leq |\log_2 \delta|} 2^{-\ell \frac{d}{2}} 2^{(\ell - \ell_k)(d-1)} \\ &\lesssim 2^{-\ell_k(d-1)} \sum_{\ell \leq |\log_2 \delta|} 2^{\ell(\frac{d}{2}-1)} \lesssim \delta^{\frac{d}{2}-1} \text{diam}(\Omega_k)^{d-1}, \end{aligned}$$

recalling that ℓ is bounded from below by a constant depending on \mathcal{Q} only. This proves (2.2).

To prove (2.3), we split dyadically,

$$\begin{aligned} \sum_{j: d(x, \Omega_j) > \text{diam}(\Omega_k)} \frac{|\Omega_j|^{\frac{1}{2}}}{d(x, \Omega_j)^{d-1}} &\lesssim \sum_{\ell \leq \ell_k} \frac{1}{(2^{-\ell})^{d-1}} \sum_{j: d(x, \Omega_j) \in [2^{-\ell}, 2^{-\ell+1})} |\Omega_j|^{\frac{1}{2}} \\ &\stackrel{(2.5)}{\lesssim} \sum_{\ell \leq \ell_k} 2^{\ell(d-1)} \sum_{\Omega_j \subset B(x, C2^{-\ell})} |\Omega_j|^{\frac{1}{2}}. \end{aligned} \quad (2.6)$$

Let us also notice that, if $\Omega_j \subseteq B(x, C2^{-\ell})$, then necessarily $\delta \leq \text{diam}(\Omega_j) \lesssim 2^{-\ell}$ (since $\text{diam}(\Omega_j)^d \sim |\Omega_j|$). Thus for ℓ' with $2^{-\ell'} \sim 2^{-\ell}$,

$$\begin{aligned} \sum_{\Omega_j \subset B(x, C2^{-\ell})} |\Omega_j|^{\frac{1}{2}} &\lesssim \sum_{\ell' \leq u \leq |\log_2 \delta|} 2^{-u \frac{d}{2}} \left| \left\{ \Omega_j \subseteq B(x, C2^{-\ell}) : \text{diam}(\Omega_j) \in [2^{-u}, 2^{-u+1}) \right\} \right| \\ &\stackrel{(2.4)}{\lesssim} \sum_{\ell' \leq u \leq |\log_2 \delta|} 2^{-u \frac{d}{2}} \cdot 2^{(u-\ell)(d-1)} = 2^{-\ell(d-1)} \sum_{\ell' \leq u \leq |\log_2 \delta|} 2^{u(\frac{d}{2}-1)} \\ &\lesssim 2^{-\ell(d-1)} \delta^{1-\frac{d}{2}}. \end{aligned}$$

In the last line we used once again that ℓ' is bounded from below. Using this bound in (2.6), we conclude that

$$\sum_{j: d(x, \Omega_j) > \text{diam}(\Omega_k)} \frac{|\Omega_j|^{\frac{1}{2}}}{d(x, \Omega_j)^{d-1}} \leq \sum_{\ell \leq \ell_k} 2^{\ell(d-1)} \cdot 2^{-\ell(d-1)} \delta^{1-\frac{d}{2}} \lesssim \delta^{1-\frac{d}{2}} |\log(\text{diam}(\Omega_k))|,$$

where in the last line we used again that ℓ is bounded from below. This concludes the proof of (2.3). \square

2.4. Sobolev norms. Given a bounded domain $\Omega \subseteq \mathbb{R}^d$ with Lipschitz boundary and $p \in (1, \infty)$, with Hölder conjugate $q = p/(p-1)$, we write $\|f\|_{L^p(\Omega)}$ for the Lebesgue norm of f , and

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

for the negative Sobolev norm. We notice in particular that if $\|f\|_{W^{-1,p}(\Omega)} < \infty$ then $\int_{\Omega} f = 0$. In this case we may also restrict the supremum to functions ϕ having also average zero. When it is clear from the context, we will drop the explicit dependence on Ω in the norms.

Let us recall that we can bound the $W^{-1,p}$ norm by the L^p norm. We give here a proof based on the embedding $L^{pd/(p+d)} \subset W^{-1,p}$ (for $p > d/(d-1)$) which is an elementary alternative to the PDE arguments used in [24, Lemma 3.4].

Lemma 2.3. *Let Ω be a bounded domain with Lipschitz boundary and let $f : \Omega \rightarrow \mathbb{R}$ such that $\int_{\Omega} f = 0$. Then, for every $p > d/(d-1)$,*

$$\|f\|_{W^{-1,p}(\Omega)} \lesssim |\Omega|^{\frac{1}{d}} \|f\|_{L^p(\Omega)}. \quad (2.7)$$

Moreover, the implicit constant depends on Ω only through the corresponding constant for the Sobolev embedding.

Proof. Let q be the Hölder conjugate of p , q^* the Sobolev conjugate of q and $p^* = pd/(p+d)$ the Hölder conjugate of q^* . We then have for every ϕ with $\|\nabla\phi\|_{L^q(\Omega)} \leq 1$,

$$\int_{\Omega} f\phi \leq \left(\int_{\Omega} |f|^{p^*} \right)^{\frac{1}{p^*}} \left(\int_{\Omega} |\phi|^{q^*} \right)^{\frac{1}{q^*}} \lesssim \left(\int_{\Omega} |f|^{p^*} \right)^{\frac{1}{p^*}} \left(\int_{\Omega} |\nabla\phi|^q \right)^{\frac{1}{q}} \leq \left(\int_{\Omega} |f|^{p^*} \right)^{\frac{1}{p^*}}.$$

Using that $p^* < p$ and Hölder inequality concludes the proof of (2.7). \square

As in [2], (2.7) will however not be precise enough when estimating the error in subadditivity in the case of general densities and domains. We will instead rely on gradient bounds for the Green kernel $(G(x, y))_{x, y \in \Omega}$ of the Laplacian with Neumann boundary conditions to obtain sharper estimates. See [5, 1, 33, 22] for related results. Let us however point out that in our case we will not rely on any stochastic cancellation in the form of Rosenthal inequality [39] but will instead use a purely deterministic estimate. We will assume that

$$|\nabla_x G(x, y)| \lesssim |x - y|^{1-d}, \quad \text{for every } x, y \in \Omega, \quad (2.8)$$

where the implicit constant depends uniquely on Ω .

Remark 2.4. This condition is satisfied for instance if Ω is C^2 or convex, see e.g. [46]. Notice that since it is a local condition it also holds for $Q \setminus \Omega$ with Ω a C^2 open set with $d(\partial Q, \partial\Omega) > 0$.

Remark 2.5. Let us point out that as in [33], instead of (2.8) it would have been enough to have L^p bounds (for the same p as for the cost \mathcal{C}_p^p) on the Riesz transform for the Neumann Laplacian. From the available results for the Dirichlet Laplacian [30, 40], we expect that for every Lipschitz domain there is $p > 3$ (depending on the domain) for which these bounds hold. In particular, this would allow to extend the validity of Theorem 1.1 to every Lipschitz domain when $d = 3$. However, since we were not able to find in the literature the corresponding results for the case of Neumann boundary conditions we kept the stronger hypothesis (2.8).

We then have

Lemma 2.6. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary, such that (2.8) holds and let ρ be a density bounded above and below on Ω . For $\delta > 0$ sufficiently small, let $(\Omega_k)_{k=1}^K = \mathcal{Q}_\delta \cup \mathcal{R}_\delta$ as in Lemma 2.1. If there exists $h > 0$ such that $|b_k| \leq h^{\frac{1}{2}} |\Omega_k|^{\frac{1}{2}}$ for $k = 1, \dots, K$, then for every $p \geq 1$,*

$$\left\| \sum_{k=1}^K \frac{b_k}{\rho(\Omega_k)} (I_{\Omega_k} - \rho(\Omega_k)) \rho \right\|_{W^{-1,p}(\Omega)} \lesssim \delta^{1-\frac{d}{2}} h^{\frac{1}{2}}. \quad (2.9)$$

Proof. Set

$$B_k = \frac{b_k}{\rho(\Omega_k)}, \quad f_k = (I_{\Omega_k} - \rho(\Omega_k)) \rho.$$

Let then ϕ_k denotes the solution to the equation $\Delta\phi_k = f_k$, with null Neumann boundary conditions on Ω and use as competitor $\xi = \sum_{k=1}^K B_k \nabla\phi_k$ in the definition of the $W^{-1,p}$ norm. We get,

$$\left\| \sum_{k=1}^K B_k f_k \right\|_{W^{-1,p}(\Omega)}^p \leq \int_{\Omega} \left| \sum_{k=1}^K B_k \nabla\phi_k \right|^p \lesssim h^{\frac{p}{2}} \int_{\Omega} \left(\sum_{k=1}^K |\Omega_k|^{-\frac{1}{2}} |\nabla\phi_k| \right)^p. \quad (2.10)$$

To bound the last term, we use the integral representation in terms of the Green's function,

$$\phi_k = \int_{\Omega} G(x, y) f_k(y) dy,$$

to obtain that, for every $x \in \Omega$,

$$|\nabla \phi_k(x)| \lesssim \min \left\{ \text{diam}(\Omega_k), \frac{|\Omega_k|}{\mathbf{d}(x, \Omega_k)^{d-1}} \right\}. \quad (2.11)$$

Indeed, by (2.8),

$$\begin{aligned} |\nabla \phi_k(x)| &\lesssim \int_{\Omega_k} \frac{dy}{|x-y|^{d-1}} + |\Omega_k| \int_{\Omega} \frac{dy}{|x-y|^{d-1}} \leq \int_{\{|y| \leq \text{diam}(\Omega_k)\}} \frac{dy}{|y|^{d-1}} + |\Omega_k| \\ &\lesssim \text{diam}(\Omega_k). \end{aligned}$$

Moreover, for $x \notin \Omega_k$, we get directly from (2.8),

$$|\nabla \phi_k(x)| \lesssim \frac{|\Omega_k|}{\mathbf{d}(x, \Omega_k)^{d-1}}.$$

For any $k = 1, \dots, K$ and $x \in \Omega_k$, we then estimate

$$\begin{aligned} \sum_{j=1}^K |\Omega_j|^{-\frac{1}{2}} |\nabla \phi_j(x)| &\stackrel{(2.11)}{\lesssim} \sum_{j: \mathbf{d}(x, \Omega_j) \leq \text{diam}(\Omega_k)} |\Omega_j|^{-\frac{1}{2}} \text{diam}(\Omega_k) \\ &\quad + \sum_{j: \mathbf{d}(x, \Omega_j) > \text{diam}(\Omega_k)} \frac{|\Omega_j|^{-\frac{1}{2}}}{\mathbf{d}(x, \Omega_j)^{d-1}} \\ &\lesssim \delta^{1-\frac{d}{2}} \left(\text{diam}(\Omega_k)^d + \log(\text{diam}(\Omega_k)) \right), \end{aligned}$$

having used inequalities (2.2) and (2.3) from Lemma 2.2.

Therefore, we can split the integration

$$\begin{aligned} \int_{\Omega} \left(\sum_{j=1}^K |\Omega_j|^{-\frac{1}{2}} |\nabla \phi_j| \right)^p &= \sum_{k=1}^K \int_{\Omega_k} \left(\sum_{j=1}^K |\Omega_j|^{-\frac{1}{2}} |\nabla \phi_j| \right)^p \\ &\lesssim \delta^{p(1-\frac{d}{2})} \sum_{k=1}^K (|\Omega_k|^{p+1} + |\Omega_k| |\log(\text{diam}(\Omega_k))|^p) \\ &\lesssim \delta^{p(1-\frac{d}{2})}, \end{aligned}$$

estimating e.g. $|\log(\text{diam}(\Omega_k))|^p \lesssim \text{diam}(\Omega_k)^{-\frac{1}{2}}$ and using (2.1) (with $\alpha = (p+1)d$ and $\alpha = d - 1/2 > d - 1$). In combination with (2.10) this concludes the proof of (2.9). \square

2.5. Optimal Transport. Given two positive Borel measures μ, λ on \mathbb{R}^d with $\mu(\mathbb{R}^d) = \lambda(\mathbb{R}^d) \in (0, \infty)$ and finite p -th moments, the optimal transport cost of order $p \geq 1$ between μ and λ is defined as the quantity

$$\mathbf{W}^p(\mu, \lambda) = \min_{\pi \in \Gamma(\mu, \lambda)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p d\pi(x, y),$$

where $\Gamma(\mu, \lambda)$ is the set of couplings between μ and λ , i.e., finite Borel measures π on the product $\mathbb{R}^d \times \mathbb{R}^d$ such that their marginals are respectively μ and λ . Notice that if $\mu(\mathbb{R}^d) = \lambda(\mathbb{R}^d) = 0$ then $\mathbf{W}^p(\mu, \lambda) = 0$, while if $\mu(\Omega) \neq \lambda(\Omega)$, we conveniently extend the definition setting $\mathbf{W}^p(\mu, \lambda) = \infty$. Let us recall that the triangle inequality for the Wasserstein distance of order p (which is defined as the the p -th root of $\mathbf{W}^p(\mu, \lambda)$) yields

$$\mathbf{W}^p(\mu, \nu) \lesssim \mathbf{W}^p(\mu, \lambda) + \mathbf{W}^p(\nu, \lambda), \quad (2.12)$$

A straightforward, but useful subadditivity inequality is

$$\mathbf{W}^p \left(\sum_k \mu_k, \sum_k \nu_k \right) \leq \sum_k \mathbf{W}^p(\mu_k, \nu_k). \quad (2.13)$$

valid for any (countable) family of measures $(\mu_k, \nu_k)_k$.

To keep notation simple, we write

$$\mathbb{W}_\Omega^p(\mu, \lambda) = \mathbb{W}^p(\mu \llcorner \Omega, \lambda \llcorner \Omega).$$

and, if a measure is absolutely continuous with respect to Lebesgue measure, we only write its density. For example, $\mathbb{W}_\Omega^p(\mu, \mu(\Omega)/|\Omega|)$ denotes the transportation cost between $\mu \llcorner \Omega$ to the uniform measure on Ω with total mass $\mu(\Omega)$.

For $q \geq p$, Jensen inequality gives

$$\mathbb{W}_\Omega^p(\mu, \nu) \leq \mu(\Omega)^{1-\frac{p}{q}} \left(\mathbb{W}_\Omega^q(\mu, \nu) \right)^{\frac{p}{q}}. \quad (2.14)$$

Our arguments make substantial use of two crucial properties of the optimal transport cost. The first one [24, Lemma 3.1] is a simple consequence of (2.12) and (2.13).

Lemma 2.7. *For every $p \geq 1$, there exists a constant $C > 0$ depending only on p such that the following holds. Let $\Omega \subseteq \mathbb{R}^d$ be Borel and $(\Omega_k)_{k \in \mathbb{N}}$ be a countable Borel partition of Ω . Then, for finite measures μ, λ , and $\varepsilon \in (0, 1)$, we have the inequality*

$$\mathbb{W}_\Omega^p(\mu, \alpha\lambda) \leq (1 + \varepsilon) \sum_i \mathbb{W}_{\Omega_k}^p(\mu, \alpha_k\lambda) + \frac{C}{\varepsilon^{p-1}} \mathbb{W}_\Omega^p\left(\sum_k \alpha_k \chi_{\Omega_k} \lambda, \alpha\lambda\right). \quad (2.15)$$

where $\alpha = \mu(\Omega)/\lambda(\Omega)$ and $\alpha_k = \mu(\Omega_k)/\lambda(\Omega_k)$.

The second one is [2, Lemma 2.2] which gives an upper bound for the Wasserstein distance in terms of a negative Sobolev norm. It follows from the Benamou-Brenier formulation of the optimal transport problem (see also [37, Corollary 3]).

Lemma 2.8. *Assume that $\Omega \subseteq \mathbb{R}^d$ is a bounded connected open set with Lipschitz boundary. If μ and λ are measures on Ω with $\mu(\Omega) = \lambda(\Omega)$, absolutely continuous with respect to the Lebesgue measure and $\inf_\Omega \lambda > 0$, then, for every $p \geq 1$,*

$$\mathbb{W}_\Omega^p(\mu, \lambda) \lesssim \frac{1}{\inf_\Omega \lambda^{p-1}} \|\mu - \lambda\|_{W^{-1,p}(\Omega)}^p. \quad (2.16)$$

As in many recent works on the matching problem, we will use this inequality to improve on the trivial bound

$$\mathbb{W}_\Omega^p(\mu, \lambda) \leq \text{diam}(\Omega)^p \mu(\Omega), \quad (2.17)$$

which holds as soon as $\mu(\Omega) = \nu(\Omega)$. Much of our effort in the proofs will be ultimately to deal with an intermediate situation, where the measures can be decomposed as the sum of a “good” part, i.e., absolutely continuous with smooth density and a “bad” remainder about which not much can be assumed. We prove here a general inequality which could also be of independent interest.

Proposition 2.9. *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded Lipschitz domain, ρ be a density bounded above and below on Ω , μ be any finite measure on Ω and $h > 0$. Then, for every $p > d/(d-1)$,*

$$\mathbb{W}_\Omega^p(\mu + h\rho, \alpha\rho) \lesssim \frac{1}{h^{\frac{p}{d}}} \mu(\Omega)^{1+\frac{p}{d}}, \quad (2.18)$$

where $\alpha = \frac{\mu(\Omega)}{\rho(\Omega)} + h$. Moreover, this inequality is invariant by rescaling of Ω .

Proof. By scaling we may assume that $|\Omega| = 1$. Notice that, by the trivial bound

$$\mathbb{W}^p(\mu + h\rho, \alpha\rho) \lesssim \mu(\Omega),$$

we can assume that $\mu(\Omega) \ll h$. Let P_t be the heat semi-group with null Neumann boundary conditions on Ω and set $\mu_t = P_t\mu$. By triangle inequality (2.12) and (2.16), we have

$$\begin{aligned} \mathbb{W}^p(\mu + h\rho, \alpha\rho) &\lesssim \mathbb{W}^p(\mu + h\rho, \mu_t + h\rho) + \mathbb{W}^p(\mu_t + h\rho, \alpha\rho) \\ &\lesssim \mathbb{W}^p(\mu, \mu_t) + \frac{1}{\eta^{p-1}} \left\| \mu_t - \frac{\mu(\Omega)}{\rho(\Omega)} \rho \right\|_{W^{-1,p}}^p \\ &\lesssim t^{\frac{p}{2}} \mu(\Omega) + \frac{1}{\eta^{p-1}} \left\| \mu_t - \frac{\mu(\Omega)}{\rho(\Omega)} \rho \right\|_{W^{-1,p}}^p. \end{aligned}$$

We now estimate the last term. For this let q be the Hölder conjugate exponent of p , i.e., $q = p/(p-1) \in (1, d)$ and $q^* = qd/(d-q)$ be the Sobolev conjugate of q . We first use triangle inequality and the fact that ρ is bounded from above and below to estimate

$$\begin{aligned} \left\| \mu_t - \frac{\mu(\Omega)}{\rho(\Omega)} \rho \right\|_{W^{-1,p}} &\leq \|\mu_t - \mu(\Omega)\|_{W^{-1,p}} + \mu(\Omega) \left\| 1 - \frac{1}{\rho(\Omega)} \rho \right\|_{W^{-1,p}} \\ &\lesssim \|\mu_t - \mu(\Omega)\|_{W^{-1,p}} + \mu(\Omega). \end{aligned}$$

Using that the Sobolev embedding is equivalent to ultra-contractivity i.e. if $\int_{\Omega} \phi = 0$,

$$\|\phi_t\|_{L^\infty(\Omega)} \lesssim t^{-\frac{d}{2q^*}} \|\phi\|_{L^{q^*}(\Omega)} \lesssim t^{-\frac{d}{2q^*}} \|\nabla \phi\|_{L^q(\Omega)},$$

we finally estimate for every ϕ with $\|\nabla \phi\|_{L^q(\Omega)} \leq 1$ and $\int_{\Omega} \phi = 0$,

$$\int_{\Omega} \phi(\mu_t - \mu(\Omega)) = \int_{\Omega} \phi_t d\mu \leq \mu(\Omega) \|\phi_t\|_{L^\infty(\Omega)} \leq \mu(\Omega) t^{-\frac{d}{2q^*}}.$$

Therefore, by taking the supremum over ϕ we find

$$\|\mu_t - \mu(\Omega)\|_{W^{-1,p}} \lesssim \mu(\Omega) t^{-\frac{d}{2q^*}}.$$

Taking the p -th power we find for $t \leq 1$,

$$\begin{aligned} W^p(\mu + h\rho, \alpha\rho) &\lesssim \mu(\Omega) \left[t^{\frac{p}{2}} + t^{-\frac{pd}{2q^*}} \left(\frac{\mu(\Omega)}{h} \right)^{p-1} \right] \\ &= \mu(\Omega) \left[t^{\frac{p}{2}} + t^{-\frac{p}{2} \left(\frac{d}{q} - 1 \right)} \left(\frac{\mu(\Omega)}{h} \right)^{p-1} \right]. \end{aligned}$$

Optimizing in t we find $t^{\frac{p}{2}} = \left(\frac{\mu(\Omega)}{h} \right)^{\frac{(p-1)q}{d}}$ which satisfies $t \ll 1$ if $\mu(\Omega) \ll h$. Since $(p-1)q = p$, this concludes the proof of (2.18). \square

Remark 2.10. Since by Hölder inequality it will be enough for us to apply Proposition 2.9 for p arbitrarily close to d , the condition $p > d/(d-1)$ will not be a limitation for us. Let us however mention that, in the critical case $p = d/(d-1)$ one can argue similarly, relying instead on the Moser-Trudinger inequality [15, Remark 1.4], to obtain (in the case $\rho = 1$ and $\Omega = Q$ a cube for simplicity)

$$W_Q^p \left(\mu + h, \frac{\mu(Q)}{|Q|} + h \right) \lesssim \mu(Q) \left| \log \left(\frac{\mu(Q)}{|Q|h} \right) \right| \left(\frac{\mu(Q)}{h} \right)^{\frac{1}{d-1}}.$$

If instead $p < d/(d-1)$, using the same proof as above but with the inclusion $W^{1,q}(Q) \subseteq L^\infty(Q)$ and letting $t \rightarrow 0$ gives the estimate

$$\frac{1}{|Q|^{\frac{p}{d}}} W_Q^p \left(\mu + h, \frac{\mu(Q)}{|Q|} + h \right) \lesssim \mu(Q) \left(\frac{\mu(Q)}{|Q|h} \right)^{p-1}.$$

We close this section with the following result easily adapted from [2, Proposition 2.4] which helps in particular to reduce the transport problem from Hölder to constant densities.

Proposition 2.11. *For $d \geq 1$, $\alpha \in (0, 1)$ and $\rho_0 > 0$, there exists $C = C(\rho_0, d, \alpha) > 0$ such that the following holds: for any $\rho \in C^\alpha((0, 1)^d)$ with*

$$\int_{(0,1)^d} \rho = 1 \quad \text{and} \quad \rho_0 \leq \rho \leq \rho_0^{-1},$$

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

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

2.6. A subadditivity lemma. We will need a slight variant of the usual convergence results for subadditive functions, see e.g. [43, 11].

Lemma 2.12. *Let $\alpha, \beta, c > 0$, $f : [1, \infty) \rightarrow [0, \infty)$ be continuous and such that the following holds: for every $\eta \in (0, 1/2]$, there exists $C(\eta) > 0$ such that, for every $m \in \mathbb{N} \setminus \{0\}$ and $L \geq C(\eta)$,*

$$f(mL) \leq f(L(1 - \eta)) + c\eta^\alpha + C(\eta)L^{-\beta}. \quad (2.19)$$

Then $\lim_{L \rightarrow \infty} f(L) \in [0, \infty)$ exists.

Proof. We use the following fact: for any open interval $(a, b) \subseteq [0, \infty)$, the union

$$\bigcup_{m=1}^{\infty} (ma, mb) \supseteq (A, +\infty)$$

contains a half-line, for some $A > 0$. Indeed, one has $(ma, mb) \cap ((m+1)a, (m+1)b) \neq \emptyset$ if $mb > (m+1)a$, which holds for every $m > a/(b-a)$.

First, we show that f is uniformly bounded. Let $\eta = 1/2$ and use the fact that both $f(L)$ and $L^{-\beta}$ are continuous for $L \in [1/2, 2]$, hence bounded, so that by (2.19), for every $m \geq 1$, $L \in [1, 2]$,

$$f(mL) \leq \sup_{\ell \in [1, 2]} \left(f(\ell/2) + c2^{-\alpha} + C(1/2)\ell^{-\beta} \right) < \infty.$$

since $\bigcup_{m=1}^{\infty} [m, 2m] = [1, \infty)$, it follows that f is uniformly bounded on $[1, \infty)$. To show that the limit exists (and is finite) we argue that

$$\limsup_{L \rightarrow \infty} f(L) \leq \liminf_{L \rightarrow \infty} f(L).$$

Given $\varepsilon \ll 1$, let $\eta = \eta(\varepsilon) \in (0, 1/2]$ such that $c\eta^\alpha = \varepsilon$ and $L_\varepsilon > 0$ such that $C(\eta)L_\varepsilon^{-\beta} = \varepsilon$, so that, for every $L \geq L_\varepsilon$,

$$C(\eta)L^{-\beta} \leq \varepsilon.$$

Let then $L^* > \max\{L_\varepsilon, C(\eta)\}$ be such that

$$f(L^*) < \liminf_{L \rightarrow \infty} f(L) + \varepsilon.$$

By continuity of f , there exists $a < L^* < b$ with $a > \max\{L_\varepsilon, C(\eta)\}$ such that the same inequality holds for $L \in (a, b)$. For every $m \geq 1$, and $L \in (a/(1-\eta), b/(1-\eta))$, we have $L \geq \max\{L_\varepsilon, C(\eta)\}$ and $L(1-\eta) \in (a, b)$, hence using (2.19) we obtain

$$f(mL) \leq f(L(1-\eta)) + c\eta^\alpha + C(\eta)L^{-\beta} \leq \liminf_{L \rightarrow \infty} f(L) + 3\varepsilon.$$

Using that $\bigcup_{m=1}^{\infty} (ma/(1-\eta), mb/(1-\eta))$ contains a half-line $(A, +\infty)$, it follows that

$$\limsup_{L \rightarrow \infty} f(L) \leq \liminf_{L \rightarrow \infty} f(L) + 3\varepsilon,$$

and the thesis follows letting $\varepsilon \rightarrow 0$. \square

2.7. Concentration inequalities. We close this section by recalling some standard concentration inequalities. Let us start with a general definition.

Definition 2.13. We say that a random variable X with $\mathbb{E}[X] = h$ satisfies (algebraic) concentration if for every $q \geq 1$ there exists $C(q) \in (0, \infty)$ such that

$$\mathbb{E}[|X - h|^q] \leq C(q)|h|^{\frac{q}{2}}.$$

We then have

Lemma 2.14. *Poisson, binomial and hypergeometric random variables satisfy concentration. More precisely, if :*

i) N is a Poisson random variable with parameter $n \geq 1$ then, for every $q \geq 1$,

$$\mathbb{E}[|N - n|^q] \lesssim_q n^{\frac{q}{2}}. \quad (2.20)$$

Hence, for every $\gamma \in (0, 1)$,

$$\mathbb{P}(N < \gamma n \quad \text{or} \quad N > (1 + \gamma)n) \lesssim_{q,\gamma} (1 - \gamma)^{-2q} n^{-q}. \quad (2.21)$$

ii) B is a binomial random variable with parameters n and $p \in (0, 1)$ (so that $\mathbb{E}[B] = np$) then, for every $q \geq 1$,

$$\mathbb{E}[|B - np|^q] \lesssim_q n^{\frac{q}{2}}. \quad (2.22)$$

iii) H is a hypergeometric random variables counting the number of red marbles extracted in z draws without replacement from an urn containing u marbles, r of which are red (so that $\mathbb{E}[H] = zr/u$) then, for every $q \geq 1$,

$$\mathbb{E}[|H - zr/u|^q] \lesssim_q r^{\frac{q}{2}}. \quad (2.23)$$

Proof. We only prove concentration in the hypergeometric case, since it is classical for both Poisson and binomial random variables. We may assume that $r \geq 1$, otherwise there is nothing to prove since $H = \mathbb{E}[H] = 0$. From [28, Theorem 1], we have, for $\lambda \geq 2$,

$$\mathbb{P}(|H - \mathbb{E}[H]| \geq \lambda) \leq 2 \exp(-\alpha \lambda^2),$$

where

$$\alpha = \min \left\{ \frac{1}{z+1} + \frac{1}{u-z+1}, \frac{1}{r+1} + \frac{1}{u-r+1} \right\} \geq \frac{u+2}{(r+1)(u-r+1)} \geq \frac{1}{r+1}.$$

As usual, writing

$$\mathbb{E}[|H - \mathbb{E}[H]|^q] = \int_0^\infty \mathbb{P}(|H - \mathbb{E}[H]| \geq \lambda) p \lambda^{q-1} d\lambda,$$

yields the bound

$$\mathbb{E}[|H - \mathbb{E}[H]|^q] \lesssim_q 1 + \alpha^{-\frac{q}{2}} \lesssim_q 1 + (r+1)^{\frac{q}{2}},$$

which is bounded from above by $r^{q/2}$, since $r \geq 1$. \square

3. COMBINATORIAL OPTIMIZATION PROBLEMS OVER BIPARTITE GRAPHS

3.1. Graphs. Although we are interested in random combinatorial optimization over Euclidean bipartite graphs, it is useful to recall some general terminology. A (finite, undirected) graph $G = (V, E)$ is defined by a finite set $V = V_G$ of vertices (or nodes) and a set of edges $E = E_G$, which is a collection of unordered pairs $e = \{x, y\} \subseteq V$ with $x \neq y$. A graph G' is a subgraph of G and we write $G' \subseteq G$, if $V_{G'} \subseteq V_G$ and $E_{G'} \subseteq E_G$. The induced subgraph over a subset of vertices $V' \subseteq V_G$ is defined as the subgraph G' with $V_{G'} = V'$ and all the edges from E_G connecting vertices in V' . It will be useful to denote by \emptyset the empty graph, i.e., $V = \emptyset$, $E = \emptyset$, which is a subgraph of any graph G .

Given a vertex $x \in V$, its neighborhood in G is the set

$$\mathcal{N}_G(x) = \{y \in V : \{x, y\} \in E\}.$$

The degree of x in G , $\deg_G(x)$, is the number of elements in $\mathcal{N}_G(x)$. Given $\kappa \in \mathbb{N}$, a graph G is κ -regular if $\deg_G(x) = \kappa$ for every $x \in V_G$. We say that a subgraph $G' \subseteq G$ spans V_G if $V_{G'} = V_G$ and $\mathcal{N}_{G'}(x) \neq \emptyset$ for ever $x \in V_{G'}$. We say that two subgraphs G_1, G_2 of G are disjoint if $V_{G_1} \cap V_{G_2} = \emptyset$. A graph G is connected if it cannot be decomposed as the union of two disjoint subgraphs $G = G_1 \cup G_2$, i.e., $V_G = V_{G_1} \cup V_{G_2}$ with both $V_{G_1}, V_{G_2} \neq \emptyset$, $V_{G_1} \cap V_{G_2} = \emptyset$ and $E_G = E_{G_1} \cup E_{G_2}$. Given $\kappa \in \mathbb{N}$, $\kappa \geq 1$, we say that a graph G is κ -connected if any subgraph $G' \subseteq G$ obtained by removing from G $(\kappa - 1)$ -edges is still connected. A cycle is a connected 2-regular graph, a tree is a connected graph which contains no cycles as subgraphs.

Given two graphs G_1, G_2 and an injective function $\sigma : V_{G_1} \rightarrow V_{G_2}$, we let $\sigma(E_1) = \{\{\sigma(x), \sigma(y)\} : \{x, y\} \in E_{G_1}\}$. If $\sigma(E_1) \subseteq E_2$, then we say that G_1 embeds into G_2 via σ . If σ is bijective and $\sigma(E_1) = E_2$, then we say that G_1 is isomorphic to G_2 via σ .

A graph G is complete if E_G consists of all the pairs $\{x, y\} \subseteq V$ with $x \neq y$. The complete graph over $V = [n]$ is commonly denoted by \mathcal{K}_n . Any complete graph G with n vertices is isomorphic to \mathcal{K}_n . We say that the graph G is bipartite over a partition $V = X \cup Y$ (i.e., $X \cap Y = \emptyset$), if every $e \in E$ can be written as $e = \{x, y\}$ with $x \in X, y \in Y$. A graph is complete bipartite if it is bipartite over a partition $V = X \cup Y$ and every pair $\{x, y\}$ with $x \in X, y \in Y$ is an edge. For any $n, m \in \mathbb{N}$, any two complete bipartite graphs with X having n elements and Y having m elements are isomorphic. To fix a representative, we define $\mathcal{K}_{n,m}$ as the complete bipartite graph over the vertex set $V = [n]_1 \cup [m]_2$.

We introduce a weight function on edges $w : E \rightarrow [0, \infty)$, $w(e) = w(x, y)$. The total weight of G is then

$$w(G) = \sum_{e \in E} w(e).$$

A subgraph $G' \subseteq G$ of a weighted graph is always intended with the restriction of w on E' . Notice that for the empty graph $\emptyset \subseteq G$ we have $w(\emptyset) = 0$.

We are interested in geometric realizations of graphs, where vertices are in correspondence with points in a metric space (Ω, d) , and the weight function is a power of the distance between the corresponding points, with a fixed exponent $p > 0$. Since we consider only complete and complete bipartite graphs, we introduce the following notation. Given $\mathbf{x} = (x_i)_{i=1}^n \subseteq \Omega$, we let $\mathcal{K}(\mathbf{x})$ be the complete graph \mathcal{K}_n endowed with the weight function $w(i, j) = d(x_i, x_j)^p$. Similarly, given $\mathbf{x} = (x_i)_{i=1}^n, \mathbf{y} = (y_j)_{j=1}^m \subseteq \Omega$, we let $\mathcal{K}(\mathbf{x}, \mathbf{y})$ denote the complete bipartite graph $\mathcal{K}_{n,m}$ endowed with the weight function $w((1, i), (2, j)) = d(x_i, y_j)^p$. Notice that the points in \mathbf{x} and \mathbf{y} may not be all distinct, but this will in fact occur with probability zero. If all the points are distinct, then we can and will identify the vertex set directly with the set of points \mathbf{x} for $\mathcal{K}(\mathbf{x})$, and with the set of points in $\mathbf{x} \cup \mathbf{y}$ for $\mathcal{K}(\mathbf{x}, \mathbf{y})$. With this convention, if $\mathbf{x} = \mathbf{x}^0 \cup \mathbf{x}^1, \mathbf{y} = \mathbf{y}^0 \cup \mathbf{y}^1$, then both $\mathcal{K}(\mathbf{x}^0, \mathbf{y}^0)$ and $\mathcal{K}(\mathbf{x}^1, \mathbf{y}^1)$ are naturally seen as subgraphs of $\mathcal{K}(\mathbf{x}, \mathbf{y})$.

3.2. Combinatorial problems. A combinatorial optimization problem \mathbf{P} on weighted graphs is informally defined by prescribing, for every graph G , a set of subgraphs $G' \subseteq G$, also called feasible solutions \mathcal{F}_G , and, after introducing a weight w , by minimizing $w(G')$ over all $G' \in \mathcal{F}_G$.

Our aim is to study problems on random geometric realizations of complete bipartite graphs $\mathcal{K}_{n,n}$, thus it is sufficient to define a combinatorial optimization problem over complete bipartite graphs as a collection of feasible solutions $\mathbf{P} = (\mathcal{F}_{n,n})_{n \in \mathbb{N}}$, with $\mathcal{F}_{n,n}$ being the feasible solutions on $\mathcal{K}_{n,n}$. We will mostly consider problems \mathbf{P} that satisfy the following assumptions:

- A1 (*isomorphism*) if σ is any isomorphism of $\mathcal{K}_{n,n}$ into itself and $G \in \mathcal{F}_{n,n}$, then $\sigma(G) = (\sigma(V_G), \sigma(E_G)) \in \mathcal{F}_{n,n}$;
- A2 (*spanning*) for every $n \in \mathbb{N}$, $\mathcal{F}_{n,n}$ is not empty and there exists $c_{A2} > 0$ such that, for $n < c_{A2}$, $\mathcal{F}_{n,n} = \{\emptyset\}$ while for $n \geq c_{A2}$, every $G \in \mathcal{F}_{n,n}$ spans $\mathcal{K}_{n,n}$;
- A3 (*bounded degree*) there exists $c_{A3} > 0$ such that, for every $n \in \mathbb{N}$ and every feasible solution $G \in \mathcal{F}_{n,n}$, one has $\deg_G(x) \leq c_{A3}$ for every $x \in G$.

Given $\mathbf{P} = (\mathcal{F}_{n,n})_{n \in \mathbb{N}}$, we canonically extend it to graphs $\mathcal{K}_{n,m}$, with $n \neq m$, defining $\mathcal{F}_{n,m}$ as the collection of all graphs $\sigma(G)$ where $G \in \mathcal{F}_{z,z}$, $z = \min\{n, m\}$ and σ is an isomorphism of $\mathcal{K}_{n,m}$ into itself.

In the geometric setting, i.e., when $\mathcal{K}_{n,m}$ is mapped into $\mathcal{K}(\mathbf{x}, \mathbf{y})$ with $\mathbf{x} = (x_i)_{i=1}^n, \mathbf{y} = (y_j)_{j=1}^m \subseteq \Omega$, with (Ω, d) metric space, we introduce the following notation for the

cost of a problem P:

$$\mathcal{C}_P^p(\mathbf{x}, \mathbf{y}) = \min_{G \in \mathcal{F}_{n,m}} \sum_{\{(1,i), (2,j)\} \in E_G} d(x_i, y_j)^p.$$

Recalling the definition of $\mathcal{F}_{n,m}$ if $n \neq m$, we also have the identity

$$\mathcal{C}_P^p(\mathbf{x}, \mathbf{y}) = \min_{\substack{\mathbf{x}' \subseteq \mathbf{x}, \mathbf{y}' \subseteq \mathbf{y} \\ |\mathbf{x}'| = |\mathbf{y}'| = \min\{|\mathbf{x}|, |\mathbf{y}|\}}} \mathcal{C}_P^p(\mathbf{x}', \mathbf{y}'). \quad (3.1)$$

Remark 3.1. Assumption A2 ensures that, if $\min\{|\mathbf{x}|, |\mathbf{y}|\} < c_{A2}$, then $\mathcal{C}_P^p(\mathbf{x}, \mathbf{y}) = 0$.

Remark 3.2. If (Ω', d') is a metric space and $S : \Omega \rightarrow \Omega'$ is Lipschitz, i.e., for some constant $\text{Lip } S$ one has $d'(S(x), S(y)) \leq (\text{Lip } S)d(x, y)$ for every $x, y \in \Omega$, then writing $S(\mathbf{x}) = (S(x_i))_{i=1}^n$, $S(\mathbf{y}) = (S(y_j))_{j=1}^m$, we clearly have the inequality

$$\mathcal{C}_P^p(S(\mathbf{x}), S(\mathbf{y})) \leq (\text{Lip } S)^p \mathcal{C}_P^p(\mathbf{x}, \mathbf{y}). \quad (3.2)$$

Remark 3.3. Similar definitions and assumptions may be given in the non-bipartite case, thus defining combinatorial optimization problems $P = (\mathcal{F}_n)_{n \in \mathbb{N}}$ over complete graphs, as a collection of feasible solutions \mathcal{F}_n over the complete graph \mathcal{K}_n .

3.3. Examples. Let us introduce some fundamental examples of these problems.

Assignment problem. The minimum weight bipartite matching problem, also called assignment problem, is defined letting $\mathcal{F}_{n,n}$ be the set of perfect matchings in $\mathcal{K}_{n,n}$, i.e., spanning subgraphs induced by a collection of edges which have no vertex in common (if $n = 0$ we simply let $\mathcal{F}_{n,n} = \{\emptyset\}$). Feasible solutions are in correspondence with permutations σ over $[n]$, letting

$$E_\sigma = \{\{(1, i), (2, \sigma(i))\} : i \in [n]\}.$$

When $n \neq m$, e.g. $n \leq m$, the same correspondence holds with the set of injective maps $\sigma : [n] \rightarrow [m]$. Therefore, given a weight w on $\mathcal{K}_{n,m}$, the cost of the assignment problem is

$$\min_{\sigma} \sum_{i=1}^n w((1, i), (2, \sigma(i))).$$

In the geometric case, i.e., on the weighted graph $\mathcal{K}(\mathbf{x}, \mathbf{y})$ with $\mathbf{x} = (x_i)_{i=1}^n$, $\mathbf{y} = (y_j)_{j=1}^m \subseteq \Omega$ and $w((1, i), (2, j)) = d(x_i, y_j)^p$, this expression becomes

$$M^p(\mathbf{x}, \mathbf{y}) = \min_{\sigma} \sum_{i=1}^n d(x_i, y_{\sigma(i)})^p.$$

If $n > m$, then one simply exchanges the roles of n and m .

Remark 3.4. If $n = m$, Birkhoff's theorem ensures equivalence between the bipartite matching problem and the optimal transport between the associated empirical measures $\mu^{\mathbf{x}} = \sum_{i=1}^n \delta_{x_i}$, $\mu^{\mathbf{y}} = \sum_{j=1}^n \delta_{y_j}$, i.e.,

$$M^p(\mathbf{x}, \mathbf{y}) = W^p(\mu^{\mathbf{x}}, \mu^{\mathbf{y}}). \quad (3.3)$$

Therefore, using the triangle inequality (2.12), we can bound from above as follows:

$$M^p(\mathbf{x}, \mathbf{y}) \lesssim W^p(\mu^{\mathbf{x}}, n\lambda) + W^p(\mu^{\mathbf{y}}, n\lambda), \quad (3.4)$$

for every probability measure λ on \mathbb{R}^d .

Travelling salesperson problem. The travelling salesperson problem (TSP) is usually defined on a general graph by prescribing as feasible solutions the cycles visiting each vertex exactly once (also called Hamiltonian cycles). In the complete bipartite case $\mathcal{K}_{n,n}$, such cycles exist for every $n \geq 2$, and assumptions **A1**, **A2** and **A3** are also clearly satisfied (letting $\mathcal{F}_{n,n} = \{\emptyset\}$ if $n \in \{0,1\}$). Similarly as in the case of the assignment problem, feasible solutions are in this case in correspondence with pairs of permutations σ, τ over $[n]$, letting

$$E_{\sigma,\tau} = \{ \{(1, \sigma(i)), (2, \tau(i))\}, \{(1, \sigma(i)), (2, \tau(i+1))\} : i \in \{1, \dots, n\} \}, \quad (3.5)$$

where we conventionally let $\tau(n+1) = \tau(1)$ (we will always use summation mod n in such cases). In words, σ and τ prescribe the order at which the vertices are visited by the cycle. When $n \neq m$, e.g. $n \leq m$, the same correspondence holds with injective maps σ, τ from $[n]$ into $[m]$.

Therefore, given a weight w on $\mathcal{K}_{n,m}$, the cost of the TSP reads

$$\min_{\sigma,\tau} \sum_{i=1}^n w((1, \sigma(i)), (2, \tau(i))) + w((1, \sigma(i)), (2, \tau(i+1))).$$

In the geometric case, i.e., on the weighted graph $\mathcal{K}(\mathbf{x}, \mathbf{y})$ with $\mathbf{x} = (x_i)_{i=1}^n$, $\mathbf{y} = (y_j)_{j=1}^m \subseteq \Omega$, this becomes

$$\mathcal{C}_{\text{TSP}}^p(\mathbf{x}, \mathbf{y}) = \min_{\sigma,\tau} \sum_{i=1}^n d(x_{\sigma(i)}, y_{\tau(i)})^p + d(x_{\sigma(i)}, y_{\tau(i+1)})^p.$$

If $n > m$, then one simply exchanges the roles of n and m .

The non-bipartite version of the TSP, i.e., on \mathcal{K}_n , feasible solutions to the TSP are in correspondence with permutations σ over $[n]$, letting

$$E_{\sigma} = \{ \{\sigma(i), \sigma(i+1)\} : i \in [n] \}.$$

In the geometric case $\mathbf{x} = (x_i)_{i=1}^n \subseteq \Omega$, it becomes

$$\mathcal{C}_{\text{TSP}}^p(\mathbf{x}) = \min_{\sigma} \sum_{i=1}^n d(x_{\sigma(i)}, x_{\sigma(i+1)})^p.$$

Connected κ -factor problem. The TSP can be generalized in many directions. For example, since a cycle is a connected graph such that every vertex has degree 2, i.e., it is 2-regular, we may instead define as feasible solutions κ -regular spanning connected subgraphs, for a fixed $\kappa \in \mathbb{N}$, $\kappa \geq 2$. This defines a non-empty set of feasible solutions $\mathcal{F}_{n,n}$ over $\mathcal{K}_{n,n}$ if $n \geq \kappa$ (otherwise we let $\mathcal{F}_{n,n} = \{\emptyset\}$) and assumptions **A1**, **A2** and **A3** are easily seen to be satisfied. We refer to such problem as the (minimum weight) connected κ -factor problem. A simpler variant is to require that feasible solutions are κ -regular but not necessarily connected: this is simply known as (minimum weight) κ -factor problem. Let us notice that, for $\kappa = 1$, this reduces to the assignment problem.

Back to the the connected κ -factor problem, a simple fact worth noticing, that we will use below, is that any connected κ -regular bipartite graph G is 2-connected, i.e., it remains connected even after removing a single edge. Assume that $V_G = X \cup Y$, with $X \cap Y = \emptyset$ and by contradiction let $x \in X, y \in Y$ be such that $\{x, y\} \in E_G$ and the subgraph $G' \subseteq G$ with edge set $E_{G'} = E_G \setminus \{x, y\}$ is not connected: there are two disjoint subgraphs G'_1, G'_2 with $x \in V_{G'_1}, y \in V_{G'_2}$ with $G' = G'_1 \cup G'_2$. All the vertices in G'_1 have degree κ , except for x , whose degree is $\kappa - 1$. However, if we let $n_X = |V_{G'_1} \cap X|$, $n_Y = |V_{G'_1} \cap Y|$, then using the fact that the graph G'_1 is bipartite we can count the number of edges as the sum of the degrees of the vertices in $V_{G'_1} \cap X$ or equivalently of those in $V_{G'_1} \cap Y$, which leads to the identity $\kappa n_X - 1 = \kappa n_Y$, from which $\kappa(n_X - n_Y) = 1$, which gives a contradiction.

κ -bounded degree minimum spanning tree. The minimum weight spanning trees (MST) problem is defined by letting feasible solutions be all spanning subgraphs that are trees, i.e., connect and acyclic, whose existence on any given connected graph is guaranteed by standard algorithms. This problem however may not have uniformly bounded degree, thus assumption **A3** may not be satisfied. Therefore, we restrict the set of feasible solutions to spanning trees over $\mathcal{K}_{n,n}$ such that that each vertex degree is less than or equal to some fixed $\kappa \geq 2$ (letting $\mathcal{F}_{0,0} = \{\emptyset\}$). This problem, known as the κ -bounded degree minimum spanning tree (κ -MST), satisfies assumptions **A1**, **A2** and **A3**: notice in particular that removing any edge from a Hamiltonian cycle, i.e., a feasible solution for the TSP, gives a 2-bounded degree minimum spanning tree.

We remark here that the κ -MST problem may be also directly defined over graphs $\mathcal{K}_{n,m}$, with $n \neq m$, with a non trivial set of feasible solutions (provided that $|n - m|$ is not too large). However, also in this case we follow our the general convention, so that if $n \neq m$, the set $\mathcal{F}_{n,m}$ does not contain spanning trees of $\mathcal{K}_{n,m}$ but only spanning trees over subgraphs isomorphic to $\mathcal{K}_{z,z}$ with $z = \min\{n, m\}$.

A simple fact that we will use below is that any $G \in \mathcal{F}_{n,n}$ contains at least one leaf (i.e., a vertex with degree 1) in $[n]_1$ and one in $[n]_2$. This is because more generally any spanning tree over $\mathcal{K}_{n,n}$ contains at least one leaf in $[n]_1$ and one in $[n]_2$. Indeed, assume by contradiction that there are no leaves in $[n]_1$. Then, since the tree spans, all the vertices in $[n]_1$ must have degree at least 2 (the graph is connected, hence every vertex has at least degree 1) and since no edges connect pairs of vertices in $[n]_1$, these are all distinct, hence the tree contains at least $2n$ edges, which contradicts the well-known fact that any tree (not necessarily bipartite) over $2n$ vertices must have $2n - 1$ edges.

In order to perform our analysis, we introduce two further assumptions that we discuss in the following subsections.

3.4. Local merging. Our analysis relies on a key subadditivity inequality, that ultimately follows by a stability assumption with respect to local merging operations, besides assumptions **A1** and **A3**. Let us give the following general definition.

Definition 3.5 (gluing). Given a graph G and two disjoint subgraphs $G_1, G_2 \subseteq G$, we say that $G' \subseteq G$ is obtained by gluing at $x_1 \in V_{G_1}, x_2 \in V_{G_2}$ if $V_{G'} = V_{G_1} \cup V_{G_2}$,

$$(E_{G_1} \cup E_{G_2}) \setminus E_{G'} \subseteq \mathcal{N}_{G_1}(x_1) \cup \mathcal{N}_{G_2}(x_2)$$

and

$$E_{G'} \setminus (E_{G_1} \cup E_{G_2}) \subseteq \{\{x_1, y\} : y \in \mathcal{N}_{G_2}(x_2)\} \cup \{\{x_2, y\} : y \in \mathcal{N}_{G_1}(x_1)\}.$$

In words, gluing at x_1, x_2 means that the two subgraphs are joined by (possibly) removing and adding edges connecting x_2 to vertices from the neighborhood of x_1 in G_1 , and similarly x_1 to vertices from the the neighborhood of x_2 in G_2 . In particular, we have that $\mathcal{N}_{G'}(x) = \mathcal{N}_{G_1}(x)$ for every $x \in V_{G_1} \setminus (\mathcal{N}_{G_1}(x_1) \cup \{x_1\})$, and similarly $\mathcal{N}_{G'}(x) = \mathcal{N}_{G_2}(x)$ for every $x \in V_{G_2} \setminus (\mathcal{N}_{G_2}(x_2) \cup \{x_2\})$.

Back to combinatorial optimization problems over bipartite graphs, our assumption is, loosely speaking, that any two (non empty) feasible solutions $G \in \mathcal{F}_{n,n}, G' \in \mathcal{F}_{n',n'}$, can be glued together yielding a feasible solution $G \in \mathcal{F}_{n+n',n+n'}$. In fact, we also allow adding up to c edges, but only connecting vertices of G , where $c \in \mathbb{N}$ is a constant (depending only on the problem P). Before giving a precise formulation of the assumption, we notice that G and G' are in general not disjoint: what we mean is that G' must be suitably “translated”. Precisely, given $n \in \mathbb{N}$, we introduce the map

$$\tau_n : V_{G'} \rightarrow [n+n']_1 \cup [n+n']_2$$

defined as

$$\tau((1, i)) = (1, n+i), \quad \tau((2, j)) = (2, n+j),$$

so that $G, \tau(G') \subseteq \mathcal{K}_{n+n',n+n'}$ are disjoint.

We consider therefore combinatorial optimization problems \mathbf{P} over bipartite graphs which satisfy the following assumption:

A4 (*local merging*) there exists $c_{A4} \geq 0$ such that, for every $n, n' \in \mathbb{N}$, and $G \in \mathcal{F}_{n,n}$, $G' \in \mathcal{F}_{n',n'}$ with both $G \neq \emptyset$ and $G' \neq \emptyset$, one can find $G'' \in \mathcal{F}_{n+n',n+n'}$ obtained by gluing G and $\tau(G')$ at the vertices $(1, 1)$, $(1, n+1)$ and possibly adding up to c_{A4} edges from those of $\mathcal{K}_{n,n}$.

The reason why we also allow up to c_{A4} additional edges is to include some problems where connectedness may be destroyed by gluing, such as the κ -MST. This should be compared with the *merging* assumption [6, (A4)], where a bounded number of edges from the whole $\mathcal{K}_{n+n',n+n'}$ instead is allowed to be added to the union $G \cup \tau(G')$ (with our notation). Notice however that, in our case, since the extra edges are from $\mathcal{K}_{n,n}$ it remains true that

$$\mathcal{N}_{G''}(x) = \mathcal{N}_{\tau(G')}(x), \text{ for every } x \in V_{\tau(G')} \setminus (\mathcal{N}_{\tau(G')}((1, n+1)) \cup \{(1, n+1)\}), \quad (3.6)$$

which is a key condition that we use below.

All the problems described in the previous section satisfy A4.

Lemma 3.6. *The TSP, the connected κ -factor problem (as well as the non connected one) and the κ -MST over complete bipartite graphs satisfy assumption A4.*

Proof. Let $G \in \mathcal{F}_{n,n}$, $G' \in \mathcal{F}_{n',n'}$ be both non empty. Then (e.g. by assumption A2) $\deg_G(1, 1) \geq 1$ but also $\deg_{\tau(G')}(1, n+1) \geq 1$. The basic idea is to pick $y \in \mathcal{N}_G(1, 1)$, $y' \in \mathcal{N}_{G'}(1, n+1)$, remove the edges $\{(1, 1), y\}$, $\{(1, n+1), y'\}$ and add instead $\{(1, 1), y'\}$, $\{(1, n+1), y\}$. This operation does not change the vertex degrees, in particular at $(1, 1)$ and $(1, n+1)$.

For the TSP and more generally the connected κ -factor problem, the resulting graph G'' is connected, because after removing a single edge, both graphs G and $\tau(G')$ are still connected, and adding the new edges has the effect of connecting the two graphs (hence in this case $c_{A4} = 0$).

For the κ -bounded degree MST, we use the fact that the tree $G \in \mathcal{F}_{n,n}$ must have at least one leaf in the set of $[n]_1$ and one in the set $[n]_2$. Therefore, we obtain a connected tree (with degree bounded by κ) if we add also one edge connecting two such leaves (hence in this case $c_{A4} = 1$). \square

3.5. Subadditivity inequality. Using all the assumptions introduced so far, in particular A4, we establish a fundamental subadditivity inequality.

Proposition 3.7 (Approximate subadditivity). *Let \mathbf{P} be a combinatorial optimization problem over bipartite graphs satisfying assumptions A1, A2, A3 and A4.*

For a metric space (Ω, d) and a finite partition $\Omega = \cup_{k=1}^K \Omega_k$, $K \in \mathbb{N}$,

- i) *let $\mathbf{x}^0, \mathbf{y}^0 \subseteq \Omega$ be such that $\min\{|\mathbf{x}^0|, |\mathbf{y}^0|\} \geq \max\{c_{A2}, K\}$,*
- ii) *for every $k = 1, \dots, K$, let $\mathbf{x}^k, \mathbf{y}^k \subseteq \Omega_k$ with $|\mathbf{x}^k| = |\mathbf{y}^k| = n_k$, with either $n_k \geq c_{A2}$ or $n_k = 0$ (i.e., both families are empty),*
- iii) *let $\mathbf{z} = (z_k)_{k=1}^K$ with $z_k \in \Omega_k$, for every $k = 1, \dots, K$.*

Then, the following inequality holds:

$$\mathcal{C}_{\mathbf{P}}^p \left(\mathbf{x}^0 \cup \bigcup_{k=1}^K \mathbf{x}^k, \mathbf{y}^0 \cup \bigcup_{k=1}^K \mathbf{y}^k \right) - \sum_{k=1}^K \mathcal{C}_{\mathbf{P}}^p(\mathbf{x}^k, \mathbf{y}^k) \lesssim \mathcal{C}_{\mathbf{P}}^p(\mathbf{x}^0, \mathbf{y}^0) + M^p(\mathbf{z}, \mathbf{x}^0) + \sum_{k=1}^K \text{diam}(\Omega_k)^p. \quad (3.7)$$

The implicit constant depends only upon p , c_{A2} , c_{A3} and c_{A4} (in particular not on K).

Remark 3.8. The role played by the points \mathbf{z} is quite marginal, and indeed if $\mathbf{x}^0(\Omega_k) > 0$ for every k , then by choosing $z_k \in \mathbf{x}^0_{\Omega_k}$, the term $M^p(\mathbf{z}, \mathbf{x}^0)$ vanishes.

Proof. Recalling (3.1), up to replacing $\mathbf{x}^0, \mathbf{y}^0$ with subsets \mathbf{x}', \mathbf{y}' with $|\mathbf{x}'| = |\mathbf{y}'| = \min\{|\mathbf{x}^0|, |\mathbf{y}^0|\}$, we may also assume that $|\mathbf{x}^0| = |\mathbf{y}^0|$. For every $k = 1, \dots, K$ let $G_k \subseteq \mathcal{K}(\mathbf{x}^k, \mathbf{y}^k)$ be a minimizer for \mathcal{P} . If $n_k = 0$, then $G_k = \emptyset$. Otherwise, $n_k \geq c_{A2}$, and by assumption A2 it is in particular non-empty and using Markov inequality, we can choose $x^k \in \mathbf{x}^k$ such that

$$\sum_{y \in \mathcal{N}_{G_k}(x^k)} d(x^k, y)^p \leq \frac{4\mathcal{C}_P^p(\mathbf{x}^k, \mathbf{y}^k)}{|\mathbf{x}^k|} \lesssim \text{diam}(\Omega_k)^p.$$

For the last estimate we used that $\deg_{x^k}(G_k) \leq c_{A3}$. Similarly, let $G_0 \subseteq \mathcal{K}(\mathbf{x}^0, \mathbf{y}^0)$ be a (also non-empty) minimizer for $\mathcal{C}_P^p(\mathbf{x}^0, \mathbf{y}^0)$ and let $\sigma : \{1, \dots, K\} \rightarrow \{1, \dots, |\mathbf{x}^0|\}$ be an optimal matching between \mathbf{z} and \mathbf{x}^0 .

We iteratively use assumptions A1 and A4 to define feasible solutions

$$\tilde{G}_k \subseteq \mathcal{K}\left(\mathbf{x}^0 \cup \bigcup_{i=1}^k \mathbf{x}^i, \mathbf{y}^0 \cup \bigcup_{i=1}^k \mathbf{y}^i\right).$$

We begin by letting $\tilde{G}_0 = G_0$. For $k = 1, \dots, K$, having already defined \tilde{G}_{k-1} , if $n_k = 0$, then we simply let $\tilde{G}_k = \tilde{G}_{k-1}$. Otherwise, we obtain a feasible solution \tilde{G}_k by gluing G_k with \tilde{G}_{k-1} at the vertices x^k, x_k^0 and adding up to c_{A4} edges from $\mathcal{K}(\mathbf{x}^k, \mathbf{y}^k)$. The fact that we can glue at any such pair of vertices is due to assumption A1: up to isomorphisms we can assume that x^k corresponds to the abstract graph vertex $(1, 1)$ and that x_k^0 to $(1, n_k + 1)$.

This construction gives the following inequality between the graph weights, if $n_k \neq 0$:

$$\begin{aligned} w(\tilde{G}_k) - w(\tilde{G}_{k-1}) - w(G_k) &\leq c_{A4} \text{diam}(\Omega_k)^p \\ &\quad + \sum_{y \in \mathcal{N}_{G_k}(x^k)} d(x_{\sigma(k)}^0, y)^p + \sum_{y \in \mathcal{N}_{\tilde{G}_{k-1}}(x_{\sigma(k)}^0)} d(x^k, y)^p, \end{aligned} \quad (3.8)$$

while if $n_k = 0$, we simply have $w(\tilde{G}_k) = w(\tilde{G}_{k-1})$. We bound from above the last two terms in (3.8) as follows: first,

$$\begin{aligned} \sum_{y \in \mathcal{N}_{G_k}(x^k)} d(x_{\sigma(k)}^0, y)^p &\lesssim \sum_{y \in \mathcal{N}_{G_k}(x^k)} d(x_{\sigma(k)}^0, z_k)^p + d(z_k, x^k)^p + d(x^k, y)^p \\ &\lesssim d(x_{\sigma(k)}^0, z_k)^p + d(z_k, x^k)^p + \sum_{y \in \mathcal{N}_{G_k}(x^k)} d(x^k, y)^p \\ &\lesssim d(x_{\sigma(k)}^0, z_k)^p + \text{diam}(\Omega_k)^p, \end{aligned}$$

where we used that $\deg_{x^k}(G_k) \leq c_{A3}$. To bound the last term, we notice that each step in the construction we are locally merging at different points in \mathbf{x}^0 : since no such points are adjacent because the graph is bipartite, using (3.6) by induction yields

$$\mathcal{N}_{\tilde{G}_{k-1}}(x_{\sigma(k)}^0) = \mathcal{N}_{G_0}(x_{\sigma(k)}^0),$$

which in particular contains at most c_{A3} elements, since G_0 is feasible. Therefore,

$$\begin{aligned} \sum_{y \in \mathcal{N}_{\tilde{G}_{k-1}}(x_{\sigma(k)}^0)} d(x^k, y)^p &= \sum_{y \in \mathcal{N}_{G_0}(x_{\sigma(k)}^0)} d(x^k, y)^p \\ &\lesssim \sum_{y \in \mathcal{N}_{G_0}(x_{\sigma(k)}^0)} d(x^k, z_k)^p + d(z_k, x_{\sigma(k)}^0)^p + d(x_{\sigma(k)}^0, y)^p \\ &\lesssim \text{diam}(\Omega_k)^p + d(z_k, x_{\sigma(k)}^0)^p + \sum_{y \in \mathcal{N}_{G_0}(x_{\sigma(k)}^0)} d(x_{\sigma(k)}^0, y)^p. \end{aligned}$$

Summing (3.8) upon $k = 1, \dots, K$, we obtain (3.7) because

$$\sum_{k=1}^K d(z_k, x_{\sigma(k)}^0)^p = M^p(\mathbf{z}, \mathbf{x}^0)$$

and, being all the points $x_{\sigma(k)}^0$ different,

$$\sum_{k=1}^K \sum_{y \in \mathcal{N}_{G_0}(x_{\sigma(k)}^0)} d(x_{\sigma(k)}^0, y)^p \leq \sum_{x \in \mathbf{x}^0} \sum_{y \in \mathcal{N}_{G_0}(x)} d(x, y)^p = \mathcal{C}_P^p(\mathbf{x}^0, \mathbf{y}^0). \quad \square$$

3.6. Growth/regularity. The last assumption that we introduce for a combinatorial optimization problem P over bipartite graphs is a general upper bound for the cost when specialized to a geometric graph in the Euclidean cube $(0, 1)^d$:

A5 (growth/regularity) There exists $c_{A5} \geq 0$ such that, for every $\mathbf{x}, \mathbf{y} \subseteq (0, 1)^d$, we have

$$\mathcal{C}_P^p(\mathbf{x}, \mathbf{y}) \leq c_{A5} \left(\min \left\{ |\mathbf{x}|^{1-\frac{p}{d}}, |\mathbf{y}|^{1-\frac{p}{d}} \right\} + M^p(\mathbf{x}, \mathbf{y}) \right). \quad (3.9)$$

Remark 3.9. Notice that if $\Omega \subset (0, 1)^d$ then (3.9) applies in particular for $\mathbf{x}, \mathbf{y} \subseteq \Omega$. By scaling we obtain that for every bounded set Ω and every $\mathbf{x}, \mathbf{y} \subseteq \Omega$,

$$\mathcal{C}_P^p(\mathbf{x}, \mathbf{y}) \leq c_{A5} \left(\text{diam}(\Omega)^p \min \left\{ |\mathbf{x}|^{1-\frac{p}{d}}, |\mathbf{y}|^{1-\frac{p}{d}} \right\} + M^p(\mathbf{x}, \mathbf{y}) \right).$$

Using (3.1), we obtain at once that in order to establish that a given problem P satisfies (3.9) it is enough to consider the case where $\mathbf{x}, \mathbf{y} \subseteq (0, 1)^d$ have the same number of elements.

Notice that this assumption seems slightly different with respect to the previous ones, as it explicitly refers to the cost for Euclidean realizations of the graph, instead of feasible solutions, and relies as well on the assignment problem. In fact, the constant c_{A5} depends upon the problem P but also on the dimension d and the exponent p , which however will be fixed in our derivations so we avoid to explicitly state it.

It is well known that quite general arguments, such as the space-filling curve heuristics [43, Chapter 2], lead to an upper bound in terms of $n^{1-p/d}$ for non-bipartite combinatorial optimization problems over n points in a cube, under very mild assumptions, including those introduced above. Simple examples show that similar bounds cannot hold for their bipartite counterparts, which explains the second term in the right-hand side of (3.9).

To establish it in our examples we follow the strategy from [13], where limit results for the random Euclidean bipartite TSP for $p = d = 2$ were first obtained.

Lemma 3.10. *The TSP, the connected κ -factor problem (as well as the non-connected one) and the κ -MST problems over complete bipartite graphs satisfy assumption A5 (with a constant c_{A5} depending on κ, p, d only).*

Proof. Let us first observe that the cost of the κ -MST problem is always bounded from above by the cost of the minimum weight connected κ -factor problem, since given any connected κ -factor, one can extract from it a MST whose degree at every vertex is then bounded by κ . Therefore it is sufficient to check that assumption A5 holds with P being the connected κ -factor problem, for any $\kappa \geq 2$ (the case $\kappa = 2$ being the TSP).

For (Ω, d) a general metric space and $\mathbf{x}, \mathbf{y} \subseteq \Omega$ we establish first the bound

$$\mathcal{C}_P^p(\mathbf{x}, \mathbf{y}) \lesssim \mathcal{C}_{\text{TSP}}^p(\mathbf{x}) + M^p(\mathbf{x}, \mathbf{y}). \quad (3.10)$$

Combining this with the fact that when (Ω, d) is the unit cube $(0, 1)^d$ with the Euclidean distance, $\mathcal{C}_{\text{TSP}}^p(\mathbf{x}) \lesssim |\mathbf{x}|^{1-p/d}$ (a well-known fact, proved e.g. via space-filling curves) this would conclude the proof of (3.9).

Assume without loss of generality that $|\mathbf{x}| = |\mathbf{y}| = n \geq \kappa$ and let ρ be a permutation over $[n]$ that induces an optimal assignment between \mathbf{x} and \mathbf{y} . Consider then an optimizer for

the TSP over $\mathcal{K}(\mathbf{x})$, which we also identify with a permutation σ over $[n]$. We then define the feasible solution $G \in \mathcal{F}_{n,n}$ for the connected κ -factor problem whose edge set is

$$E_G = \{(1, \sigma(i)), (2, \rho(\sigma(i+\ell)))\} : i \in [n], \ell \in \{0, 1, \dots, \kappa-1\},$$

which generalizes $E_{\sigma,\tau}$ from (3.5) with $\tau = \rho \circ \sigma$ in the $\kappa = 2$ case, and as in (3.5) we use the summation mod n , i.e., $i + \ell = i + \ell - n$ if $i + \ell > n$. Clearly, any vertex has degree κ and the graph is connected, since $E_G \supset E_{\sigma,\tau}$.

It follows that

$$C_P^p(\mathbf{x}, \mathbf{y}) \leq \sum_{i=1}^n \sum_{\ell=0}^{\alpha-1} d(x_{\sigma(i)}, y_{\rho(\sigma(i+\ell))})^p.$$

Using the triangle inequality for every i and ℓ , we bound from above

$$d(x_{\sigma(i)}, y_{\rho(\sigma(i+\ell))})^p \lesssim \sum_{j=0}^{\ell-1} d(x_{\sigma(i+j)}, x_{\sigma(i+j-1)})^p + d(x_{\sigma(i+\ell)}, y_{\rho(\sigma(i+\ell))})^p.$$

Summation upon i (keeping ℓ fixed) gives

$$\sum_{j=0}^{\ell-1} \sum_{i=1}^n d(x_{\sigma(i+j)}, x_{\sigma(i+j-1)})^p + \sum_{i=1}^n d(x_{\sigma(i+\ell)}, y_{\rho(\sigma(i+\ell))})^p \lesssim \ell \text{TSP}^p(\mathbf{x}) + M^p(\mathbf{x}, \mathbf{y}),$$

hence, after summing upon $\ell = 0, \dots, \kappa-1$, we obtain (3.10). \square

4. CONVERGENCE RESULTS FOR POISSON POINT PROCESSES

4.1. Point processes. We define a point process on \mathbb{R}^d as a random finite family of points $\mathcal{N} = (X_i)_{i=1}^N \subseteq \mathbb{R}^d$, i.e. a N -uple of random variables with values in \mathbb{R}^d , where the total number of points N is also random and a.s. finite (if $N = 0$, then $\mathcal{N} = \emptyset$). We extend the notation for families of points to point processes (naturally defined for each realization of the random variables): for a process $\mathcal{N} = (X_i)_{i=1}^N$, write $\mu^{\mathcal{N}} := \sum_{i=1}^N \delta_{X_i}$ and, given a Borel $\Omega \subseteq \mathbb{R}^d$, let $\mathcal{N}(\Omega) = \mu^{\mathcal{N}}(\Omega)$ be the (random) number of variables belonging to Ω , while \mathcal{N}_Ω denotes its restriction to Ω , i.e., the collection of the variables such that $X_i \in \Omega$ (naturally re-indexed over $i = 1, \dots, \mathcal{N}(\Omega)$, with the order inherited from the original process). Given two point processes $\mathcal{N} = (X_i)_{i=1}^N$, $\mathcal{M} = (Y_j)_{j=1}^M$, their union is $\mathcal{N} \cup \mathcal{M} = (X_1, \dots, X_N, Y_1, \dots, Y_M)$.

Given a finite Borel measure λ on \mathbb{R}^d , a Poisson point process \mathcal{N}^λ with intensity λ can be constructed from a random collection of i.i.d. variables $(X_i)_{i=1}^\infty$ with common law $\lambda/\lambda(\mathbb{R}^d)$ and, after introducing a further independent Poisson variable N^λ with mean $\lambda(\mathbb{R}^d)$, by considering only the first N^λ variables, i.e.,

$$\mathcal{N}^\lambda := (X_i)_{i=1}^{N^\lambda}.$$

A key property of a Poisson point process (with intensity λ) is that, given any countable Borel partition $\mathbb{R}^d = \cup_k \Omega_k$, the variables $(\mathcal{N}^\lambda(\Omega_k))_k$ are independent Poisson variables, each with mean $\lambda(\Omega_k)$ and, conditionally upon their value, the points in each Ω_k are i.i.d. variables with common probability law $\lambda \llcorner \Omega_k / \lambda(\Omega_k)$. This property can be summarized by stating that the restrictions $(\mathcal{N}_{\Omega_k}^\lambda)_k$ are independent Poisson point processes, with each $\mathcal{N}_{\Omega_k}^\lambda$ having intensity given by the restriction $\lambda \llcorner \Omega_k$.

We will use the well-known thinning operation, which apparently dates back to Rényi [38], to split a Poisson point process \mathcal{N}^λ with intensity λ into two independent Poisson point processes, each containing approximately a given fraction of points: for $\eta \in [0, 1]$, the η -thinning of a Poisson point process $\mathcal{N}^\lambda = (X_i)_{i=1}^{N^\lambda}$ defines the two processes

$$\mathcal{N}^{(1-\eta)\lambda} = (X_i)_{i=1}^{N^{(1-\eta)\lambda}} \quad \text{and} \quad \mathcal{N}^{\eta\lambda} = (X_{N^{(1-\eta)\lambda}+i})_{i=1}^{N^\lambda - N^{(1-\eta)\lambda}},$$

where $N^{(1-\eta)\lambda} = \sum_{i=1}^N Z_i$ is defined using a further sequence of i.i.d. Bernoulli random variables $(Z_i)_{i=1}^\infty$ with $\mathbb{P}(Z_i = 1) = 1 - \eta$ (independent from the variables $(X_i)_i$ and N^λ). Clearly, $\mathcal{N}^\lambda = \mathcal{N}^{(1-\eta)\lambda} \cup \mathcal{N}^{\eta\lambda}$, and it is straightforward to prove that both are independent and Poisson point processes with intensities respectively $(1 - \eta)\lambda$ and $\eta\lambda$.

4.2. Statement. The aim of this section is to prove the analogue of Theorem 1.1 for Poisson point processes (instead of i.i.d. points).

Theorem 4.1. *Let $d \geq 3$, $p \in [1, d)$ and let $\mathbf{P} = (\mathcal{F}_{n,n})_{n \in \mathbb{N}}$ be a combinatorial optimization problem over complete bipartite graphs such that assumptions A1, A2, A3, A4 and A5 hold. Then, there exists $\beta_{\mathbf{P}} \in (0, \infty)$ (depending on p and d) such that the following holds.*

Let $\Omega \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary and such that (2.8) holds. Let ρ be a Hölder continuous probability density on Ω , uniformly strictly positive and bounded from above. For every $n \in (0, \infty)$, let $\mathcal{N}^{n\rho}$, $\mathcal{M}^{n\rho}$ be independent Poisson point processes with intensity $n\rho$ on Ω . Then,

$$\limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_{\mathbf{P}}^p(\mathcal{N}^{n\rho}, \mathcal{M}^{n\rho})] \leq \beta_{\mathbf{P}} \int_{\Omega} \rho^{1-\frac{p}{d}}. \quad (4.1)$$

Moreover, if ρ is the uniform density and Ω is a cube or its boundary is C^2 , then the limit exists and equals the right-hand side.

After having introduced some general notation and proved some basic facts, we split the proof into four main cases. We deal first with the case of a uniform density on a cube and establish existence of the limit via subadditivity. Then, we consider Hölder densities on a cube and move next to general domains. Finally, we establish existence of the limit for uniform densities on domains with C^2 boundary.

4.3. General facts. Although each case has its distinctive features, the underlying strategy is common and relies on Proposition 3.7 in combination with a preliminary application of the thinning operation. To avoid repetitions and introduce a general notation, we give a description of the construction and show a first lemma which uses the fundamental ideas upon which we elaborate in the next sections.

Let \mathcal{N} , \mathcal{M} be two independent Poisson point processes on Ω with common intensity given by a finite measure λ . In our applications, λ is Lebesgue measure or $\lambda = n\rho$, but for simplicity here we omit to specify it. We apply the η -thinning to $\mathcal{N} = \mathcal{N}^{1-\eta} \cup \mathcal{N}^\eta$, obtaining independent Poisson point processes with respective intensities $(1 - \eta)\lambda$, $\eta\lambda$, and similarly to $\mathcal{M} = \mathcal{M}^{1-\eta} \cup \mathcal{M}^\eta$. Given a finite Borel partition $\Omega = \bigcup_{k=1}^K \Omega_k$, for each $k = 1, \dots, K$, we pick a minimizer $G_k \subseteq \mathcal{K}(\mathcal{N}_{\Omega_k}^{1-\eta}, \mathcal{M}_{\Omega_k}^{1-\eta})$ for the problem

$$\mathcal{C}_{\mathbf{P}}^p(\mathcal{N}_{\Omega_k}^{1-\eta}, \mathcal{M}_{\Omega_k}^{1-\eta}).$$

Writing

$$Z_k = \min \left\{ |\mathcal{N}_{\Omega_k}^{1-\eta}|, |\mathcal{M}_{\Omega_k}^{1-\eta}| \right\},$$

we notice that $G_k = \emptyset$ if and only if $Z_k < c_{A2}$ (by Remark 5.5 for $p > 1$, G_k is a.s. unique. For $p = 1$ we can consider a measurable selection).

We define point processes \mathcal{U} , \mathcal{V} on Ω by setting $\mathcal{U}_{\Omega_k} \subseteq \mathcal{N}_{\Omega_k}^{1-\eta}$, $\mathcal{V}_{\Omega_k} \subseteq \mathcal{M}_{\Omega_k}^{1-\eta}$, given by all the points, respectively in $\mathcal{N}_{\Omega_k}^{1-\eta}$ and $\mathcal{M}_{\Omega_k}^{1-\eta}$, which do not belong to the set of vertices of G_k . In particular, if $G_k = \emptyset$, then $\mathcal{U}_{\Omega_k} = \mathcal{N}_{\Omega_k}^{1-\eta}$, $\mathcal{V}_{\Omega_k} = \mathcal{M}_{\Omega_k}^{1-\eta}$. Notice that by construction the K pairs of processes $((\mathcal{U}_{\Omega_k}, \mathcal{V}_{\Omega_k}))_{k=1}^K$ are independent, but for any k the two processes \mathcal{U}_{Ω_k} , \mathcal{V}_{Ω_k} are not in general independent. For later use, we prove:

Lemma 4.2. *For every $k = 1, \dots, K$ such that*

$$\lambda(\Omega_k) > 4c_{A2}, \quad (4.2)$$

we have, for every $q \geq 1$,

$$\mathbb{E} [|\mathcal{U}_{\Omega_k}|^q + |\mathcal{V}_{\Omega_k}|^q] \lesssim_q \lambda(\Omega_k)^{\frac{q}{2}}. \quad (4.3)$$

Proof. In the event

$$A_k = \{Z_k \geq (1 - \eta)\lambda(\Omega_k)/2\},$$

since $\eta \in (0, 1/2)$ we have that $Z_k \geq \mathbf{c}_{A2}$ hence by assumption **A2**, every feasible solution (in particular the optimal solution G_k) spans a subgraph of $\mathcal{K}_{\mathcal{N}^{1-\eta}(\Omega_k), \mathcal{M}^{1-\eta}(\Omega_k)}$ isomorphic to \mathcal{K}_{Z_k, Z_k} , so that

$$|\mathcal{U}_{\Omega_k}| \leq |\mathcal{N}_{\Omega_k}^{1-\eta}| - Z_k \leq \left| |\mathcal{M}_{\Omega_k}^{1-\eta}| - |\mathcal{N}_{\Omega_k}^{1-\eta}| \right|.$$

Using (2.20), we have

$$\mathbb{E} [|\mathcal{U}_{\Omega_k}|^q I_{A_k}] \leq \mathbb{E} \left[\left| |\mathcal{M}_{\Omega_k}^{1-\eta}| - |\mathcal{N}_{\Omega_k}^{1-\eta}| \right|^q \right] \lesssim_q \lambda(Q_L)^{\frac{q}{2}}.$$

By the union bound and (2.21) with $n = (1 - \eta)\lambda(\Omega_k)$, $\gamma = 1/2$, we have

$$\begin{aligned} \mathbb{P}(A_k^c) &\leq \mathbb{P}(|\mathcal{N}_{\Omega_k}^{1-\eta}| < (1 - \eta)\lambda(\Omega_k)/2) + \mathbb{P}(|\mathcal{M}_{\Omega_k}^{1-\eta}| < (1 - \eta)\lambda(\Omega_k)/2) \\ &\lesssim_q \lambda(\Omega_k)^{-q}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{E} [|\mathcal{U}_{\Omega_k}|^q I_{A_k^c}] &\leq \mathbb{E} \left[|\mathcal{N}_{\Omega_k}^{1-\eta}|^q I_{A_k^c} \right] \leq \mathbb{E} \left[|\mathcal{N}_{\Omega_k}^{1-\eta}|^{2q} \right]^{\frac{1}{2}} \mathbb{P}(A_k^c)^{\frac{1}{2}} \\ &\lesssim_q \lambda(\Omega_k)^{\frac{q}{2}}. \end{aligned}$$

Arguing similarly for $|\mathcal{V}_{\Omega_k}|$, we obtain (4.3). \square

For $k = 1, \dots, K$, we define

$$\mathbf{x}^k = \mathcal{N}_{\Omega_k}^{1-\eta} \setminus \mathcal{U}_{\Omega_k}, \quad \mathbf{y}^k = \mathcal{M}_{\Omega_k}^{1-\eta} \setminus \mathcal{V}_{\Omega_k},$$

so that by construction $|\mathbf{x}^k| = |\mathbf{y}^k| = n_k$, with

$$n_k = \begin{cases} Z_k & \text{if } Z_k \geq \mathbf{c}_{A2}, \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, since the optimizer G_k is a feasible solution in $\mathcal{K}(\mathbf{x}^k, \mathbf{y}^k)$, we have

$$\mathcal{C}_P^p(\mathcal{N}_{\Omega_k}^{1-\eta}, \mathcal{M}_{\Omega_k}^{1-\eta}) = \mathcal{C}_P^p(\mathbf{x}^k, \mathbf{y}^k).$$

We then let $\mathbf{x}^0 = \mathcal{N}^\eta \cup \mathcal{U}$, $\mathbf{y}^0 = \mathcal{M}^\eta \cup \mathcal{V}$. In the event

$$\{\min\{|\mathcal{N}^\eta|, |\mathcal{M}^\eta|\} \geq \min\{K, \mathbf{c}_{A2}\}\}, \quad (4.4)$$

Proposition 3.7 applies for any choice of points $\mathbf{z} = (z_k)_{k=1}^K$ with $z_k \in \Omega_k$, yielding the inequality

$$\mathcal{C}_P^p(\mathcal{N}, \mathcal{M}) - \sum_{k=1}^K \mathcal{C}_P^p(\mathcal{N}_{\Omega_k}^{1-\eta}, \mathcal{M}_{\Omega_k}^{1-\eta}) \lesssim \mathcal{C}_P^p(\mathcal{N}^\eta \cup \mathcal{U}, \mathcal{M}^\eta \cup \mathcal{V}) + \mathbb{M}^p(\mathcal{N}^\eta \cup \mathcal{U}, \mathbf{z}) + \sum_{k=1}^K \text{diam}(\Omega_k)^p. \quad (4.5)$$

By Remark 3.8, if also

$$\left\{ \min_{k=1, \dots, K} \min \left\{ |\mathcal{N}_{\Omega_k}^\eta|, |\mathcal{M}_{\Omega_k}^\eta| \right\} \geq 1 \right\} \quad (4.6)$$

then the term $\mathbb{M}^p(\mathcal{N}^\eta \cup \mathcal{U}, \mathbf{z})$ can be removed in (4.5).

Once (4.5) is established, the next step is to take expectation and carefully estimate the ‘‘error terms’’ in the right-hand side. To convey the main ideas, we start with the simplest case when K is kept fixed as we let $n \rightarrow \infty$ in the intensity of the process $\lambda = n\rho$.

Lemma 4.3. *With the notation and assumptions of Theorem 4.1, fix $K \in \mathbb{N}$ and consider a Borel partition $\Omega = \bigcup_{k=1}^K \Omega_k$. Then,*

$$\limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}^{n\rho}, \mathcal{M}^{n\rho})] \leq \sum_{k=1}^K \limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}_{\Omega_k}^{n\rho}, \mathcal{M}_{\Omega_k}^{n\rho})]. \quad (4.7)$$

Proof. We can assume that each Ω_k is not negligible. Then, condition (4.2) with $\lambda = n\rho$ holds if n is sufficiently large. Letting

$$A = \bigcap_{k=1}^K \left\{ \min \left\{ |\mathcal{N}_{\Omega_k}^{n\rho}|, |\mathcal{M}_{\Omega_k}^{n\rho}| \right\} \geq c_{A2} \right\} = \bigcap_{k=1}^K A_k, \quad (4.8)$$

we have that both (4.4) and (4.6) hold A .

By the union bound in combination with (2.21), we estimate, for every $q \geq 1$,

$$\mathbb{P}(A^c) \leq \sum_{k=1}^K \mathbb{P}(A_k^c) \lesssim_{q,\eta,K} n^{-q}. \quad (4.9)$$

Combined with the trivial inequality $\mathcal{C}_P^p(\mathcal{N}^{n\rho}, \mathcal{M}^{n\rho}) \lesssim |\mathcal{N}^{n\rho}|$ we obtain that

$$\mathbb{E} [\mathcal{C}_P^p(\mathcal{N}^{n\rho}, \mathcal{M}^{n\rho}) I_{A^c}] \leq \mathbb{E} [|\mathcal{N}^{n\rho}|^2]^{\frac{1}{2}} \mathbb{P}(A^c)^{\frac{1}{2}} \lesssim_{q,K} n^{\frac{1-q}{2}},$$

which is infinitesimal if $q > 1$ (even without dividing by $n^{1-p/d}$). Therefore,

$$\limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}^{n\rho}, \mathcal{M}^{n\rho}) I_A] = \limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}^{n\rho}, \mathcal{M}^{n\rho})]$$

and we only need to prove the following inequality, for fixed η ,

$$\limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}^{n\rho}, \mathcal{M}^{n\rho}) I_A] - \sum_{k=1}^K \limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}_{\Omega_k}^{n\rho}, \mathcal{M}_{\Omega_k}^{n\rho})] \lesssim_K \eta^{1-\frac{p}{d}},$$

and finally let $\eta \rightarrow 0$ to obtain the thesis. To this aim, we multiply (4.5) by I_A and take expectation, obtaining the inequality

$$\begin{aligned} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}^{n\rho}, \mathcal{M}^{n\rho}) I_A] - \sum_{k=1}^K \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}_{\Omega_k}^{(1-\eta)n\rho}, \mathcal{M}_{\Omega_k}^{(1-\eta)n\rho})] \\ \lesssim \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}^{\eta n\rho} \cup \mathcal{U}, \mathcal{M}^{\eta n\rho} \cup \mathcal{V})] + K. \end{aligned} \quad (4.10)$$

Since, for each $k = 1, \dots, K$,

$$\begin{aligned} \limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}_{\Omega_k}^{(1-\eta)n\rho}, \mathcal{M}_{\Omega_k}^{(1-\eta)n\rho})] \\ = (1-\eta)^{1-\frac{p}{d}} \limsup_{n \rightarrow \infty} ((1-\eta)n)^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}_{\Omega_k}^{(1-\eta)n\rho}, \mathcal{M}_{\Omega_k}^{(1-\eta)n\rho})] \\ \leq \limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}_{\Omega_k}^{n\rho}, \mathcal{M}_{\Omega_k}^{n\rho})], \end{aligned}$$

we need to focus only on the terms in the right-hand side of (4.10). Since the last term is constant, we are left with the proof of

$$\limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}^{\eta n\rho} \cup \mathcal{U}, \mathcal{M}^{\eta n\rho} \cup \mathcal{V})] \lesssim \eta^{1-\frac{p}{d}}. \quad (4.11)$$

We first notice that by (4.3) and Hölder inequality we have for every $q \geq 1$,

$$\mathbb{E} [|\mathcal{U}|^q + |\mathcal{V}|^q] \lesssim_q K^{\frac{q}{2}} n^{\frac{q}{2}}. \quad (4.12)$$

We now use assumption A5 so that

$$\mathbb{E} [\mathcal{C}_P^p(\mathcal{N}^{\eta n\rho} \cup \mathcal{U}, \mathcal{M}^{\eta n\rho} \cup \mathcal{V})] \lesssim \mathbb{E} [|\mathcal{N}^{\eta n\rho} \cup \mathcal{U}|^{1-\frac{p}{d}}] + \mathbb{E} [\mathcal{M}^p(\mathcal{N}^{\eta n\rho} \cup \mathcal{U}, \mathcal{M}^{\eta n\rho} \cup \mathcal{V})]. \quad (4.13)$$

To estimate the first term in the right-hand side, we use Hölder inequality and (4.12) with $q = 1$,

$$\begin{aligned} \mathbb{E} \left[|\mathcal{N}^{\eta\rho} \cup \mathcal{U}|^{1-\frac{p}{d}} \right] &\lesssim \mathbb{E} [|\mathcal{N}^{\eta\rho}|]^{1-\frac{p}{d}} + \mathbb{E} [|\mathcal{U}|]^{1-\frac{p}{d}} \\ &\lesssim n^{1-\frac{p}{d}} \left(\eta^{1-\frac{p}{d}} + C_K n^{-\frac{1}{2}(1-\frac{p}{d})} \right). \end{aligned} \quad (4.14)$$

For the second term, thanks to (4.12) we may use Proposition 6.3 with $H = n^{1/2}$ and $h = \min \{ \mathbb{E} [|\mathcal{N}^{\eta\rho}|], \mathbb{E} [|\mathcal{M}^{\eta\rho}|] \} \sim n\eta$ so that for some $\alpha < 2$ and $\beta > 0$

$$\mathbb{E} [M^p(\mathcal{N}^{\eta\rho} \cup \mathcal{U}, \mathcal{M}^{\eta\rho} \cup \mathcal{V})] \lesssim n^{1-\frac{p}{d}} \left(\eta^{1-\frac{p}{d}} + C_{K,\eta} n^{-\frac{\beta}{2}(2-\alpha)} \right).$$

Plugging this and (4.14) in (4.13) concludes the proof of (4.11). \square

Remark 4.4. We notice that the proof above yields also the inequality

$$\begin{aligned} \liminf_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_{\mathbb{P}}^p(\mathcal{N}^{n\rho}, \mathcal{M}^{n\rho})] &\leq \liminf_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p \left(\mathcal{N}_{\Omega_1}^{n\rho}, \mathcal{M}_{\Omega_1}^{n\rho} \right) \right] \\ &\quad + \sum_{k=2}^K \limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p \left(\mathcal{N}_{\Omega_k}^{n\rho}, \mathcal{M}_{\Omega_k}^{n\rho} \right) \right]. \end{aligned} \quad (4.15)$$

This follows by repeating the argument only along a subsequence $n_\ell \rightarrow \infty$ such that

$$\lim_{\ell \rightarrow \infty} n_\ell^{\frac{p}{d}-1} \mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p \left(\mathcal{N}_{\Omega_1}^{n_\ell\rho}, \mathcal{M}_{\Omega_1}^{n_\ell\rho} \right) \right] = \liminf_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p \left(\mathcal{N}_{\Omega_1}^{n\rho}, \mathcal{M}_{\Omega_1}^{n\rho} \right) \right].$$

4.4. Uniform density on a cube. In this section we consider the case of a uniform measure on a cube. Up to rescaling (see (4.18)) it is equivalent to consider two independent Poisson point processes \mathcal{N}_{Q_L} and \mathcal{M}_{Q_L} with intensity one on Q_L and prove that

$$f(L) = \frac{1}{|Q_L|} \mathbb{E} [\mathcal{C}_{\mathbb{P}}^p(\mathcal{N}_{Q_L}, \mathcal{M}_{Q_L})]$$

has a limit as $L \rightarrow \infty$.

Proposition 4.5. *Let $d \geq 3$, $p \in [1, d)$ and let $\mathbb{P} = (\mathcal{F}_{n,n})_{n \in \mathbb{N}}$ be a combinatorial optimization problem over complete bipartite graphs such that assumptions A1, A2, A3, A4 and A5 hold. Then, there exists $\beta_{\mathbb{P}} \in (0, \infty)$ (depending on p and d) such that*

$$\lim_{L \rightarrow \infty} f(L) = \beta_{\mathbb{P}}. \quad (4.16)$$

Proof. We split the proof into several steps. In the first two steps we establish basic properties of f , before moving to the main argument. This follows the strategy of the previous section and ultimately relies upon an application of Lemma 2.12.

Step 1. Continuity and upper bound. Writing $z = \min \{n, m\}$, we first notice that by Assumption A5 and (6.2) of Proposition 6.1,

$$\mathbb{E} [\mathcal{C}_{\mathbb{P}}^p((X_i)_{i=1}^n, (Y_j)_{j=1}^m)] \lesssim z^{1-\frac{p}{d}} + \mathbb{E} [M^p((X_i)_{i=1}^n, (Y_j)_{j=1}^m)] \lesssim z^{1-\frac{p}{d}}.$$

This proves on the one hand that f is bounded from above as

$$f(L) \lesssim L^{p-d} \mathbb{E} \left[\min \{ |\mathcal{N}_{Q_L}|, |\mathcal{M}_{Q_L}| \}^{1-\frac{p}{d}} \right] \lesssim L^{p-d} \mathbb{E} [|\mathcal{N}_{Q_L}|]^{1-\frac{p}{d}} \lesssim 1. \quad (4.17)$$

On the other hand, combining it with dominated convergence it also gives continuity of f thanks to the representation formula

$$\begin{aligned} f(L) &= \sum_{n,m=0}^{\infty} \frac{1}{L^d} \mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p(\mathcal{N}_{Q_L}, \mathcal{M}_{Q_L}) \mid |\mathcal{N}_{Q_L}| = n, |\mathcal{M}_{Q_L}| = m \right] e^{-2L^d} \frac{(L^d)^{n+m}}{n!m!} \\ &= L^{p-d} \sum_{n,m=0}^{\infty} \mathbb{E} [\mathcal{C}_{\mathbb{P}}^p((X_i)_{i=1}^n, (Y_j)_{j=1}^m)] e^{-2L^d} \frac{(L^d)^{n+m}}{n!m!} \end{aligned}$$

where $(X_i)_{i=1}^n, (Y_j)_{j=1}^m$ are i.i.d. points on Q_L . We also notice that by a simple scaling argument, if $\mathcal{N}^\lambda, \mathcal{M}^\lambda$ are independent Poisson processes of intensity $\lambda > 0$ on Q_L then

$$\frac{1}{|Q_L|} \mathbb{E} \left[\mathcal{C}_P^p \left(\mathcal{N}_{Q_L}^\lambda, \mathcal{M}_{Q_L}^\lambda \right) \right] = \frac{\lambda^{1-\frac{p}{d}}}{|Q_{\lambda L}|} \mathbb{E} \left[\mathcal{C}_P^p \left(\mathcal{N}_{\frac{1}{\lambda} Q_L}^1, \mathcal{M}_{\frac{1}{\lambda} Q_L}^1 \right) \right] = \lambda^{1-\frac{p}{d}} f(\lambda^{\frac{1}{d}} L). \quad (4.18)$$

Combined with (4.17), it yields that for any cube Q and $\lambda > 0$,

$$\mathbb{E} \left[\mathcal{C}_P^p \left(\mathcal{N}_Q^\lambda, \mathcal{M}_Q^\lambda \right) \right] \lesssim |Q| \lambda^{1-\frac{p}{d}}. \quad (4.19)$$

Step 2. Lower bound. The spanning assumption A2 yields that, if e.g. $c_{A2} \leq n \leq m$, then

$$\mathcal{C}_P^p \left((X_i)_{i=1}^n, (Y_j)_{j=1}^m \right) \geq \sum_{i=1}^n \min_{j=1, \dots, m} |X_i - Y_j|^p.$$

The following classical lower bound, e.g. proved in [43, Chapter 2],

$$\mathbb{E} \left[\min_{j=1, \dots, m} |X_i - Y_j|^p \right] \gtrsim m^{-\frac{p}{d}}$$

entails that

$$\mathcal{C}_P^p \left((X_i)_{i=1}^n, (Y_j)_{j=1}^m \right) \gtrsim m^{-\frac{p}{d}} \cdot n.$$

Writing $Z = \min \{ |\mathcal{N}_{Q_L}|, |\mathcal{M}_{Q_L}| \}$, we deduce that

$$f(L) \gtrsim L^{p-d} \mathbb{E} \left[\max \{ |\mathcal{N}_{Q_L}|, |\mathcal{M}_{Q_L}| \}^{-\frac{p}{d}} Z I_{\{Z \geq c_{A2}\}} \right].$$

Let

$$A = \{ |Q_L|/2 \leq Z \leq \max \{ |\mathcal{N}_{Q_L}|, |\mathcal{M}_{Q_L}| \} \leq 3|Q_L|/2 \}.$$

By (2.21) with $\eta = 1/2$, we have $\mathbb{P}(A) \rightarrow 1$ as $L \rightarrow \infty$. Therefore $\text{id } L$ is large enough,

$$f(L) \gtrsim L^{p-d} \mathbb{E} \left[\max \{ |\mathcal{N}_{Q_L}|, |\mathcal{M}_{Q_L}| \}^{-\frac{p}{d}} Z I_A \right] \gtrsim L^{p-d} \mathbb{E} \left[L^{d-p} I_A \right] \gtrsim 1.$$

In the remaining steps we prove the following claim. There exists $\beta = \beta(p, d) > 0$ such that for every $\eta \in (0, 1/2)$, there exists $C(\eta) > 0$ such that, for every $m \in \mathbb{N}$, $m \geq 1$ and $L \geq C(\eta)$,

$$f(mL) - f((1-\eta)L) \lesssim \eta^{1-\frac{p}{d}} + C(\eta)L^{-\beta}. \quad (4.20)$$

This would conclude the proof of (4.16) by Lemma 2.12.

Step 3. Partitioning and exclusion of the event in which few points are sampled. Using the notation from Section 4.3, we partition $\Omega = Q_{mL}$ into $K = m^d$ cubes $Q_i = Q_L + Lz_i \subseteq Q_{mL}$ with $z_i \in \mathbb{Z}^d$ and two independent Poisson processes \mathcal{N}, \mathcal{M} of unit intensity on Q_{mL} .

We first reduce to the event

$$A = \left\{ \min \left\{ |\mathcal{N}_{Q_{mL}}^\eta|, |\mathcal{M}_{Q_{mL}}^\eta| \right\} \geq \eta |Q_{mL}|/2 \right\},$$

which contains (4.4) provided L is sufficiently large (depending on η only, not on m). We first argue that A^c is of small probability. Indeed, using a union bound we find that for every $q \geq 1$,

$$\mathbb{P}(A^c) \leq \mathbb{P} \left(|\mathcal{N}_{Q_{mL}}^\eta| < |Q_{mL}|/2 \right) + \mathbb{P} \left(|\mathcal{M}_{Q_{mL}}^\eta| < |Q_{mL}|/2 \right) \stackrel{(2.21)}{\lesssim_{\eta, q}} |Q_{mL}|^{-q}.$$

If A^c holds, we use the trivial bound that follows from Assumption A3:

$$\mathcal{C}_P^p(\mathcal{N}_{Q_{mL}}, \mathcal{M}_{Q_{mL}}) \lesssim |\mathcal{N}_{Q_{mL}}| |Q_{mL}|^{\frac{p}{d}},$$

so that for any given $\beta > 0$ and provided we choose q sufficiently large.

$$\begin{aligned} \frac{1}{|Q_{mL}|} \mathbb{E} \left[\mathcal{C}_P^p(\mathcal{N}_{Q_{mL}}, \mathcal{M}_{Q_{mL}}) I_{A^c} \right] &\lesssim |Q_{mL}|^{\frac{p}{d}-1} \mathbb{E} \left[|\mathcal{N}_{Q_{mL}}|^2 \right]^{\frac{1}{2}} \mathbb{P}(A^c)^{\frac{1}{2}} \\ &\lesssim_{\eta, q} |Q_{mL}|^{\frac{p}{d}-1} |Q_{mL}| \cdot |Q_{mL}|^{-q} \lesssim_\eta L^{-\beta}. \end{aligned} \quad (4.21)$$

If A holds, letting \mathbf{z} be the set of centres of the m^d cubes, inequality (4.5) reads

$$\begin{aligned} \mathcal{C}_{\mathbb{P}}^p(\mathcal{N}_{Q_{mL}}, \mathcal{M}_{Q_{mL}}) - \sum_{i=1}^{m^d} \mathcal{C}_{\mathbb{P}}^p(\mathcal{N}_{Q_i}^{1-\eta}, \mathcal{M}_{Q_i}^{1-\eta}) &\lesssim \mathcal{C}_{\mathbb{P}}^p(\mathcal{N}^\eta \cup \mathcal{U}, \mathcal{M}^\eta \cup \mathcal{V}) \\ &+ \mathbb{M}^p(\mathcal{N}^\eta \cup \mathcal{U}, \mathbf{z}) + m^d L^p. \end{aligned} \quad (4.22)$$

Notice that by the properties of the Poisson point process, the law of $\mathcal{C}_{\mathbb{P}}^p(\mathcal{N}_{Q_i}^{1-\eta}, \mathcal{M}_{Q_i}^{1-\eta})$ equals that of $\mathcal{C}_{\mathbb{P}}^p(\mathcal{N}_{Q_L}^{1-\eta}, \mathcal{M}_{Q_L}^{1-\eta})$. In particular

$$\begin{aligned} \frac{1}{|Q_L|} \mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p(\mathcal{N}_{Q_i}^{1-\eta}, \mathcal{M}_{Q_i}^{1-\eta}) I_A \right] &\leq \frac{1}{|Q_L|} \mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p(\mathcal{N}_{Q_i}^{1-\eta}, \mathcal{M}_{Q_i}^{1-\eta}) \right] \\ &= \frac{1}{|Q_L|} \mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p(\mathcal{N}_{Q_L}^{1-\eta}, \mathcal{M}_{Q_L}^{1-\eta}) \right] \\ &\stackrel{(4.18)}{=} (1-\eta)^{1-\frac{p}{d}} f((1-\eta)^{\frac{1}{d}} L) \leq f((1-\eta)^{\frac{1}{d}} L). \end{aligned}$$

We thus obtain from (4.22),

$$\begin{aligned} \frac{1}{|Q_{mL}|} \mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p(\mathcal{N}_{Q_{mL}}, \mathcal{M}_{Q_{mL}}) I_A \right] - f((1-\eta)^{\frac{1}{d}} L) &\lesssim \frac{1}{|Q_{mL}|} \mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p(\mathcal{N}^\eta \cup \mathcal{U}, \mathcal{M}^\eta \cup \mathcal{V}) \right] \\ &+ \frac{1}{|Q_{mL}|} \mathbb{E} \left[\mathbb{M}^p(\mathcal{N}^\eta \cup \mathcal{U}, \mathbf{z}) I_A \right] + L^{p-d}. \end{aligned}$$

In the final two steps we prove that

$$\frac{1}{|Q_{mL}|} \mathbb{E} \left[\mathbb{M}^p(\mathcal{N}^\eta \cup \mathcal{U}, \mathbf{z}) I_A \right] \lesssim L^{p-d} \quad (4.23)$$

and

$$\frac{1}{|Q_{mL}|} \mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p(\mathcal{N}^\eta \cup \mathcal{U}, \mathcal{M}^\eta \cup \mathcal{V}) \right] \lesssim \eta^{1-\frac{p}{d}} + C(\eta) L^{-\beta}. \quad (4.24)$$

In combination with (4.21) this would conclude the proof of (4.20).

Step 4. Proof of (4.23). On A , we have $|\mathcal{N}_{Q_{mL}}^\eta| \geq m^d$, thus (randomly) choosing m^d points from \mathcal{N}^η , we find after relabelling a family $(X_i)_{i=1}^{m^d}$ of points i.i.d. and uniformly distributed on Q_{mL} . Recalling that \mathbf{z} denotes the set of centres of the m^d cubes Q_i we can bound

$$\mathbb{E} \left[\mathbb{M}^p(\mathcal{N}^\eta \cup \mathcal{U}, \mathbf{z}) I_A \right] \leq \mathbb{E} \left[\mathbb{M}^p((X_i)_{i=1}^{m^d}, \mathbf{z}) \right].$$

We then use (3.4) with $n = m^d$ and λ the uniform density on the cube Q_{mL} , so that

$$\mathbb{E} \left[\mathbb{M}^p((X_i)_{i=1}^{m^d}, \mathbf{z}) \right] \lesssim \mathbb{E} \left[\mathbb{W}_{Q_{mL}}^p \left(\sum_{i=1}^{m^d} \delta_{X_i}, \frac{m^d}{|Q_{mL}|} \right) \right] + \mathbb{E} \left[\mathbb{W}_{Q_{mL}}^p \left(\mu^{\mathbf{z}}, \frac{m^d}{|Q_{mL}|} \right) \right] \lesssim m^d L^p,$$

having used (6.1) to bound the first term (the second term is trivially estimated by transporting the mass on each cube Q_i to its center). This proves (4.23).

Step 5. Proof of (4.24). We use Assumption A5 (on Q_{mL} instead of Q_1 , see Remark 3.9), so that

$$\mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p(\mathcal{N}^\eta \cup \mathcal{U}, \mathcal{M}^\eta \cup \mathcal{V}) \right] \lesssim (mL)^p \mathbb{E} \left[(|\mathcal{N}_{Q_{mL}}^\eta| + |\mathcal{U}|)^{1-\frac{p}{d}} \right] + \mathbb{E} \left[\mathbb{M}^p(\mathcal{N}^\eta \cup \mathcal{U}, \mathcal{M}^\eta \cup \mathcal{V}) \right].$$

We further bound the first contribution using Hölder inequality

$$\begin{aligned} \mathbb{E} \left[(|\mathcal{N}_{Q_{mL}}^\eta| + |\mathcal{U}|)^{1-\frac{p}{d}} \right] &\leq \mathbb{E} \left[|\mathcal{N}_{Q_{mL}}^\eta| \right]^{1-\frac{p}{d}} + \mathbb{E} \left[|\mathcal{U}| \right]^{1-\frac{p}{d}} \\ &\lesssim \eta^{1-\frac{p}{d}} (mL)^{d-p} + \mathbb{E} \left[|\mathcal{U}| \right]^{1-\frac{p}{d}}. \end{aligned} \quad (4.25)$$

To proceed further, let us recall that in Section 4.3 we argued that $((\mathcal{U}_{Q_i}, \mathcal{V}_{Q_i}))_{i=1}^{m^d}$ are independent (and also independent from $\mathcal{N}^\eta, \mathcal{M}^\eta$). Moreover, since the law of each $(\mathcal{N}_{Q_i}^{1-\eta}, \mathcal{M}_{Q_i}^{1-\eta})$ coincides with that of $(\mathcal{N}_{Q_L}^{1-\eta}, \mathcal{M}_{Q_L}^{1-\eta})$ (up to a translation by $-Lz_i$, since $Q_i = Q_L + Lz_i$) it follows that the same property holds for the processes $(\mathcal{U}_{Q_i}, \mathcal{V}_{Q_i})$: their law coincides with that of $(\mathcal{U}_{Q_L}, \mathcal{V}_{Q_L})$ (also up to translating by $-Lz_i$).

Using (4.3) with $q = 1$, we obtain

$$\mathbb{E}[|\mathcal{U}|] = m^d \mathbb{E}[|\mathcal{U}_{Q_L}|] \lesssim m^d L^{\frac{d}{2}},$$

thus (4.25) yields

$$\frac{(mL)^p}{|Q_{mL}|} \mathbb{E} \left[(|\mathcal{N}_{Q_{mL}}^\eta| + |\mathcal{U}|)^{1-\frac{p}{d}} \right] \lesssim \eta^{1-\frac{p}{d}} + L^{\frac{p-d}{2}}.$$

Combining this with Proposition 6.5, concludes the proof of (4.24). \square

4.5. Hölder density on a cube. In this section, we still assume that $\Omega = Q$ is a cube, but consider the case of a general Hölder continuous density ρ , uniformly bounded from above and below. Up to rescaling and translation, it is sufficient to consider the case $\Omega = (0, 1)^d$.

The proof of (4.1) in this case is obtained by combining the case of constant density treated above together with Lemma 4.3 and the following claim: there exists a constant $C = C(\rho) > 0$ such that, for $r < C$ and for every cube $Q \subseteq (0, 1)^d$ with side length r the following inequality holds:

$$\mathbb{E} \left[\mathcal{C}_P^p \left(\mathcal{N}_Q^{n\rho}, \mathcal{M}_Q^{n\rho} \right) \right] \leq (1 + C^{-1}r^\alpha) \mathbb{E} \left[\mathcal{C}_P^p \left(\mathcal{N}^{n\rho(Q)/r^d}, \mathcal{M}^{n\rho(Q)/r^d} \right) \right], \quad (4.26)$$

where $\mathcal{N}^{n\rho(Q)/r^d}, \mathcal{M}^{n\rho(Q)/r^d}$ are two independent Poisson point processes with constant intensity $n\rho(Q)/r^d$ on the cube $(0, r)^d$, and α denotes the Hölder exponent of ρ .

Indeed, assume that the claim holds and let us prove (4.1). Given any $r < C(p, \rho)$ of the form $r = 1/K^{1/d}$, we consider a partition of $(0, 1)^d = \bigcup_{k=1}^K Q_k$ into K disjoint sub-cubes of side length r , so that

$$\begin{aligned} & \limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} \left[\mathcal{C}_P^p \left(\mathcal{N}^{n\rho}, \mathcal{M}^{n\rho} \right) \right] \\ & \stackrel{(4.7)}{\leq} \sum_{k=1}^K \limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} \left[\mathcal{C}_P^p \left(\mathcal{N}_{Q_k}^{n\rho}, \mathcal{M}_{Q_k}^{n\rho} \right) \right] \\ & \stackrel{(4.26)}{\leq} (1 + C^{-1}r^\alpha) \sum_{k=1}^K \limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} \left[\mathcal{C}_P^p \left(\mathcal{N}^{n\rho(Q_k)/r^d}, \mathcal{M}^{n\rho(Q_k)/r^d} \right) \right] \\ & = \beta_P (1 + C^{-1}r^\alpha) \sum_{k=1}^K \rho(Q_k)^{1-\frac{p}{d}} r^p, \end{aligned}$$

where the last line follows from (4.1) in the case of a cube and constant intensity. Letting $K \rightarrow \infty$, we have that $r \rightarrow 0$ and the easily seen convergence

$$\lim_{K \rightarrow \infty} \sum_{k=1}^K \rho(Q_k)^{1-\frac{p}{d}} r^p = \lim_{K \rightarrow \infty} \int_{(0,1)^d} \sum_{k=1}^K I_{Q_k} \left(\frac{\rho(Q_k)}{r^d} \right)^{-\frac{p}{d}} \rho = \int_{(0,1)^d} \rho^{1-\frac{p}{d}}.$$

This would conclude the proof of (4.1) also in this case.

We now prove (4.26) for which we closely follow [2, Lemma 2.5]. Up to translating, we may assume that $Q = (0, r)^d$. We write $\rho_0 = \min_{(0,1)^d} \rho$ and 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 $x, y \in (0, 1)^d$,

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

thus $\|\rho^r - 1\|_{C^\alpha} \lesssim r^\alpha$ if r is sufficiently small. We define $S : Q \rightarrow Q$ as $S(x) = rT^{-1}(x/r)$, where T is the map provided by Proposition 2.11. It holds $\text{Lip } S = \text{Lip } T^{-1}$, and $S_{\#}1/r^d = \rho/\rho(Q)$. Therefore, $S\left(\mathcal{N}^{n\rho(Q)/r^d}\right) = (S(X_i))_{i=1}^{N^{n\rho(Q)/r^d}(Q)}$, which is a Poisson point process on Q with intensity $n\rho$, i.e., it has the same law as $\mathcal{N}_Q^{n\rho}$, and similarly $S\left(\mathcal{M}^{n\rho(Q)/r^d}\right)$ has the same law as $\mathcal{M}_Q^{n\rho}$. Therefore,

$$\begin{aligned} \mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p \left(\mathcal{N}_Q^{n\rho}, \mathcal{M}_Q^{n\rho} \right) \right] &= \mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p \left(S(\mathcal{N}^{n\rho(Q)/r^d}), S(\mathcal{M}^{n\rho(Q)/r^d}) \right) \right] \\ &\stackrel{(3.2)}{\leq} (\text{Lip } S)^p \mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p \left(\mathcal{N}^{n\rho(Q)/r^d}, \mathcal{M}^{n\rho(Q)/r^d} \right) \right]. \end{aligned}$$

This proves the claim since $(\text{Lip } S)^p = (\text{Lip } T^{-1})^p \leq 1 + Cr^\alpha$ if r is sufficiently small.

Remark 4.6. Let us notice that the fact that Ω is a cube is not used in the proof of (4.26), which therefore holds true for every bounded domain Ω and Hölder continuous density ρ uniformly bounded from above and below. In particular, combining (4.26) with (4.19) we obtain that there exists $C = C(\rho) > 0$ such that, for every cube $Q \subseteq \Omega$ with side length $r < C$,

$$\mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p \left(\mathcal{N}_Q^{n\rho}, \mathcal{M}_Q^{n\rho} \right) \right] \lesssim |Q|n^{1-\frac{p}{d}}, \quad (4.27)$$

where the implicit constant depends on p , d and ρ only.

4.6. General density on a domain. We prove (4.1) for a domain Ω and a Hölder density ρ . The main difficulty here is that since we rely on the result established in the previous section we need to partition Ω into cubes. This is accomplished relying on the Whitney-type decomposition provided by Lemma 2.1. We begin by fixing a Whitney decomposition $\mathcal{Q} = (Q_i)_i$ such that every cube Q_i has side length $r < C$, where $C = C(\rho)$ is as in Remark 4.6. Then, by Lemma 2.1, for every sufficiently small $\delta > 0$ we have a finite Borel partition of $\Omega = \bigcup_{k=1}^K \Omega_k$, whose elements are collected into the two disjoint sets $\mathcal{Q}_\delta, \mathcal{R}_\delta$.

We fix $\eta \in (0, 1/2)$ and use the construction from Section 4.3. We set $\delta = n^{-\gamma}$ for $\gamma > 0$ to be fixed below. The first constraint is that (4.2) holds with $\lambda = n\rho$ so that we need $n\delta^d \gg 1$, i.e.

$$\gamma d < 1. \quad (4.28)$$

We first reduce to the case when there are many points in each Ω_k . Defining the event A as in (4.8) and arguing as in (4.9) gives here, for every $q > 0$, the inequality

$$\mathbb{P}(A^c) \lesssim_{q,\eta} \sum_{k=1}^K (n|\Omega_k|)^{-q} \lesssim_{q,\eta} n^{-q} \delta^{1-d-dq} = n^{-q(1-d\gamma)+(d-1)\gamma},$$

where we used (2.1) with $\alpha = -qd$ in the second inequality. Under the assumption (4.28) this is infinitesimal provided q is chosen sufficiently large. Arguing exactly as before we can thus reduce ourselves to the case where A holds. In that case, both (4.4) and (4.6) hold and thus by (4.5)

$$\begin{aligned} \mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p \left(\mathcal{N}^{n\rho}, \mathcal{M}^{n\rho} \right) I_A \right] &= \sum_{k=1}^K \mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p \left(\mathcal{N}_{\Omega_k}^{(1-\eta)n\rho}, \mathcal{M}_{\Omega_k}^{(1-\eta)n\rho} \right) \right] \\ &\lesssim \mathbb{E} \left[\mathcal{C}_{\mathbb{P}}^p \left(\mathcal{N}^{\eta n\rho} \cup \mathcal{U}, \mathcal{M}^{\eta n\rho} \cup \mathcal{V} \right) \right] + \sum_{k=1}^K \text{diam}(\Omega_k)^p. \end{aligned} \quad (4.29)$$

We start by considering the left-hand side of (4.29). For $\Omega_k \in \mathcal{R}_\delta$ we use the simple bound $\mathcal{C}_P^p \left(\mathcal{N}_{\Omega_k}^{(1-\eta)n\rho}, \mathcal{M}_{\Omega_k}^{(1-\eta)n\rho} \right) \lesssim \text{diam}(\Omega_k)^p |\mathcal{N}_{\Omega_k}^{(1-\eta)n\rho}|$, to estimate

$$\begin{aligned} n^{\frac{p}{d}-1} \sum_{\Omega_k \in \mathcal{R}_\delta} \mathbb{E} \left[\mathcal{C}_P^p \left(\mathcal{N}_{\Omega_k}^{(1-\eta)n\rho}, \mathcal{M}_{\Omega_k}^{(1-\eta)n\rho} \right) \right] &\lesssim n^{\frac{p}{d}-1} \delta^p \sum_{\Omega_k \in \mathcal{R}_\delta} \mathbb{E} \left[|\mathcal{N}_{\Omega_k}^{(1-\eta)n\rho}| \right] \\ &\lesssim n^{\frac{p}{d}-1} \delta^p \cdot \delta^{1-d} \cdot n\delta^d = n^{-\gamma + \frac{p}{d}(1-d\gamma)}. \end{aligned}$$

This tends to zero provided $\gamma d > p/(p+1)$ which is in particular true if (recall that $p < d$)

$$\gamma d > d/(d+1). \quad (4.30)$$

Notice that this condition is compatible with (4.28). Under condition (4.30) we thus have

$$\begin{aligned} \limsup_{n \rightarrow \infty} n^{1-\frac{p}{d}} \sum_{k=1}^K \mathbb{E} \left[\mathcal{C}_P^p \left(\mathcal{N}_{\Omega_k}^{(1-\eta)n\rho}, \mathcal{M}_{\Omega_k}^{(1-\eta)n\rho} \right) \right] \\ = \limsup_{n \rightarrow \infty} \sum_{\Omega_k \in \mathcal{Q}_\delta} n^{1-\frac{p}{d}} \mathbb{E} \left[\mathcal{C}_P^p \left(\mathcal{N}_{\Omega_k}^{(1-\eta)n\rho}, \mathcal{M}_{\Omega_k}^{(1-\eta)n\rho} \right) \right]. \end{aligned}$$

Since every $\Omega_k \in \mathcal{Q}_\delta$ is a cube, we may combine (4.1) in Ω_k together with the precise limit procedure, justified by the domination given in (4.27) (this is why each cube Q_i in the Whitney partition has side length $r < C$), to obtain

$$\limsup_{n \rightarrow \infty} n^{1-\frac{p}{d}} \sum_{k=1}^K \mathbb{E} \left[\mathcal{C}_P^p \left(\mathcal{N}_{\Omega_k}^{(1-\eta)n\rho}, \mathcal{M}_{\Omega_k}^{(1-\eta)n\rho} \right) \right] \leq (1-\eta)^{1-\frac{p}{d}} \int_{\Omega} \rho^{1-\frac{p}{d}}.$$

We now turn to the right-hand side of (4.29). The last term is easily estimated using directly (2.1) with $\alpha = p$. In particular, if $p < d-1$ we notice that

$$n^{\frac{p}{d}-1} \sum_{k=1}^K \text{diam}(\Omega_k)^p \lesssim n^{\frac{p}{d}-1} \delta^{1-(d-p)} = (n\delta^d)^{-(1-\frac{p}{d})} \delta$$

which goes to zero if (4.28) holds.

We finally estimate the first term in the right-hand side of (4.29). We argue as in (4.13) and (4.14) which we combine with Proposition 6.4 to obtain that for every $\varepsilon > 0$,

$$\begin{aligned} n^{1-\frac{p}{d}} \mathbb{E} \left[\mathcal{C}_P^p \left(\mathcal{N}^{\eta n\rho} \cup \mathcal{U}, \mathcal{M}^{\eta n\rho} \cup \mathcal{V} \right) \right] &\lesssim \mathbb{E} \left[|\mathcal{U}|/n \right]^{1-\frac{p}{d}} + \eta^{1-\frac{p}{d}} \\ &\quad + C(\eta, \varepsilon, \gamma) n^\varepsilon \left(\left(\max \left\{ n^{\frac{p}{d}} \delta^{p+1}, n^{\frac{2}{d}} \delta^3 \right\} \right)^\alpha + (n\delta^d)^{-\beta} \right). \end{aligned} \quad (4.31)$$

Using (2.1) with $\alpha = d/2 < d-1$ we have

$$\mathbb{E} \left[|\mathcal{U}|/n \right]^{1-\frac{p}{d}} \lesssim \left(\sum_{k=1}^K (|\Omega_k|/n)^{\frac{1}{2}} \right)^{1-\frac{p}{d}} \lesssim n^{\frac{1}{2}} \delta^{1-\frac{d}{2}} = \left(\delta(n\delta^d)^{-\frac{1}{2}} \right)^{1-\frac{p}{d}}.$$

Under condition (4.28) this term goes to zero. Regarding the term inside brackets in (4.31) we notice that if $q = \max \{p, 2\}$, then under condition (4.28),

$$\max \left\{ n^{\frac{p}{d}} \delta^{p+1}, n^{\frac{2}{d}} \delta^3 \right\} = n^{-\gamma + \frac{q}{d}(1-d\gamma)}.$$

In particular, as above this term goes to zero under condition (4.30).

We can thus choose first γ satisfying both (4.28) and (4.30) and then $\varepsilon = \varepsilon(\alpha, \beta, \gamma, p) > 0$ such that

$$\lim_{n \rightarrow \infty} n^\varepsilon \left(\left(\max \left\{ n^{\frac{p}{d}} \delta^{p+1}, n^{\frac{2}{d}} \delta^3 \right\} \right)^\alpha + (n\delta^d)^{-\beta} \right) = 0.$$

With this choice we find

$$\limsup_{n \rightarrow \infty} n^{1-\frac{p}{d}} \mathbb{E} \left[\mathcal{C}_P^p \left(\mathcal{N}^{\eta n\rho} \cup \mathcal{U}, \mathcal{M}^{\eta n\rho} \cup \mathcal{V} \right) \right] \lesssim \eta^{1-\frac{p}{d}},$$

from which we conclude the proof of (4.1) after sending $\eta \rightarrow 0$.

4.7. Uniform density on a domain. In this last case, we assume that Ω is a bounded domain with C^2 boundary and $\rho = I_\Omega/|\Omega|$ is uniform. After a simple rescaling, it is more convenient to argue with Poisson point processes $\mathcal{N}_\Omega^n, \mathcal{M}_\Omega^n$ with constant intensity n (on Ω) so that the thesis reduces to

$$\lim_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}_\Omega^n, \mathcal{M}_\Omega^n)] = \beta_P |\Omega|.$$

Since the boundary of Ω is C^2 , we can apply the result from the previous section and obtain the upper bound

$$\limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}_\Omega^n, \mathcal{M}_\Omega^n)] \leq \beta_P |\Omega|.$$

To prove the corresponding lower bound, we follow closely the argument of [6, Theorem 24]: we fix a cube Q sufficiently large so that $\Omega \subseteq Q$ and introduce a Poisson point process \mathcal{N}_Q^n with intensity n on Q . For $k = 2, \dots, K$, let Ω_k be the connected components of $Q \setminus \Omega$ so that $Q \setminus \Omega = \cup_{k=2}^K \Omega_k$. Notice that for every k either $\partial\Omega_k$ is C^2 or is the union of ∂Q and a C^2 surface. In particular each Ω_k satisfies (2.8). Using (4.15) with the decomposition $Q = \Omega \cup \bigcup_{k=2}^K \Omega_k$, we obtain

$$\begin{aligned} \beta_P |Q| &= \liminf_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}_Q^n, \mathcal{M}_Q^n)] \leq \liminf_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}_\Omega^n, \mathcal{M}_\Omega^n)] \\ &\quad + \sum_{k=2}^K \limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}_{\Omega_k}^n, \mathcal{M}_{\Omega_k}^n)]. \end{aligned}$$

Now for every k , using (4.1) we have

$$\limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}_{\Omega_k}^n, \mathcal{M}_{\Omega_k}^n)] \leq \beta_P |\Omega_k|.$$

Therefore,

$$\liminf_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}_\Omega^n, \mathcal{M}_\Omega^n)] \geq \beta_P |Q| - \beta_P \sum_{k=2}^K |\Omega_k| = \beta_P |Q|,$$

which is the desired conclusion.

5. PROOF OF MAIN RESULT

From Theorem 4.1, we deduce our main result Theorem 1.1. We follow a relatively standard strategy, using de-Poissonization and concentration of measure arguments with the necessary adjustments to deal with our setting. First, we argue that Theorem 4.1 yields similar convergence in the case of a deterministic number of independent points.

Proposition 5.1. *Let $d \geq 3$, $p \in [1, d)$ and let $P = (\mathcal{F}_{n,n})_{n \in \mathbb{N}}$ be a combinatorial optimization problem over complete bipartite graphs such that assumptions A1, A2, A3, A4 and A5 hold. Then, with $\beta_P \in (0, \infty)$ given by Theorem 4.1 the following hold.*

Let $\Omega \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary and such that (2.8) holds and let ρ be a Hölder continuous probability density on Ω , uniformly strictly positive and bounded from above.

Given i.i.d. random variables $(X_i)_{i=1}^\infty, (Y_j)_{j=1}^\infty$ with common law ρ , we have

$$\limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p((X_i)_{i=1}^n, (Y_j)_{j=1}^n)] \leq \beta_P \int_\Omega \rho^{1-\frac{p}{d}}. \quad (5.1)$$

Moreover, if ρ is the uniform density and Ω is either a cube or has C^2 boundary, the limit exists and is equal to the right-hand side.

Remark 5.2. The only properties we used to established Proposition 5.1 are the subadditivity property (3.7), the growth condition (3.9) as well as the p -homogeneity of the problem. In particular it holds for every bipartite p -homogeneous functional \mathcal{C} satisfying

- For every $\Omega \subset \mathbb{R}^d$ and every partition $\Omega = \cup_{k=1}^K \Omega_k$, $K \in \mathbb{N}$, if $\mathbf{x}^0, \mathbf{y}^0 \subseteq \Omega$ are such that $\min \{|\mathbf{x}^0|, |\mathbf{y}^0|\} \geq \max \{c_{A2}, K\}$, for every $k = 1, \dots, K$, $\mathbf{x}^k, \mathbf{y}^k \subseteq \Omega_k$ are such that $|\mathbf{x}^k| = |\mathbf{y}^k| = n_k$, with either $n_k \geq c_{A2}$ or $n_k = 0$ and $\mathbf{z} = (z_k)_{k=1}^K$ with $z_k \in \Omega_k$, for every $k = 1, \dots, K$ then

$$\mathcal{C} \left(\mathbf{x}^0 \cup \bigcup_{k=1}^K \mathbf{x}^k, \mathbf{y}^0 \cup \bigcup_{k=1}^K \mathbf{y}^k \right) - \sum_{k=1}^K \mathcal{C}(\mathbf{x}^k, \mathbf{y}^k) \lesssim \mathcal{C}(\mathbf{x}^0, \mathbf{y}^0) + M^p(\mathbf{z}, \mathbf{x}^0) + \sum_{k=1}^K \text{diam}(\Omega_k)^p. \quad (5.2)$$

- There exists $c_{A5} \geq 0$ such that, for every $\mathbf{x}, \mathbf{y} \subseteq (0, 1)^d$, we have

$$\mathcal{C}(\mathbf{x}, \mathbf{y}) \leq c_{A5} \left(\min \left\{ |\mathbf{x}|^{1-\frac{p}{d}}, |\mathbf{y}|^{1-\frac{p}{d}} \right\} + M^p(\mathbf{x}, \mathbf{y}) \right). \quad (5.3)$$

Proof. The proof is similar to the proof of Lemma 4.3. We set $\mathbf{x} = (X_i)_{i=1}^n$ and $\mathbf{y} = (Y_j)_{j=1}^n$. Let $\eta \in (0, 1/2)$ and consider two independent copies $\mathcal{N}^{(1-\eta)n\rho}$ and $\mathcal{M}^{(1-\eta)n\rho}$ of Poisson point processes with intensity $(1-\eta)n\rho$ on Ω . We claim that

$$\limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathbf{x}, \mathbf{y})] - \limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}^{(1-\eta)n\rho}, \mathcal{M}^{(1-\eta)n\rho})] \lesssim \eta^{1-\frac{p}{d}}. \quad (5.4)$$

By Theorem 4.1, this would conclude the proof of (5.1) since η is arbitrary. We introduce the random variables $N = \max \{n - |\mathcal{N}^{(1-\eta)n\rho}|, 0\}$ and $M = \max \{n - |\mathcal{M}^{(1-\eta)n\rho}|, 0\}$ and notice that by the concentration properties of Poisson random variables, also N and M have the concentration property. Moreover, the event

$$A = \{|N - \eta n| \leq \eta n/2\} \cap \{|M - \eta n| \leq \eta/2\}$$

is of overwhelming small probability and thus arguing exactly as in the proof of Lemma 4.3 we have

$$\limsup_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathbf{x}, \mathbf{y})] = \limsup_{n \rightarrow \infty} n^{1-\frac{p}{d}} \mathbb{E} [\mathcal{C}_P(\mathbf{x}, \mathbf{y}) I_A].$$

We let $\mathcal{N} = (X_i)_{i=n-N+1}^n$ and $\mathcal{M} = (Y_j)_{j=n-M+1}^n$ so that in A , $\mathbf{x} = \mathcal{N}^{(1-\eta)n\rho} \cup \mathcal{N}$, $\mathbf{y} = \mathcal{M}^{(1-\eta)n\rho} \cup \mathcal{M}$ and $\min \{|\mathcal{N}|, |\mathcal{M}|\} \gtrsim \eta n$.

In A we let $\mathbf{x}^1 \subset \mathcal{N}^{(1-\eta)n\rho}$ and $\mathbf{y}^1 \subset \mathcal{M}^{(1-\eta)n\rho}$ be such that $|\mathbf{x}^1| = |\mathbf{y}^1|$ and

$$\mathcal{C}_P(\mathcal{N}^{(1-\eta)n\rho}, \mathcal{M}^{(1-\eta)n\rho}) = \mathcal{C}_P(\mathbf{x}^1, \mathbf{y}^1).$$

We then set $\mathcal{U} = \mathcal{N}^{(1-\eta)n\rho} \setminus \mathbf{x}^1$, $\mathcal{V} = \mathcal{M}^{(1-\eta)n\rho} \setminus \mathbf{y}^1$, $\mathbf{x}^0 = \mathcal{U} \cup \mathcal{N}$ and $\mathbf{y}^0 = \mathcal{V} \cup \mathcal{M}$. Using Lemma 2.12 on Ω with $K = 1$, i.e. a trivial partition, we find that in A ,

$$\begin{aligned} \mathcal{C}_P(\mathbf{x}, \mathbf{y}) - \mathcal{C}_P(\mathcal{N}^{(1-\eta)n\rho}, \mathcal{M}^{(1-\eta)n\rho}) &\lesssim \mathcal{C}_P(\mathbf{x}^0, \mathbf{y}^0) + 1 \\ &\stackrel{(3.9)}{\lesssim} \min \left\{ |\mathbf{x}^0|^{1-\frac{p}{d}}, |\mathbf{y}^0|^{1-\frac{p}{d}} \right\} + M^p(\mathbf{x}^0, \mathbf{y}^0) + 1. \end{aligned}$$

Multiplying by I_A , taking expectation and arguing exactly as in (4.11) (using in particular Proposition 6.3) we conclude the proof of (5.4).

With a similar argument one can prove that

$$\liminf_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathcal{N}^{n\rho}, \mathcal{M}^{n\rho})] \leq \liminf_{n \rightarrow \infty} n^{\frac{p}{d}-1} \mathbb{E} [\mathcal{C}_P^p(\mathbf{x}, \mathbf{y})],$$

which concludes the proof of Proposition 5.1. \square

To conclude the proof of Theorem 1.1, we prove a concentration bound, which improves (5.1) to complete convergence. The argument requires minimal assumptions on the combinatorial optimization problem and relies essentially on the validity of a Poincaré inequality.

Proposition 5.3. *Let $d \geq 3$, $p \in [1, d)$ and let $\mathbf{P} = (\mathcal{F}_{n,n})_{n \in \mathbb{N}}$ be a combinatorial optimization problem over complete bipartite graphs such that assumptions **A3** and **A5** hold. Let $\Omega \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary and let ρ be a probability density on Ω , uniformly strictly positive and bounded from above. Let $(X_i)_{i=1}^\infty, (Y_j)_{j=1}^\infty$ be i.i.d. random variables with common law ρ .*

For every $q \geq 2$ and $\varepsilon > 0$,

$$\mathbb{P} \left(n^{\frac{p}{d}-1} |\mathcal{C}_{\mathbf{P}}^p((X_i)_{i=1}^n, (Y_j)_{j=1}^n) - \mathbb{E}[\mathcal{C}_{\mathbf{P}}^p((X_i)_{i=1}^n, (Y_j)_{j=1}^n)]| > \varepsilon \right) \leq_q \frac{1}{\varepsilon^q n^{\frac{\alpha q}{2}}}, \quad (5.5)$$

with

$$\alpha = \begin{cases} 1 - 2/d & \text{if } p \in [1, 2), \\ 1 - p/d & \text{if } p \geq 2. \end{cases}$$

In particular, complete (hence \mathbb{P} -a.s.) convergence holds:

$$\lim_{n \rightarrow \infty} n^{\frac{p}{d}-1} |\mathcal{C}_{\mathbf{P}}^p((X_i)_{i=1}^n, (Y_j)_{j=1}^n) - \mathbb{E}[\mathcal{C}_{\mathbf{P}}^p((X_i)_{i=1}^n, (Y_j)_{j=1}^n)]| = 0.$$

Remark 5.4 (Poincaré inequality). We first recall that for every Lipschitz function $F : \Omega^{2n} \rightarrow \mathbb{R}$ we have the following L^q -Poincaré inequality,

$$\mathbb{E} \left[|F((X_i)_{i=1}^n, (Y_j)_{j=1}^n) - \mathbb{E}[F((X_i)_{i=1}^n, (Y_j)_{j=1}^n)]|^q \right] \lesssim_q \mathbb{E} \left[|\nabla F((X_i)_{i=1}^n, (Y_j)_{j=1}^n)|^q \right]. \quad (5.6)$$

Here $|\nabla F|$ denotes the usual Euclidean norm of the gradient. We stress the fact that the implicit constant in (5.6) does not depend upon n .

Inequality (5.6) is a consequence of well-known facts: first, the assumptions on Ω yield the L^2 -Poincaré inequality with respect to the uniform measure,

$$\int_{\Omega} \left| u - \frac{1}{|\Omega|} \int_{\Omega} u \right|^2 \lesssim \int_{\Omega} |\nabla u|^2.$$

Using that the constant $c = \int_{\Omega} u \rho$ minimizes $\int_{\Omega} |u - c|^2 \rho$ and that ρ is bounded from above and below, and, we obtain the weighted version

$$\int_{\Omega} \left| u - \int_{\Omega} u \rho \right|^2 \rho \leq \int_{\Omega} \left| u - \frac{1}{|\Omega|} \int_{\Omega} u \right|^2 \rho \leq C \int_{\Omega} |\nabla u|^2 \rho.$$

for some $C = C(\rho, \Omega) \in (0, \infty)$. Then, a standard tensorization argument [32, Corollary 5.7] entails that the inequality holds also on the product space Ω^{2n} , endowed with the product measure $\rho^{\otimes 2n}$, with the same constant C . This yields (5.6) with $q = 2$.

The general case $q \geq 2$ follows finally from the chain rule. Preliminarily, we notice that if μ is a probability measure on \mathbb{R}^D , then the validity of the inequality

$$\int \left| u - \int u d\mu \right|^q d\mu \lesssim \int |\nabla u|^q d\mu \quad (5.7)$$

for every Lipschitz function $u : \mathbb{R}^D \rightarrow \mathbb{R}$ is equivalent to

$$\int |u - m_u|^q d\mu \lesssim \int |\nabla u|^q d\mu, \quad (5.8)$$

where m_u denotes a median of (the law) of u , i.e. any $m \in \mathbb{R}$ such that $\mu(u \leq m) \geq 1/2$ and $\mu(u \geq m) \geq 1/2$. Indeed,

$$\left| m_u - \int u d\mu \right| \leq \left| \int (m_u - u) d\mu \right| \leq \int |m_u - u| d\mu.$$

Since m_u can be characterized as a minimizer for $c \mapsto \int |u - c| d\mu$, we also have

$$\int |m_u - u| d\mu \leq \int \left| u - \int u d\mu \right| d\mu.$$

Using Jensen's inequality, we obtain

$$\left| m_u - \int u d\mu \right|^q \leq \min \left\{ \int |m_u - u|^q d\mu, \int \left| u - \int u d\mu \right|^q d\mu \right\}$$

Then, assuming that (5.7) or (5.8) holds, using the triangle inequality and the bound above, we obtain the validity of the other inequality.

To conclude, we assume that (5.8) holds for $q = 2$ and argue that it also holds for any $q \geq 2$. Up to adding a suitable constant, we can assume that u is Lipschitz with $m_u = 0$. We then consider the Lipschitz function $v = |u|^{q/2} \text{sign}(u)$ (recall that in our case the support of $\mu = \rho^{\otimes 2n}$ is bounded, hence we can assume that also u is bounded), so that $m_v = 0$ and apply the $q = 2$ case of (5.8):

$$\begin{aligned} \int |u|^q d\mu &= \int |v|^2 d\mu \lesssim \int |\nabla v|^2 d\mu \lesssim \int |u|^{q-2} |\nabla u|^2 d\mu \\ &\lesssim \left(\int |u|^q d\mu \right)^{1-2/q} \left(\int |\nabla u|^q d\mu \right)^{2/q}. \end{aligned}$$

Dividing both sides by $(\int |u|^q d\mu)^{1-2/q}$ yields the desired conclusion.

Proof of Proposition 5.3. The second statement follows choosing q sufficiently large in (5.5) so that the right-hand side in (5.5) is summable. We thus focus on the proof of (5.5). Given a feasible $G \subseteq \mathcal{K}_{n,n}$, i.e., $G \in \mathcal{F}_{n,n}$, and $\mathbf{x} = (x_i)_{i=1}^n$, $\mathbf{y} = (y_j)_{j=1}^n \subseteq \Omega$, write

$$w_G(\mathbf{x}, \mathbf{y}) = \sum_{\{(1,i),(2,j)\} \in E_G} |x_i - y_j|^p.$$

Since $p \geq 1$, w_G is Lipschitz with a.e. derivative given by

$$\nabla_{x_i} w_G(\mathbf{x}, \mathbf{y}) = \sum_{(2,j) \in \mathcal{N}_G((1,i))} p |x_i - y_j|^{p-2} (x_i - y_j),$$

and

$$\nabla_{y_j} w_G(\mathbf{x}, \mathbf{y}) = - \sum_{(1,i) \in \mathcal{N}_G((1,j))} p |x_i - y_j|^{p-2} (x_i - y_j).$$

Notice also that w_G is differentiable at every (\mathbf{x}, \mathbf{y}) such that $x_i \neq y_j$ for every i, j . Since $G \in \mathcal{F}_{n,n}$, assumption A3 yields that the sums above contain at most c_3 terms, hence we bound, using Cauchy-Schwarz inequality,

$$|\nabla_{x_i} w_G(\mathbf{x}, \mathbf{y})|^2 \lesssim \sum_{(2,j) \in \mathcal{N}_G((1,i))} |x_i - y_j|^{2(p-1)},$$

and similarly

$$|\nabla_{y_j} w_G(\mathbf{x}, \mathbf{y})|^2 \lesssim \sum_{(2,j) \in \mathcal{N}_G((1,i))} |x_i - y_j|^{2(p-1)}.$$

Summing upon i and $j \in \{1, \dots, n\}$, we obtain, for the Euclidean norm of the gradient, the inequality

$$|\nabla w_G(\mathbf{x}, \mathbf{y})|^2 \lesssim \sum_{\{(1,i),(2,j)\} \in E_G} |x_i - y_j|^{2(p-1)}.$$

If $p \geq 2$, we simply bound each term $|x_i - y_j|^{2(p-1)} \leq \text{diam}(\Omega)^{p-2} |x_i - y_j|^p$, obtaining

$$|\nabla w_G(\mathbf{x}, \mathbf{y})|^2 \lesssim \sum_{\{(1,i),(2,j)\} \in E_G} |x_i - y_j|^p = w_G(\mathbf{x}, \mathbf{y}).$$

If $p \in [1, 2)$, we use Hölder inequality and the fact that $|E_G| \lesssim n$ (again by assumption **A3**), to obtain

$$|\nabla w_G(\mathbf{x}, \mathbf{y})|^2 \lesssim \left(\sum_{\{(1,i),(2,j)\} \in E_G} |x_i - y_j|^p \right)^{\frac{1}{r}} n^{1-\frac{1}{r}} = w_G(\mathbf{x}, \mathbf{y})^{\frac{1}{r}} n^{1-\frac{1}{r}},$$

with $r = p/(2(p-1))$.

Using the trivial bound $w_G(\mathbf{x}, \mathbf{y}) \lesssim n$, it follows in particular that each $w_G(\mathbf{x}, \mathbf{y})$ has a Lipschitz constant bounded independently of G (although the bound depends upon n). Therefore, also

$$\mathcal{C}_p^p(\mathbf{x}, \mathbf{y}) = \inf_{G \in \mathcal{F}_{n,n}} w_G(\mathbf{x}, \mathbf{y}),$$

is Lipschitz, hence differentiable at Lebesgue a.e. (\mathbf{x}, \mathbf{y}) , by Rademacher theorem. Let (\mathbf{x}, \mathbf{y}) be a point of differentiability for both w_G and \mathcal{C}_p^p (which holds for Lebesgue a.e. point). Let $G = G(\mathbf{x}, \mathbf{y}) \in \mathcal{F}_{n,m}$ be any minimizer for the problem on the graph $\mathcal{K}(\mathbf{x}, \mathbf{y})$ (which is a.e. unique if $p > 1$ by Remark **5.5**). For every $(\mathbf{x}', \mathbf{y}')$, we have the inequality

$$\mathcal{C}_p^p(\mathbf{x}', \mathbf{y}') \leq w_G(\mathbf{x}', \mathbf{y}'),$$

with equality at (\mathbf{x}, \mathbf{y}) , hence we obtain the identities,

$$\nabla_{x_i} \mathcal{C}_p^p(\mathbf{x}, \mathbf{y}) = \nabla_{x_i} w_G(\mathbf{x}, \mathbf{y}), \quad \nabla_{y_j} \mathcal{C}_p^p(\mathbf{x}, \mathbf{y}) = \nabla_{y_j} w_G(\mathbf{x}, \mathbf{y}). \quad (5.9)$$

Therefore,

$$|\nabla \mathcal{C}_p^p(\mathbf{x}, \mathbf{y})|^2 \lesssim \begin{cases} \mathcal{C}_p^p(\mathbf{x}, \mathbf{y})^{\frac{1}{r}} n^{1-\frac{1}{r}} & \text{if } p \in [1, 2), \\ \mathcal{C}_p^p(\mathbf{x}, \mathbf{y}) & \text{if } p \geq 2. \end{cases}$$

If now $\mathbf{x} = (X_i)_{i=1}^n$ and $\mathbf{y} = (Y_j)_{j=1}^n$, combining this with (5.6) and (3.9) yields

$$\begin{aligned} \mathbb{E} [|\mathcal{C}_p^p(\mathbf{x}, \mathbf{y}) - \mathbb{E}[\mathcal{C}_p^p(\mathbf{x}, \mathbf{y})]|^q] &\lesssim_q \mathbb{E} [|\nabla \mathcal{C}_p^p(\mathbf{x}, \mathbf{y})|^q] \\ &\lesssim \begin{cases} \mathbb{E} \left[\left(n^{1-\frac{p}{d}} + M^p(\mathbf{x}, \mathbf{y}) \right)^{\frac{q}{2r}} \right] n^{(1-\frac{1}{r})\frac{q}{2}} & \text{if } p \in [1, 2), \\ \mathbb{E} \left[\left(n^{1-\frac{p}{d}} + M^p(\mathbf{x}, \mathbf{y}) \right)^{\frac{q}{2}} \right] & \text{if } p \geq 2. \end{cases} \end{aligned}$$

By the equivalence between M^p and W^p (recall (3.3)), the triangle inequality (2.12) and (2.14) and finally using (6.2) with qp instead of p , we bound from above

$$\mathbb{E} \left[\left(M^p((X_i)_{i=1}^n, (Y_j)_{j=1}^n) \right)^{\frac{q}{2}} \right] \lesssim n^{(1-\frac{p}{d})\frac{q}{2}}.$$

If $p \geq 2$, we conclude at once that

$$\mathbb{E} [|\mathcal{C}_p^p(\mathbf{x}, \mathbf{y}) - \mathbb{E}[\mathcal{C}_p^p(\mathbf{x}, \mathbf{y})]|^q] \lesssim_q n^{(1-\frac{p}{d})\frac{q}{2}},$$

hence (5.5) by Markov inequality. If $p \in [1, 2)$, we bound similarly and obtain, after simple computations,

$$\mathbb{E} [|\mathcal{C}_p^p(\mathbf{x}, \mathbf{y}) - \mathbb{E}[\mathcal{C}_p^p(\mathbf{x}, \mathbf{y})]|^q] \lesssim n^{((1-\frac{p}{d})(1-\frac{1}{p})+\frac{1}{p}-\frac{1}{2})q},$$

which leads to the corresponding case of (5.5) by Markov inequality. \square

Remark 5.5 (uniqueness of minimizers). If $p > 1$, for Lebesgue a.e. (\mathbf{x}, \mathbf{y}) , the minimizer $G \in \mathcal{F}_{n,m}$ for the problem on $\mathcal{K}(\mathbf{x}, \mathbf{y})$ is unique. This in particular yields that it is unique a.s., when $\mathbf{x} = (X_i)_{i=1}^n$, $\mathbf{y} = (Y_j)_{j=1}^m$ are random i.i.d. with a common density ρ . For simplicity, we argue in the case of $|\mathbf{x}| = |\mathbf{y}|$ only, but the same result holds in general.

Let (\mathbf{x}, \mathbf{y}) be a differentiability point for $\mathcal{C}_p^p(\mathbf{x}, \mathbf{y})$ with $X_i \neq Y_j$ for every i, j . Notice that by the previous proof this holds a.s. . Let $G, G' \in \mathcal{F}_{n,n}$ be minimizers for the problem

on $\mathcal{K}(\mathbf{x}, \mathbf{y})$, so that by (5.9) we obtain that, for every $i \in [n]$, $\nabla_{x_i} w_G(\mathbf{x}, \mathbf{y}) = \nabla_{x_i} w_{G'}(\mathbf{x}, \mathbf{y})$, i.e.,

$$\sum_{(2,j) \in \mathcal{N}_G((1,i))} |x_i - y_j|^{p-2} (x_i - y_j) = \sum_{(2,j) \in \mathcal{N}_{G'}((1,i))} |x_i - y_j|^{p-2} (x_i - y_j)$$

Assuming that $E_G \neq E_{G'}$, we can find $i, j \in [n]$ such that $(2, j) \in \mathcal{N}_{G'}((1, i)) \setminus \mathcal{N}_G((1, i))$ (up to exchanging the roles of G and G'). Then,

$$\begin{aligned} |x_i - y_j|^{p-2} (x_i - y_j) &= \sum_{(2,k) \in \mathcal{N}_G((1,i))} |x_i - y_k|^{p-2} (x_i - y_k) \\ &\quad - \sum_{(2,k) \in \mathcal{N}_{G'}((1,i)) \setminus \{(2,j)\}} |x_i - y_k|^{p-2} (x_i - y_k). \end{aligned} \quad (5.10)$$

We notice that the right-hand side above is a function $U(\mathbf{x}, \mathbf{y})$ which however does not depend on the variable y_j . The map

$$z \in \mathbb{R}^d \mapsto |z|^{p-2} z \in \mathbb{R}^d$$

is invertible, with a Borel inverse which we denote by f , hence we can rewrite (5.10) equivalently as the identity

$$y_j = x_i - f(U(\mathbf{x}, \mathbf{y})),$$

where right-hand side is a Borel function of (\mathbf{x}, \mathbf{y}) which does not depend on y_j . This identity however cannot hold on a set of positive Lebesgue measure.

6. BOUNDS FOR THE EUCLIDEAN ASSIGNMENT PROBLEM

In this section we establish some novel upper bounds for the random Euclidean assignment problem, in the case of not necessarily i.i.d. uniformly distributed points.

6.1. Matching of i.i.d. points. We begin with a general upper bound for the Wasserstein distance between the empirical measure of i.i.d. points and the corresponding common law when $d \geq 3$ and $p \geq 1$. As a consequence, we also obtain a similar bound for the Euclidean assignment problem. We derive the general case of a Hölder continuous law bounded above and below on an open connected set with Lipschitz boundary Ω from the case of the uniform law on a cube $Q \subseteq \mathbb{R}^d$. In that case, it is a well-known result, marginally discussed in [3], where the focus is on the $d = 2$ case. However, we point out that the case $d \geq 3$, $p \geq d/2$ was, to our knowledge, not explicitly covered in the literature until the proof provided by [33], which clearly extends to any $p \neq 2$ (see also [24]).

Proposition 6.1. *Let $d \geq 3$, $p \geq 1$ and Ω be a bounded connected open set with Lipschitz boundary. For every Hölder continuous density $\rho : \Omega \mapsto \mathbb{R}$ bounded above and below and independent sequences $(X_i)_{i=1}^\infty, (Y_j)_{j=1}^\infty$ of i.i.d. random variables with common law ρ ,*

$$\mathbb{E} \left[\mathbf{W}^p \left(\sum_{i=1}^n \delta_{X_i}, n\rho \right) \right] \lesssim |\Omega|^{\frac{p}{d}} n^{1-\frac{p}{d}}, \quad (6.1)$$

and therefore

$$\mathbb{E} \left[\mathbf{M}^p \left((X_i)_{i=1}^n, (Y_j)_{j=1}^m \right) \right] \lesssim |\Omega|^{\frac{p}{d}} \min \{n, m\}^{1-\frac{p}{d}}. \quad (6.2)$$

Proof. Inequality (6.2) follows from (6.1) assuming e.g. $n \leq m$ and (3.4) with $\lambda = \rho$. Hence, we focus on the proof of (6.1). By Jensen inequality (2.14), it is enough to prove this bound for large p so that we may assume without loss of generality that $p > d/(d-1)$. We then set $\mu = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$.

We first prove the statement in the case $\Omega = Q$ is a cube. By scaling we may assume that $Q = (0, 1)^d$ is the unit cube. By Proposition 2.11, there is a bi-Lipschitz map $T : Q \mapsto Q$

with Lipschitz constant depending only on ρ such that $T\#\rho = 1$. Then, $X'_i = T(X_i)$ are i.i.d. uniformly distributed on Q . Letting $\mu' = T\#\mu = \frac{1}{n} \sum_{i=1}^n \delta_{X'_i}$ we have

$$W^p(\mu, \rho) \lesssim W^p(\mu', 1)$$

and the statement follows from [33].

Consider now Ω a general bounded connected open set with Lipschitz boundary. We say that Ω is well-partitioned if there exists convex polytopes $(\Omega_k)_{k=1}^K$ covering Ω , with $|\Omega_k \cap \Omega_{k'}| = 0$ for $k \neq k'$ and such that each Ω_k is bi-Lipschitz homeomorphic to a cube. By [45], every connected and Lipschitz domain is bi-Lipschitz homeomorphic to a well-partitioned and smooth domain so that arguing exactly as above we may assume that Ω itself is smooth and well-partitioned. Let $T_k : \Omega_k \mapsto Q_k$ be Lipschitz homeomorphisms between Ω_k and some cubes Q_k . We then define $\rho_k = T_k\#\rho$, $n_k = \mu(\Omega_k)$ and $\mu_k = \frac{n}{n_k} T_k\#\mu$. Notice in particular that we may write $\mu_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \delta_{Y_i}$ where $(Y_i)_{i=1}^{\infty}$ are i.i.d. with common law $\rho_k/\rho_k(Q_k)$ and that n_k is a Binomial random variable with parameters n and $\rho_k(Q_k) = \rho(\Omega_k)$. Using (2.15) with $\varepsilon = 1$ we thus find

$$\begin{aligned} W^p(\mu, \rho) &\lesssim \sum_{k=1}^K W_{\Omega_k}^p \left(\mu, \frac{n_k}{n\rho(\Omega_k)} \rho \right) + W^p \left(\sum_{k=1}^K \frac{n_k}{n\rho(\Omega_k)} \rho I_{\Omega_k}, \rho \right) \\ &\stackrel{(2.16)}{\lesssim} \sum_{k=1}^K \frac{n_k}{n} W_{Q_k}^p(\mu_k, \rho_k) + \left\| \sum_{k=1}^K \left(\frac{n_k}{n\rho(\Omega_k)} - 1 \right) I_{\Omega_k} \rho \right\|_{W^{-1,p}(\Omega)}^p \\ &\stackrel{(2.7)}{\lesssim} \sum_{k=1}^K \frac{n_k}{n} W_{Q_k}^p(\mu_k, \rho_k) + \sum_{k=1}^K |\Omega_k| \left| \frac{n_k}{n\rho(\Omega_k)} - 1 \right|^p. \end{aligned}$$

Taking the expectation and using the concentration properties of binomial random variables (2.22) we find

$$\mathbb{E} [W^p(\mu, \rho)] \lesssim \sum_{k=1}^K \mathbb{E} \left[\frac{n_k}{n} W_{Q_k}^p(\mu_k, \rho_k) \right] + \frac{1}{n^{\frac{p}{2}}}.$$

By the first part of the proof and the concentration properties of Binomial random variables we get

$$\mathbb{E} \left[\frac{n_k}{n} W_{Q_k}^p(\mu_k, \rho_k) \right] \lesssim \frac{1}{n^{\frac{p}{d}}}$$

which concludes the proof of (6.2) since $p/2 > p/d$. \square

Remark 6.2. By translation and scaling invariance, when $\Omega = Q$ and $\rho = \frac{1}{|\Omega|} I_Q$ is the uniform measure of a cube $Q \subset \mathbb{R}^d$, the implicit constant in (6.2) does not depend on Q .

6.2. Matching with a fraction of i.i.d. points. In this section we extend the bound (6.2) for the matching to the case where most of the points are still i.i.d. but essentially no assumption is made on the remaining points. This is used in Theorem 4.1 and in the de-Poissonization procedure (see Proposition 5.1). Just like in Theorem 4.1 we will have to consider three different situations. Let us however set some common notation. Letting \mathcal{N} , \mathcal{M} , \mathcal{U} and \mathcal{V} be point processes on Ω (\mathcal{N} and \mathcal{M} will contain the i.i.d. points), we want to estimate

$$\mathbb{E} [M^p(\mathcal{U} \cup \mathcal{N}, \mathcal{V} \cup \mathcal{M})].$$

Setting

$$Z = \min \{ |\mathcal{U}| + |\mathcal{N}|, |\mathcal{V}| + |\mathcal{M}| \},$$

we want to construct two (random) subsets $\mathcal{S} \subseteq \mathcal{U} \cup \mathcal{N}$, $\mathcal{T} \subseteq \mathcal{V} \cup \mathcal{M}$, both containing Z points, so that

$$M^p(\mathcal{U} \cup \mathcal{N}, \mathcal{V} \cup \mathcal{M}) \leq W^p(\mu^{\mathcal{S}}, \nu^{\mathcal{T}}) \stackrel{(2.12)}{\lesssim} W^p(\mu^{\mathcal{S}}, Z\rho) + W^p(\nu^{\mathcal{T}}, Z\rho), \quad (6.3)$$

where $\mu^{\mathcal{S}}, \mu^{\mathcal{T}}$ are the associated empirical measures. We then separately estimate the two terms on the right-hand side of (6.3). Since the construction is completely symmetric, we detail it only for $\mathcal{S} \subseteq \mathcal{U} \cup \mathcal{N}$. It is given as the union of two sets, a “good” set $\mathcal{G} \subseteq \mathcal{N}$ and a “bad” set $\mathcal{B} \subseteq \mathcal{U}$. We first define the set \mathcal{G} by sampling without replacement

$$|\mathcal{G}| = \min \{|\mathcal{N}|, Z\}$$

points from \mathcal{N} . Similarly, the set \mathcal{B} is constructed by sampling without replacement

$$|\mathcal{B}| = \max \{Z - |\mathcal{N}|, 0\}$$

points from \mathcal{U} . Notice that

$$Z = |\mathcal{G}| + |\mathcal{B}| \tag{6.4}$$

and that when conditioned on $|\mathcal{G}|$, the points in \mathcal{G} are still i.i.d. with common law ρ . We then write $\mu^{\mathcal{S}} = \mu^{\mathcal{G}} + \mu^{\mathcal{B}}$ for the associated empirical measure. Using the triangle inequality (2.12) and (2.13), we then split the estimate in two:

$$\begin{aligned} \mathbb{W}^p(\mu^{\mathcal{S}}, Z\rho) &\lesssim \mathbb{W}^p(\mu^{\mathcal{G}} + \mu^{\mathcal{B}}, |\mathcal{G}|\rho + \mu^{\mathcal{B}}) + \mathbb{W}^p(|\mathcal{G}|\rho + \mu^{\mathcal{B}}, Z\rho) \\ &\lesssim \mathbb{W}^p(\mu^{\mathcal{G}}, |\mathcal{G}|\rho) + \mathbb{W}^p(|\mathcal{G}|\rho + \mu^{\mathcal{B}}, Z\rho). \end{aligned}$$

Taking expectation we find

$$\mathbb{E} [\mathbb{W}^p(\mu^{\mathcal{S}}, Z\rho)] \lesssim \mathbb{E} [\mathbb{W}^p(\mu^{\mathcal{G}}, |\mathcal{G}|\rho)] + \mathbb{E} [\mathbb{W}^p(|\mathcal{G}|\rho + \mu^{\mathcal{B}}, Z\rho)]. \tag{6.5}$$

To estimate the first term in the right-hand side, we will rely on (6.1). It is in the estimate of the last term that we need to argue differently depending on the cases. In the first one (see Proposition 6.3), since we have a good control on the moments of $|\mathcal{U}|$ we can directly appeal to Proposition 2.9. In the two other cases (see Propositions 6.4 and 6.5) we need to combine it with a localization argument.

We start with the first case.

Proposition 6.3. *Let $d \geq 3$, $p \geq 1$ and Ω be a bounded domain with Lipschitz boundary, $\rho : \Omega \mapsto \mathbb{R}$ be a Hölder continuous density bounded above and below and $(X_i)_{i=1}^{\infty}, (Y_j)_{j=1}^{\infty}$ be independent sequences of i.i.d. random variables with common law ρ .*

Then, there exists $\alpha = \alpha(p, d) < 2$ and $\beta = \beta(p) > 0$ such that the following holds. Let $M, N \in \mathbb{N}$ be random variables satisfying concentration (recall Definition 2.13) and set $\mathcal{N} = (X_i)_{i=1}^N, \mathcal{M} = (Y_j)_{j=1}^M$ and $h = \min \{\mathbb{E}[M], \mathbb{E}[N]\}$. Then, for every point processes \mathcal{U}, \mathcal{V} , for which there exists $1 \leq H \leq h$ such that for every $q \geq 1$

$$\mathbb{E} [|\mathcal{U}|^q + |\mathcal{V}|^q] \leq C(q)H^q \tag{6.6}$$

for some $C(q) > 0$, we have

$$\mathbb{E} [\mathbb{M}^p(\mathcal{U} \cup \mathcal{N}, \mathcal{V} \cup \mathcal{M})] \lesssim h^{1-\frac{p}{d}} \left(1 + \left(\frac{H^\alpha}{h} \right)^\beta \right).$$

Here the implicit constant depends only on p, d , the constants involved in the concentration properties of M, N and $(C(q))_{q \geq 1}$ from (6.6).

Proof. Starting from (6.3) and (6.5) we first estimate by (6.2) and Hölder inequality,

$$\mathbb{E} [\mathbb{W}^p(\mu^{\mathcal{G}}, |\mathcal{G}|\rho)] \lesssim \mathbb{E} [|\mathcal{G}|^{1-\frac{p}{d}}] \leq \mathbb{E} [|\mathcal{G}|]^{1-\frac{p}{d}}.$$

Since $|\mathcal{G}| \leq \min \{M, N\} + |\mathcal{V}|$, by (6.6) with $q = 1$ and $H \leq h$, we have $\mathbb{E} [|\mathcal{G}|] \lesssim h$ and thus

$$\mathbb{E} [\mathbb{W}^p(\mu^{\mathcal{G}}, |\mathcal{G}|\rho)] \lesssim h^{1-\frac{p}{d}}.$$

We are then left with the proof of

$$\mathbb{E} [\mathbb{W}^p(|\mathcal{G}|\rho + \mu^{\mathcal{B}}, Z\rho)] \lesssim h^{1-\frac{p}{d}} \left(\frac{H^\alpha}{h} \right)^\beta. \tag{6.7}$$

We first single out the event

$$A = \{|\mathcal{G}| \geq h/2\}$$

and claim that for $q \geq 1$

$$\mathbb{P}[A^c] \lesssim_q h^{-q}. \quad (6.8)$$

Indeed, since $A^c \subset \{N \leq \mathbb{E}[N]/2\} \cup \{M \leq \mathbb{E}[M]/2\}$, (6.8) follows by combining a union bound together with the concentration properties of M and N . Since

$$\mathbb{W}^p(|\mathcal{G}|_\rho + \mu^{\mathcal{B}}, Z_\rho) \leq \mathbb{W}^p(\mu^{\mathcal{B}}, |\mathcal{B}|_\rho) \lesssim |\mathcal{B}| \leq |\mathcal{U}|,$$

we then find

$$\mathbb{E}[\mathbb{W}^p(|\mathcal{G}|_\rho + \mu^{\mathcal{B}}, Z_\rho) I_{A^c}] \lesssim \mathbb{E}[|\mathcal{U}|^2]^{\frac{1}{2}} \mathbb{P}[A^c]^{\frac{1}{2}} \stackrel{(6.6)\&(6.8)}{\lesssim_q} h^{-q} H.$$

By taking q large enough, in order to prove (6.7) it is therefore sufficient to show

$$\mathbb{E}[\mathbb{W}^p(|\mathcal{G}|_\rho + \mu^{\mathcal{B}}, Z_\rho) I_A] \lesssim h^{1-\frac{p}{d}} \left(\frac{H^\alpha}{h}\right)^\beta. \quad (6.9)$$

We start with the case $p > d/(d-1)$. By Proposition 2.9,

$$\mathbb{E}[\mathbb{W}^p(|\mathcal{G}|_\rho + \mu^{\mathcal{B}}, Z_\rho) I_A] \lesssim \mathbb{E}\left[\frac{|\mathcal{B}|^{1+\frac{p}{d}}}{|\mathcal{G}|^{\frac{p}{d}}} I_A\right] \lesssim h^{-\frac{p}{d}} \mathbb{E}[|\mathcal{U}|^{1+\frac{p}{d}} I_A] \stackrel{(6.6)}{\lesssim} h^{1-\frac{p}{d}} \frac{H^{1+\frac{p}{d}}}{h}.$$

This proves (6.9) in this case with $\alpha = 1 + p/d$ and $\beta = 1$.

If now $p < d/(d-1) < 2$, we use Jensen's inequality (2.14) to obtain

$$\begin{aligned} \mathbb{E}[\mathbb{W}^p(|\mathcal{G}|_\rho + \mu^{\mathcal{B}}, Z_\rho) I_A] &\lesssim \mathbb{E}\left[Z^{1-\frac{p}{2}} (\mathbb{W}^2(|\mathcal{G}|_\rho + \mu^{\mathcal{B}}, Z_\rho) I_A)^{\frac{p}{2}}\right] \\ &\leq \mathbb{E}[Z]^{1-\frac{p}{2}} \mathbb{E}[\mathbb{W}^2(|\mathcal{G}|_\rho + \mu^{\mathcal{B}}, Z_\rho) I_A]^{\frac{p}{2}}. \end{aligned}$$

Recalling (6.4) we find $\mathbb{E}[Z] \lesssim h + H \lesssim h$. Using finally (6.9) with $p = 2$ we conclude that

$$\mathbb{E}[\mathbb{W}^p(|\mathcal{G}|_\rho + \mu^{\mathcal{B}}, Z_\rho) I_A] \lesssim h^{1-\frac{p}{2}} \left(h^{1-\frac{2}{d}} \left(\frac{H^\alpha}{h}\right)^\beta\right)^{\frac{p}{2}} = h^{1-\frac{p}{d}} \left(\frac{H^\alpha}{h}\right)^{\beta \frac{p}{2}}.$$

This proves (6.9) also in this case. \square

We now consider the case when the moment bounds for \mathcal{U} and \mathcal{V} are only valid after restricting on a Whitney-type decomposition from Lemma 2.1.

Proposition 6.4. *Let $d \geq 3$, $p \geq 1$ and $\Omega \subseteq \mathbb{R}^d$ be a bounded connected open set with Lipschitz boundary and such that (2.8) holds. Fix a Whitney partition $\mathcal{Q} = (Q_i)_i$, and for $\delta > 0$ let $(\Omega_k)_{k=1}^K = \mathcal{Q}_\delta \cup \mathcal{R}_\delta$ be given by Lemma 2.1. Let finally ρ be a Hölder continuous probability density on Ω , bounded above and below.*

Then, there exist $\alpha = \alpha(p, d) > 0$ and $\beta = \beta(p, d) > 0$ such that the following holds. For every $\eta \in (0, 1)$, $\varepsilon > 0$ and $\gamma \in (0, 1/d)$, there exists $C(\eta, \varepsilon, \gamma)$ such that for every Poisson point processes $\mathcal{N}^{\eta\rho}, \mathcal{M}^{\eta\rho}$ with intensity $\eta\rho$ and every point processes \mathcal{U} and \mathcal{V} on Ω such that

$$\mathbb{E}[|\mathcal{U}_{\Omega_k}|^q + |\mathcal{V}_{\Omega_k}|^q] \lesssim_q (n|\Omega_k|)^{\frac{q}{2}} \quad \forall q > 0, \quad (6.10)$$

if $\delta = n^{-\gamma}$ then

$$\begin{aligned} n^{\frac{p}{d}-1} \mathbb{E}[\mathbb{M}^p(\mathcal{U} \cup \mathcal{N}^{\eta\rho}, \mathcal{V} \cup \mathcal{M}^{\eta\rho})] &\lesssim \eta^{1-\frac{p}{d}} \\ &+ C(\eta, \varepsilon, \gamma) n^\varepsilon \left(\left(\max \left\{ n^{\frac{p}{d}} \delta^{p+1}, n^{\frac{2}{d}} \delta^3 \right\} \right)^\alpha + \left(n\delta^d \right)^{-\beta} \right). \end{aligned}$$

Proof. Using the notation from the beginning of this section, we start as above from (6.3) and (6.5) and estimate by (6.2),

$$\mathbb{E} [\mathbb{W}^p (\mu^{\mathcal{G}}, |\mathcal{G}| \rho)] \lesssim \mathbb{E} [|\mathcal{G}|]^{1-\frac{p}{d}}.$$

Since $|\mathcal{G}| \leq |\mathcal{N}^{\eta n \rho}|$ we get

$$\mathbb{E} [\mathbb{W}^p (\mu^{\mathcal{G}}, |\mathcal{G}| \rho)] \lesssim (\eta n)^{1-\frac{p}{d}}.$$

In order to conclude the proof it is thus enough to show

$$n^{\frac{p}{d}-1} \mathbb{E} [\mathbb{W}^p (\mu^{\mathcal{B}} + |\mathcal{G}| \rho, Z \rho)] \leq C(\eta, \varepsilon, \gamma) n^\varepsilon \left(\left(\max \left\{ n^{\frac{p}{d}} \delta^{p+1}, n^{\frac{2}{d}} \delta^3 \right\} \right)^\alpha + (n \delta^d)^{-\beta} \right). \quad (6.11)$$

Step 1. Reduction to a “good” event. We let

$$A = \{|\mathcal{G}| \in [\eta n/2, 3\eta n]\} \cap \bigcap_{k=1}^K \left\{ \max \{|\mathcal{U}_{\Omega_k}|, |\mathcal{V}_{\Omega_k}|\} \leq (n|\Omega_k|)^{\frac{1}{2}} \cdot n^\varepsilon \right\}.$$

and claim that

$$n^{\frac{p}{d}-1} \mathbb{E} [\mathbb{W}^p (\mu^{\mathcal{B}} + |\mathcal{G}| \rho, Z \rho) I_{A^c}] \leq C(\eta, \varepsilon, \gamma) (n \delta^d)^{-\beta}. \quad (6.12)$$

We first prove that for every $q > 0$,

$$\mathbb{P}[A^c] \lesssim_q C(\eta) n^{-q} + \delta^{1-d} n^{-\varepsilon q}. \quad (6.13)$$

To prove this we use a union bound and split

$$\begin{aligned} \mathbb{P}[A^c] &\leq \mathbb{P}[|\mathcal{G}| \notin [\eta n/2, 3\eta n]] \\ &\quad + \sum_{k=1}^K \mathbb{P} \left[|\mathcal{U}_{\Omega_k}| \geq (n|\Omega_k|)^{\frac{1}{2}} \cdot n^\varepsilon \right] + \sum_{k=1}^K \mathbb{P} \left[|\mathcal{V}_{\Omega_k}| \geq (n|\Omega_k|)^{\frac{1}{2}} \cdot n^\varepsilon \right]. \end{aligned}$$

Regarding the first term we notice that

$$\{|\mathcal{G}| \notin [\eta n/2, 3\eta n]\} \subset \{|\mathcal{N}^{\eta n \rho}| < \eta n/2\} \cup \{|\mathcal{N}^{\eta n \rho}| > 3\eta n\} \cup \{|\mathcal{M}^{\eta n \rho}| < \eta n/2\}.$$

Using once more a union bound and (2.21), we find

$$\mathbb{P}[|\mathcal{G}| \notin [\eta n/2, 3\eta n]] \lesssim_q C(\eta) n^{-q}.$$

Regarding the two sums, by (6.10), we have for every $k \in [1, K]$

$$\mathbb{P} \left[|\mathcal{U}_{\Omega_k}| \geq (n|\Omega_k|)^{\frac{1}{2}} \cdot n^\varepsilon \right] \lesssim_q n^{-\varepsilon q}$$

and similarly for \mathcal{V} . Since $K \lesssim \delta^{1-d}$ by (2.1) this concludes the proof of (6.13).

We now turn to (6.12). As above by the bound $\mathbb{W}^p (\mu^{\mathcal{B}} + |\mathcal{G}| \rho, Z \rho) \lesssim |\mathcal{U}|$ and Cauchy-Schwarz, we have

$$\mathbb{E} [\mathbb{W}^p (\mu^{\mathcal{B}} + |\mathcal{G}| \rho, Z \rho) I_{A^c}] \lesssim \mathbb{E} [|\mathcal{U}|^2]^{\frac{1}{2}} \mathbb{P}[A^c]^{\frac{1}{2}}.$$

Using once more Cauchy-Schwarz together with (6.10) with $q = 2$ we have

$$\mathbb{E} [|\mathcal{U}|^2]^{\frac{1}{2}} \lesssim K^{\frac{1}{2}} n^{\frac{1}{2}}$$

so that by (6.13) and $K \lesssim \delta^{1-d}$

$$n^{\frac{p}{d}-1} \mathbb{E} [\mathbb{W}^p (\mu^{\mathcal{B}} + |\mathcal{G}| \rho, Z \rho) I_{A^c}] \lesssim_q C(\eta) \delta^{\frac{1}{2}(1-d)} (n^{-q} + \delta^{1-d} n^{-\varepsilon q})^{\frac{1}{2}} n^{\frac{p}{d}-\frac{1}{2}}.$$

Since $\delta = n^{-\gamma}$, this concludes the proof of (6.12) provided we choose q large enough depending on ε and γ .

In the remaining two steps we prove that in A ,

$$n^{\frac{p}{d}-1} \mathbb{W}^p(\mu^{\mathcal{B}} + |\mathcal{G}|_\rho, Z\rho) \lesssim_\eta n^\varepsilon \left(\left(\max \left\{ n^{\frac{p}{d}} \delta^{p+1}, n^{\frac{2}{d}} \delta^3 \right\} \right)^\alpha + (n\delta^d)^{-\beta} \right). \quad (6.14)$$

After taking expectation and in combination with (6.12) this would conclude the proof of (6.11). From this point all the estimates are deterministic.

Step 2. Estimate for $p > d/(d-1)$. We first use (2.15), e.g. with $\varepsilon = 1$, to obtain

$$\mathbb{W}_\Omega^p(\mu^{\mathcal{B}} + |\mathcal{G}|_\rho, Z\rho) \lesssim \sum_{k=1}^K \mathbb{W}_{\Omega_k}^p(\mu^{\mathcal{B}} + |\mathcal{G}|_\rho, \alpha_k \rho) + \mathbb{W}_\Omega^p \left(\sum_{k=1}^K \alpha_k I_{\Omega_k} \rho, Z\rho \right), \quad (6.15)$$

with

$$\alpha_k = \frac{\mu^{\mathcal{B}}(\Omega_k)}{\rho(\Omega_k)} + |\mathcal{G}|. \quad (6.16)$$

We bound the terms in the right-hand side separately. For the sum of ‘‘local’’ terms, we estimate differently according to $\Omega_k \in \mathcal{R}_\delta$ or $\Omega_k \in \mathcal{Q}_\delta$. In the first case we use the naive bound

$$\begin{aligned} \mathbb{W}_{\Omega_k}^p(\mu^{\mathcal{B}} + |\mathcal{G}|_\rho, \alpha_k \rho) &\stackrel{(2.13)}{\leq} \mathbb{W}_{\Omega_k}^p \left(\mu^{\mathcal{B}}, \frac{\mu^{\mathcal{B}}(\Omega_k)}{\rho(\Omega_k)} \rho \right) \stackrel{(2.17)}{\leq} \text{diam}(\Omega_k)^p |\mathcal{U}_{\Omega_k}| \\ &\lesssim n^{\frac{1}{2}+\varepsilon} \delta^{p+\frac{d}{2}}. \end{aligned}$$

Since $K \lesssim \delta^{1-d}$ we find

$$n^{\frac{p}{d}-1} \sum_{\Omega_k \in \mathcal{R}_\delta} \mathbb{W}_{\Omega_k}^p(\mu^{\mathcal{B}} + |\mathcal{G}|_\rho, \alpha_k \rho) \lesssim n^\varepsilon n^{\frac{p}{d}} \delta^{1+p} (n\delta^{-d})^{-\frac{1}{2}} \leq n^\varepsilon n^{\frac{p}{d}} \delta^{1+p}. \quad (6.17)$$

If $\Omega_k \in \mathcal{Q}_\delta$ is a cube, we use instead Proposition 2.9 with $\mu^{\mathcal{B}}$ instead of μ and $|\mathcal{G}|$ instead of h , so that

$$\begin{aligned} \mathbb{W}_{\Omega_k}^p(\mu^{\mathcal{B}} + |\mathcal{G}|_\rho, \alpha_k \rho) &\lesssim_\eta \frac{\mu^{\mathcal{B}}(\Omega_k)^{1+\frac{p}{d}}}{n^{\frac{p}{d}}} \lesssim_\eta n^{-\frac{p}{d}} |\mathcal{U}_{\Omega_k}|^{1+\frac{p}{d}} \\ &\lesssim_\eta n^{(1+\frac{p}{d})\varepsilon} n^{\frac{1}{2}(1-\frac{p}{d})} |\Omega_k|^{\frac{1}{2}(1+\frac{p}{d})}. \end{aligned}$$

Summing this inequality yields

$$\begin{aligned} n^{\frac{p}{d}-1} \sum_{\Omega_k \in \mathcal{Q}_\delta} \mathbb{W}_{\Omega_k}^p(\mu^{\mathcal{B}} + |\mathcal{G}|_\rho, \alpha_k \rho) &\lesssim_\eta n^{(1+\frac{p}{d})\varepsilon} n^{-\frac{1}{2}(1-\frac{p}{d})} \sum_{k=1}^K |\Omega_k|^{\frac{1}{2}(1+\frac{p}{d})} \\ &\stackrel{(2.1)}{\lesssim_\eta} n^{2\varepsilon} n^{-\frac{1}{2}(1-\frac{p}{d})} \max \left\{ 1, \delta^{\frac{1}{2}(d-2-p)} \right\}. \end{aligned}$$

Notice that since $n\delta^d \geq 1$,

$$n^{-\frac{1}{2}(1-\frac{p}{d})} \max \left\{ 1, \delta^{\frac{1}{2}(d-2-p)} \right\} \leq (n\delta^d)^{-\frac{1}{2}(1-\frac{p}{d})}$$

so that

$$n^{\frac{p}{d}-1} \sum_{\Omega_k \in \mathcal{Q}_\delta} \mathbb{W}_{\Omega_k}^p(\mu^{\mathcal{B}} + |\mathcal{G}|_\rho, \alpha_k \rho) \lesssim_\eta n^{2\varepsilon} (n\delta^d)^{-\frac{1}{2}(1-\frac{p}{d})}. \quad (6.18)$$

We then consider the last term in (6.15). Using Lemma 2.8 with $Z\rho \gtrsim \eta n$ in place of λ (recall that we assume here that A holds), we get

$$\mathbb{W}_\Omega^p \left(\sum_{k=1}^K \alpha_k I_{\Omega_k} \rho, Z\rho \right) \lesssim_\eta n^{1-p} \left\| \sum_{k=1}^K \alpha_k I_{\Omega_k} \rho - Z\rho \right\|_{W^{-1,p}(\Omega)}^p. \quad (6.19)$$

Recalling (6.16) and that $Z = \mu^{\mathcal{B}}(\Omega) + |\mathcal{G}|$, we can rewrite

$$\sum_{k=1}^K \alpha_k I_{\Omega_k} \rho - Z\rho = \sum_{k=1}^K \frac{\mu^{\mathcal{B}}(\Omega_k)}{\rho(\Omega_k)} (I_{\Omega_k} - \rho(\Omega_k)) \rho.$$

By (2.9) of Lemma 2.6 with $h = n^{1+2\varepsilon}$ we thus have in A

$$\left\| \sum_{k=1}^K \alpha_k I_{\Omega_k} \rho - Z\rho \right\|_{W^{-1,p}(\Omega)} \lesssim n^\varepsilon \delta^{1-\frac{d}{2}} n^{\frac{1}{2}}.$$

Combining this with (6.19) we get that in A ,

$$n^{\frac{p}{d}-1} \mathbb{W}_\Omega^p \left(\sum_{k=1}^K \alpha_k I_{\Omega_k} \rho, Z\rho \right) \lesssim n^{p\varepsilon} (n\delta^d)^{-p\frac{(d-2)}{2d}}.$$

Inserting this estimate, (6.17) and (6.18) in (6.15) we finally obtain (notice that $p(d-2) > d-p$ for $p > d/(d-1)$) that in A ,

$$n^{\frac{p}{d}-1} \mathbb{W}^p(\mu^{\mathcal{B}} + |\mathcal{G}| \rho, Z\rho) \lesssim_\eta n^{\max\{p,2\}\varepsilon} \left(n^{\frac{p}{d}} \delta^{1+p} + (n\delta^d)^{-\frac{(d-p)}{2d}} \right).$$

Up to replacing ε by $\max\{p,2\}\varepsilon$, this concludes the proof of (6.14) if $p > d/(d-1)$.

Step 3. Estimate for $p \leq d/(d-1)$. Since $2 > d/(d-1) \geq p$, we may use Jensen's inequality (2.14) to infer that in A ,

$$\begin{aligned} n^{\frac{p}{d}-1} \mathbb{W}^p(\mu^{\mathcal{B}} + |\mathcal{G}| \rho, Z\rho) &\leq n^{\frac{p}{d}-1} Z^{1-\frac{p}{2}} \left(\mathbb{W}^2(\mu^{\mathcal{B}} + |\mathcal{G}| \rho, Z\rho) \right)^{\frac{p}{2}} \\ &\lesssim (\eta + n^{-\frac{1}{2}})^{1-\frac{p}{2}} \left(n^{\frac{2}{d}-1} \mathbb{W}^2(\mu^{\mathcal{B}} + |\mathcal{G}| \rho, Z\rho) \right)^{\frac{p}{2}} \\ &\lesssim \left(n^{\frac{2}{d}-1} \mathbb{W}^2(\mu^{\mathcal{B}} + |\mathcal{G}| \rho, Z\rho) \right)^{\frac{p}{2}}. \end{aligned}$$

Using (6.14) for $p = 2$ concludes the proof of (6.14) also in this case. \square

Finally, we consider the case of a cube Q_{mL} decomposed into cubes of sidelength L . The difficulty compared to the previous two cases is to obtain bounds which are independent of m . This is achieved using the additional independence for the point processes \mathcal{U}, \mathcal{V} . While we believe that a direct proof combining Green kernel bounds in the spirit of the proof of Lemma 2.6 together with a Rosenthal type inequality for the (non independent) random variables $\mu^{\mathcal{B}}(Q_i)$ should be possible we give a more elementary proof based on subadditivity and concentration.

Proposition 6.5. *Let $d \geq 3$, $\eta \in (0, 1/2)$, $L \geq 1$ and $m \in \mathbb{N} \setminus \{0\}$. Let \mathcal{U}, \mathcal{V} be point processes on Q_{mL} such that the restrictions $(\mathcal{U}_{Q_i}, \mathcal{V}_{Q_i})_i$ on all sub-cubes $Q_i = Q_L + Lz_i \subseteq Q_{mL}$, with $z_i \in \mathbb{Z}^d$, are independent copies (translated by the vector Lz_i) of the pair of processes $(\mathcal{U}_{Q_L}, \mathcal{V}_{Q_L})$ and such that for every $q \geq 1$, there exists $C(q) > 0$ such that*

$$\mathbb{E} [|\mathcal{U}_{Q_i}|^q + |\mathcal{V}_{Q_i}|^q] \leq C(q) L^{\frac{q}{2}}. \quad (6.20)$$

Let $\mathcal{N}^\eta, \mathcal{M}^\eta$ independent Poisson processes on Q_{mL} with constant intensity η , also independent from $(\mathcal{U}, \mathcal{V})$. Then, for every $p \in [1, d)$, there exists $C(\eta) = C(\eta, p, d, (C(q))_{q \geq 1}) > 0$ and $\alpha = \alpha(p, d) > 0$, such that, if $L \geq C(\eta)$,

$$\mathbb{E} \left[\frac{1}{|Q_{mL}|} \mathbb{M}^p(\mathcal{U} \cup \mathcal{N}^\eta, \mathcal{V} \cup \mathcal{M}^\eta) \right] \lesssim \eta^{1-\frac{p}{d}} + \frac{C(\eta)}{L^\alpha}.$$

Remark 6.6. Let us preliminarily notice that, for any $R \subseteq Q_{mL}$ that is the disjoint union of k cubes among the cubes $Q_i = Q_L + Lz_i$, $z_i \in \mathbb{Z}^d$, we have the upper bound, if $q \geq 1$,

$$\begin{aligned} \mathbb{E}[|\mathcal{U}_R|^q] &= k^q \mathbb{E} \left[\left(\frac{1}{k} \sum_{Q_i \subseteq R} |\mathcal{U}_{Q_i}| \right)^q \right] \leq k^q \mathbb{E} \left[\frac{1}{k} \sum_{Q_i \subseteq R} |\mathcal{U}_{Q_i}|^q \right] \\ &\lesssim k^q L^{d\frac{q}{2}} \lesssim (kL^d)^q = |R|^q. \end{aligned} \quad (6.21)$$

In particular, we have

$$\mathbb{E}[|\mathcal{U}|^q] \lesssim m^{dq} L^{d\frac{q}{2}} \lesssim |Q_{mL}|^q. \quad (6.22)$$

Moreover, by Rosenthal inequalities [39], if $q \geq 2$,

$$\begin{aligned} \mathbb{E}[||\mathcal{U}_R| - \mathbb{E}[|\mathcal{U}_R|]|^q] &= \mathbb{E} \left[\left| \sum_{Q_i \subseteq R} (|\mathcal{U}_{Q_i}| - \mathbb{E}[|\mathcal{U}_{Q_i}|]) \right|^q \right] \\ &\lesssim k \mathbb{E}[||\mathcal{U}_{Q_L}| - \mathbb{E}[|\mathcal{U}_{Q_L}|]|^q] + k^{\frac{q}{2}} \mathbb{E}[||\mathcal{U}_{Q_L}| - \mathbb{E}[|\mathcal{U}_{Q_L}|]|^2]^{\frac{q}{2}} \\ &\lesssim kL^{d\frac{q}{2}} + k^{\frac{q}{2}} L^{d\frac{q}{2}} \lesssim |R|^{\frac{q}{2}}. \end{aligned}$$

We will use all these bounds in the proof below.

Proof of Proposition 6.5. For simplicity, we write throughout the proof Q instead of Q_{mL} . As in the previous two proofs, we start from (6.3) and (6.5) (with $\rho = I_Q/|Q|$) and estimate by (6.2), see also Remark 6.2,

$$\mathbb{E} \left[\frac{1}{|Q|} W_Q^p \left(\mu^{\mathcal{G}}, \frac{|\mathcal{G}|}{|Q|} \right) \right] \lesssim \mathbb{E}[|\mathcal{G}|^{1-\frac{p}{d}} |Q|^{\frac{p}{d}-1}].$$

Since $|\mathcal{G}| \leq |\mathcal{N}^\eta|$ we get

$$\mathbb{E} \left[\frac{1}{|Q|} W_Q^p \left(\mu^{\mathcal{G}}, \frac{|\mathcal{G}|}{|Q|} \right) \right] \lesssim \eta^{1-\frac{p}{d}}.$$

In order to conclude the proof it is thus enough to show

$$\mathbb{E} \left[\frac{1}{|Q|} W_Q^p \left(\frac{|\mathcal{G}|}{|Q|} + \mu^{\mathcal{B}}, \frac{Z}{|Q|} \right) \right] \lesssim \frac{C(\eta)}{L^\alpha}. \quad (6.23)$$

We split the proof into several steps. We first consider the case $p \geq 2 \geq d/(d-1)$.

Step 1. Concentration bounds for $\mu^{\mathcal{B}}$. In this intermediate step, we collect some facts about $\mu^{\mathcal{B}}(R)$, where $R \subseteq Q$ is a disjoint union of k cubes $Q_i = Q_L + Lz_i$, $z_i \in \mathbb{Z}^d$. First of all, the construction of \mathcal{B} ensures that $\mathbb{E}[\mu^{\mathcal{B}}(Q_i)]$ does not depend on Q_i (one could in fact prove that $(\mu^{\mathcal{B}}(Q_i))_i$ is an exchangeable sequence). We deduce that

$$\mathbb{E}[\mu^{\mathcal{B}}(Q_i)] = \frac{\mathbb{E}[\mu^{\mathcal{B}}(Q)]}{m^d}, \quad \text{hence} \quad \frac{\mathbb{E}[\mu^{\mathcal{B}}(R)]}{|R|} = \frac{\mathbb{E}[\mu^{\mathcal{B}}(Q)]}{|Q|}. \quad (6.24)$$

Indeed, when conditioned on $Z = z$, $|\mathcal{N}^\eta| = n$, $|\mathcal{U}| = u_Q$, and the number of points $|\mathcal{U}_R| = u_R \leq u_Q$, $\mu^{\mathcal{B}}(R)$ is the number of ‘‘successes’’ in the random sampling procedure, without replacement which we used to define \mathcal{B} , with $b = \max\{z - n, 0\}$ draws from an urn containing u_Q marbles, u_R of which have the desired feature (their extraction defines a success). This is explicitly given by a hypergeometric distribution with parameters (u_Q, u_R, b) : given $s_R \leq u_R$,

$$\mathbb{P}(\mu^{\mathcal{B}}(R) = s_R | B) = \binom{u_R}{s_R} \binom{u_Q - u_R}{b - s_R} / \binom{u_Q}{b},$$

where for brevity we write

$$B = \{Z = z, |\mathcal{N}^\eta| = n_Q, |\mathcal{U}| = u_Q, |\mathcal{U}_R| = u_R\}.$$

Specializing to $R = Q_i$, we see that this quantity does not depend on Q_i , since $|\mathcal{U}_{Q_i}|$ are i.i.d. variables, hence the joint laws of the variables $(Z, |\mathcal{N}^\eta|, |\mathcal{U}|, |\mathcal{U}_{Q_i}|)$ involved in the definition of the law of $\mu^{\mathcal{B}}(Q_i)$ do not depend on i .

Using the concentration inequality (2.23) for hypergeometric random variables, we have

$$\mathbb{E} \left[\left| \mu^{\mathcal{B}}(R) - \mathbb{E} [\mu^{\mathcal{B}}(R)|B] \right|^p |B \right] \lesssim (u_R)^{\frac{p}{2}},$$

from which we find, thanks to (6.21) (recall that $p \geq 2$),

$$\mathbb{E} \left[\left| \mu^{\mathcal{B}}(R) - \mathbb{E} [\mu^{\mathcal{B}}(R)] \right|^p \right] \lesssim |R|^{\frac{p}{2}}. \quad (6.25)$$

Step 2. Subadditivity bound. Using (6.24) and (6.25) above, we are in a position to follow closely the main argument of [24, Proposition 5.4]. We define, for a rectangle $R \subseteq Q$ that is a union of cubes Q_i 's,

$$f(R) = \mathbb{E} \left[\frac{1}{|R|} \mathbb{W}_R^p \left(\mu^{\mathcal{B}} + \frac{|\mathcal{G}|}{|Q|}, \frac{\mu^{\mathcal{B}}(R)}{|R|} + \frac{|\mathcal{G}|}{|Q|} \right) \right].$$

We say that \mathcal{R} is an admissible partition of R if it is made of rectangles satisfying the following conditions. Each $R_k \in \mathcal{R}$ is a union of cubes Q_i , it is of moderate aspect ratio and $3^{-d}|R| \leq |R_j| \leq |R|$. We claim that there exists $C_\eta = C(d, p, \eta) > 0$ such that for every admissible partition \mathcal{R} of R and every $\varepsilon \in (0, 1)$, we have

$$f(R) \leq (1 + \varepsilon) \sum_i \frac{|R_i|}{|R|} f(R_i) + \frac{C_\eta}{\varepsilon^{p-1}} \frac{1}{|R|^{\frac{p(d-2)}{2d}}}. \quad (6.26)$$

Setting

$$\alpha = \frac{\mu^{\mathcal{B}}(R)}{|R|} + \frac{|\mathcal{G}|}{|Q|}, \quad \alpha_i = \frac{\mu^{\mathcal{B}}(R_i)}{|R_i|} + \frac{|\mathcal{G}|}{|Q|}$$

and using (2.15), this reduces to

$$\mathbb{E} \left[\frac{1}{|R|} \mathbb{W}_R^p \left(\sum_i \alpha_i I_{R_i}, \alpha \right) \right] \leq \frac{C_\eta}{|R|^{\frac{p(d-2)}{2d}}}. \quad (6.27)$$

First, we single out the event

$$A = \{ \min \{ |\mathcal{N}^\eta|, |\mathcal{M}^\eta| \} \geq \eta |Q| / 2 \}.$$

Notice that on A , we have $\alpha \gtrsim \eta$. By the concentration bound (2.21), for every $q \geq 1$, $\mathbb{P}(A^c) \lesssim_q (\eta |Q|)^{-q} \leq (\eta |R|)^{-q}$. Therefore, if A^c holds, we can use the trivial bound

$$\begin{aligned} \frac{1}{|R|} \mathbb{W}_R^p \left(\sum_i \alpha_i I_{R_i}, \alpha \right) &\leq \frac{1}{|R|} \mathbb{W}_R^p \left(\sum_i \frac{\mu^{\mathcal{B}}(R_i)}{|R_i|} I_{R_i}, \frac{\mu^{\mathcal{B}}(R)}{|R|} \right) \\ &\leq |R|^{\frac{p}{d}-1} \mu^{\mathcal{B}}(R) \leq |R|^{\frac{p}{d}-1} |\mathcal{U}_R|. \end{aligned}$$

Using Cauchy-Schwarz inequality and (6.21) with $q = 2$, we get for any $q \geq 1$,

$$\mathbb{E} \left[\frac{1}{|R|} \mathbb{W}_R^p \left(\sum_i \alpha_i I_{R_i}, \alpha \right) I_{A^c} \right] \lesssim_{\eta, q} |R|^{\frac{p}{d}-q},$$

which is estimated by the right-hand side of (6.27) provided we choose q large enough.

If A holds, we use (2.16) in combination with (2.7) (recall that for rectangles of moderate aspect ratio the Sobolev constant is uniformly bounded) to get

$$\begin{aligned} \frac{1}{|R|} \mathbb{W}_R^p \left(\sum_i \alpha_i I_{R_i}, \alpha \right) &\lesssim \frac{|R|^{\frac{p}{d}-1}}{\alpha^{p-1}} \sum_i |R_i| |\alpha_i - \alpha|^p \\ &\lesssim \eta^{1-p} |R|^{\frac{p}{d}} \sum_i |\alpha_i - \alpha|^p. \end{aligned}$$

We thus have

$$\begin{aligned} \mathbb{E} \left[\frac{1}{|R|} \mathbb{W}_R^p \left(\sum_i \alpha_i I_{R_i}, \alpha \right) I_A \right] &\lesssim \eta^{1-p} |R|^{\frac{p}{d}} \sum_i \mathbb{E} [|\alpha_i - \alpha|^p I_A] \\ &\leq \eta^{1-p} |R|^{\frac{p}{d}} \sum_i \mathbb{E} [|\alpha_i - \alpha|^p]. \end{aligned}$$

Using that $\alpha_i - \alpha = \frac{\mu^{\mathcal{B}}(R_i)}{|R_i|} - \frac{\mu^{\mathcal{B}}(R)}{|R|}$, (6.24) and triangle inequality we have

$$\begin{aligned} \sum_i \mathbb{E} [|\alpha_i - \alpha|^p] &\lesssim \sum_i \mathbb{E} \left[\left| \frac{\mu^{\mathcal{B}}(R_i)}{|R_i|} - \mathbb{E} \left[\frac{\mu^{\mathcal{B}}(R_i)}{|R_i|} \right] \right|^p \right] + \mathbb{E} \left[\left| \frac{\mu^{\mathcal{B}}(R)}{|R|} - \mathbb{E} \left[\frac{\mu^{\mathcal{B}}(R)}{|R|} \right] \right|^p \right] \\ &\stackrel{(6.25)}{\lesssim} |R|^{-\frac{p}{2}}. \end{aligned}$$

This proves

$$\mathbb{E} \left[\frac{1}{|R|} \mathbb{W}_R^p \left(\sum_i \alpha_i I_{R_i}, \alpha \right) I_A \right] \lesssim \frac{\eta^{1-p}}{|R|^{\frac{p(d-2)}{2d}}},$$

concluding the proof of (6.27).

Step 3. Dyadic approximation. Starting from the cube $Q = Q_{mL}$, we build a sequence of finer and finer partitions of Q_{mL} by rectangles of moderate aspect ratios that are unions of sub-cubes Q_i 's. We let $\mathcal{R}_0 = \{Q_{mL}\}$ and define \mathcal{R}_k inductively as follows. Let $R \in \mathcal{R}_k$. Up to translation we may assume that $R = \prod_{i=1}^d (0, m_i L)$ for some $m_i \in \mathbb{N}$. We then split each interval $(0, m_i L)$ into $(0, \lfloor \frac{m_i}{2} \rfloor L) \cup (\lfloor \frac{m_i}{2} \rfloor L, m_i L)$. It is readily seen that this induces an admissible partition of R . Let us point out that when $m_i = 1$ for some i , the corresponding interval $(0, \lfloor \frac{m_i}{2} \rfloor L)$ is empty. This procedure stops after a finite number of steps K once $\mathcal{R}_K = \{Q_L + z_i, z_i \in [0, m-1]^d \cap \mathbb{Z}^d\}$. It is also readily seen that $2^{K-1} < m \leq 2^K$ and that for every $k \in [0, K]$ and every $R \in \mathcal{R}_k$ we have $|R| \sim (2^{K-k} L)^d$.

We prove via a downward induction the existence of $\Lambda_\eta > 0$ such that for every $k \in [0, K]$ and every $R \in \mathcal{R}_k$,

$$f(R) \leq f(Q_L) + \Lambda_\eta (1 + f(Q_L)) L^{-\frac{d-2}{2}} \sum_{j=K-k}^K 2^{-j \frac{d-2}{2}}. \quad (6.28)$$

The statement is clearly true for $k = K$, since the law of the point process on each cube $Q_i = Q_L + z_i$ is the same, hence $f(Q_i) = f(Q_L)$. Assume that it holds true for $k+1$. Let

$R \in \mathcal{R}_k$. Applying (6.26) with $\varepsilon = (2^{K-k}L)^{-(d-2)/2} \ll 1$, we get

$$\begin{aligned}
 f(R) &\leq (1 + \varepsilon) \sum_{R_i \in \mathcal{R}_{k+1}, R_i \subset R} \frac{|R_i|}{|R|} f(R_i) + \frac{C_\eta}{\varepsilon^{p-1}} \frac{1}{|R|^{\frac{p(d-2)}{2d}}} \\
 &\stackrel{(6.28)}{\leq} (1 + \varepsilon) \left(f(Q_L) + \Lambda_\eta(1 + f(Q_L))L^{-\frac{d-2}{2}} \sum_{j=K-k+1}^K 2^{-j\frac{d-2}{2}} \right) \\
 &\quad + C_\eta(2^{K-k}L)^{-\frac{d-2}{2}} \\
 &\leq f(Q_L) + \Lambda_\eta(1 + f(Q_L))L^{-\frac{d-2}{2}} \\
 &\quad \cdot \left[\sum_{j=K-k+1}^K 2^{-j\frac{d-2}{2}} + 2^{-(K-k)\frac{d-2}{2}} \left(\frac{C_\eta + 1}{\Lambda_\eta} + L^{-\frac{d-2}{2}} \sum_{j=K-k+1}^K 2^{-j\frac{d-2}{2}} \right) \right].
 \end{aligned}$$

If L is large enough (depending on η) then

$$\left(\sum_{j=K-k+1}^K 2^{-j\frac{d-2}{2}} \right) \eta^{1-3p} L^{-\frac{(d-2)}{2}} \lesssim_\eta \left(\sum_{j=0}^{\infty} 2^{-j\frac{d-2}{2}} \right) L^{-\frac{(d-2)}{2}} \leq \frac{1}{2}.$$

Finally, choosing $\Lambda_\eta \geq 2(C+1)$ yields (6.28). Applying (6.28) to $R = Q_{mL}$ and using that $\sum_{j \geq 0} 2^{-j\frac{d-2}{2}} < \infty$, we get

$$f(Q_{mL}) \leq f(Q_L) + \Lambda_\eta(1 + f(Q_L)) \frac{1}{L^{\frac{d-2}{2}}}. \quad (6.29)$$

Step 4. Conclusion in the case $p \geq 2$. We finally claim that

$$f(Q_L) \leq \frac{C_\eta}{L^{\frac{1}{2}(d-p)}}. \quad (6.30)$$

Arguing verbatim as in the proof of (6.27) of *Step 5*, we see that it is enough to assume that we are in the event $A = \{\min\{|\mathcal{N}^\eta|, |\mathcal{M}^\eta|\} \geq \eta|Q|/2\}$. Since in this case $|\mathcal{G}|/|Q| \gtrsim \eta$, Proposition 2.9 yields

$$\begin{aligned}
 \mathbb{E} \left[\frac{1}{|Q_L|} W_{Q_L}^p \left(\mu^{\mathcal{B}} + \frac{|\mathcal{G}|}{|Q|}, \frac{\mu^{\mathcal{B}}(Q_L)}{|Q_L|} + \frac{|\mathcal{G}|}{|Q|} \right) I_A \right] &\lesssim \frac{1}{|Q_L| \eta^{\frac{p}{d}}} \mathbb{E} \left[(\mu^{\mathcal{B}}(Q_L))^{1+\frac{p}{d}} I_A \right] \\
 &\leq \frac{1}{|Q_L| \eta^{\frac{p}{d}}} \mathbb{E} \left[(|\mathcal{U}_{Q_L}|)^{1+\frac{p}{d}} \right] \\
 &\stackrel{(6.20)}{\lesssim} \frac{L^{\frac{d+p}{2}}}{L^d \eta^{\frac{p}{d}}} \lesssim \frac{1}{\eta^{\frac{p}{d}} L^{\frac{d-p}{2}}}.
 \end{aligned}$$

This proves (6.30). Inserting this into (6.29) finally gives (recall that $p > 2$)

$$f(Q) \leq \frac{C_\eta}{L^{\frac{1}{2}(d-p)}}.$$

This concludes the proof of (6.23) with $\alpha = (d-p)/2$ when $p \geq 2$.

Step 5. The case $p \leq 2$. If $p \leq 2$, we argue as in the previous two proofs and use (2.14) to obtain (recall that $Z = |\mathcal{G}| + |\mathcal{B}| \leq |\mathcal{U}| + |\mathcal{N}^\eta|$)

$$\begin{aligned} \mathbb{E} \left[\frac{1}{|Q|} \mathcal{W}_Q^p \left(\frac{|\mathcal{G}|}{|Q|} + \mu^{\mathcal{B}}, \frac{Z}{|Q|} \right) \right] &\leq \mathbb{E} \left[\left(\frac{Z}{|Q|} \right)^{1-\frac{p}{2}} \left(\frac{1}{|Q|} \mathcal{W}_Q^2 \left(\frac{|\mathcal{G}|}{|Q|} + \mu^{\mathcal{B}}, \frac{Z}{|Q|} \right) \right)^{\frac{p}{2}} \right] \\ &\leq \left(\frac{\mathbb{E}[Z]}{|Q|} \right)^{1-\frac{p}{2}} \mathbb{E} \left[\frac{1}{|Q|} \mathcal{W}_Q^2 \left(\frac{|\mathcal{G}|}{|Q|} + \mu^{\mathcal{B}}, \frac{Z}{|Q|} \right) \right]^{\frac{p}{2}} \\ &\lesssim \left(L^{-\frac{d}{2}} + \eta \right)^{1-\frac{p}{2}} \left(\frac{C(\eta)}{L^\alpha} \right)^{\frac{p}{2}} \lesssim \left(\frac{C(\eta)}{L^\alpha} \right)^{\frac{p}{2}}, \end{aligned}$$

where in the last step we used (6.22) and (6.23) with $p = 2$. This concludes the proof of (6.23) for any $p < d$. \square

M.G.: CMAP, CNRS, ÉCOLE POLYTECHNIQUE, INSTITUT POLYTECHNIQUE DE PARIS, 91120 PALAISEAU, FRANCE

Email address: michael.goldman@cnrs.fr

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

Email address: dario.trevisan@unipi.it