Asymptotics for Random Quadratic Transportation Costs

Martin Huesmann
Universität Münster

Michael Goldman
CNRS and École polytechnique

Dario Trevisan
Università di Pisa

February 17, 2025

Abstract

We establish the validity of asymptotic limits for the general transportation problem between random i.i.d. points and their common distribution, with respect to the squared Euclidean distance cost, in any dimension larger than three. Previous results were essentially limited to the two (or one) dimensional case, or to distributions whose absolutely continuous part is uniform.

The proof relies upon recent advances in the stability theory of optimal transportation, combined with functional analytic techniques and some ideas from quantitative stochastic homogenization. The key tool we develop is a quantitative upper bound for the usual quadratic optimal transportation problem in terms of its boundary variant, where points can be freely transported along the boundary. The methods we use are applicable to more general random measures, including occupation measure of Brownian paths, and may open the door to further progress on challenging problems at the interface of analysis, probability, and discrete mathematics.

Keywords: matching problem, optimal transport, geometric probability, stochastic homogenization.

1 Introduction

The assignment problem (or bipartite matching) is a classic optimization problem that involves finding an optimal correspondence between two sets of objects, such that a total matching cost is minimized. It counts innumerate applications and is subject of intense research by several communities, from operation research and algorithm theory, combinatorics and graph theory [37], probability and statistics [51] and even theoretical physics [40, 39]. In the Euclidean formulation of the problem, the two sets of objects are families of points $(x_i)_{i=1}^n, (y_i)_{i=1}^n \subseteq \mathbb{R}^d$, and the matching cost is defined, for a given parameter $p > 0$, as

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

where $\mathcal{S}_n$ denotes the set of permutations and $|\cdot|$ the Euclidean norm.

By the Birkhoff-Von Neumann Theorem, a natural linear programming reformulation of (1.1) is given by the optimal transport problem [53], where σ is replaced by a transport plan or coupling, and the cost (1.1) becomes a special instance of the Wasserstein cost of order p, denoted by $\mathsf{W}^p(\mu, \lambda)$, between the two empirical measures $\mu := \sum_{i=1}^n \delta_{x_i}$ and $\lambda := \sum_{i=1}^n \delta_{y_i}$ (see Section 3.1 for precise notation). Such "relaxed" formulation allows also for non-discrete measures, and opens the way to more general applications. For example, by letting λ be absolutely continuous and keeping μ discrete over a set of n points, one can interpret $\mathsf{W}^p(\mu, \lambda)$ as a quantization cost [33, 24].

In this context it is then quite natural to assume e.g. that $x_i = X_i$ are randomly sampled from a given probability distribution, e.g., λ normalized to be a probability.

There is in fact a rich literature studying random instances of combinatorial optimization problems in Euclidean spaces [48, 61]. For bipartite problems, like the assignment problem, it is well-known that classical methods encounter limitations due to local fluctuations in the number of samples between the two sets. This can lead to unexpected scaling behaviors of the cost as n grows, as first observed in [1] for uniform i.i.d. samples on the unit square $(0,1)^2$ and $p = 1$. Subsequent contributions by several authors [50, 25, 15, 51, 9, 24, 12, 35, 11, 13] lead to a more complete understanding of the whole picture, for general dimensions d, cost exponents p and sample distributions λ, providing asymptotic upper and lower bounds. We report here the combination of [24, 9, 32] (see also [18]).

Theorem 1.1. *For any $d \geq 3$ and $p \geq 1$, there exist constants $0 < \underline{\mathsf{c}}(p,d) \leq \overline{\mathsf{c}}(p,d) < \infty$ such that the following holds for every probability measure λ on $\mathbb{R}^d$. Assume that either*

(i) $p < d/2$ and λ has finite q-th moment, for some $q > dp/(d-p)$;

(ii) or λ is absolutely continuous with respect to the Lebesgue measure on a bounded connected C^2-smooth or convex domain, with Hölder continuous density uniformly bounded from above and below by a strictly positive constant.

Then denoting by λ_a the absolutely continuous part of λ with respect to the Lebesgue measure,

$$\begin{aligned} \underline{\mathsf{c}}(p,d) \int_{\mathbb{R}^d} \lambda_a^{1-p/d} &\leq \liminf_{n\to\infty} \mathbb{E}\left[\mathsf{W}^p\left(\frac{1}{n}\sum_{i=1}^n \delta_{X_i}, \lambda\right)\right] n^{p/d} \\ &\leq \limsup_{n\to\infty} \mathbb{E}\left[\mathsf{W}^p\left(\frac{1}{n}\sum_{i=1}^n \delta_{X_i}, \lambda\right)\right] n^{p/d} \leq \overline{\mathsf{c}}(p,d) \int_{\mathbb{R}^d} \lambda_a^{1-p/d}, \end{aligned} \tag{1.2}$$

where $(X_i)_{i=1}^n$ are i.i.d. random variables with common distribution λ. Moreover, the constants $0 < \underline{\mathsf{c}}(p,d) \leq \overline{\mathsf{c}}(p,d) < \infty$ are given by

$$\overline{\mathsf{c}}(p,d) = \lim_{L\to\infty} \frac{1}{L^d} \mathbb{E}\left[\mathsf{W}^p_{(0,L)^d}\left(\mu, \frac{\mu((0,L)^d)}{L^d}\right)\right]$$

$$\text{and} \quad \underline{\mathsf{c}}(p,d) = \lim_{L\to\infty} \frac{1}{L^d} \mathbb{E}\left[\mathsf{Wb}^p_{(0,L)^d}\left(\mu, \frac{\mu((0,L)^d)}{L^d}\right)\right],$$

where μ is a Poisson point process with unit intensity on $\mathbb{R}^d$, $\mathsf{W}^p_{(0,L)^d}$ denotes the Wasserstein cost of order p between the measures restricted on the cube $(0,L)^d$, while $\mathsf{Wb}^p_{(0,L)^d}$ denotes instead its "boundary" variant (see Section 3.2 for precise definitions).

The boundary transport problem $\mathsf{Wb}^p_{(0,L)^d}$ was first introduced in the optimal transport literature in [27] to study the heat semigroup with Dirichlet boundary conditions as a gradient flow of the entropy. In the context of random Euclidean bipartite matching, it is proposed in [9] as a natural bipartite counterpart of the "dual" boundary functionals studied in [48, 61].

Let us point out that under the assumptions (i), the statement of Theorem 1.1 does not appear in this precise form in the literature. It can be however easily obtained by combining the proofs of [24, 9]. Indeed, while [9] treats the case of the random Euclidean assignment problem (or bipartite matching), the proofs extend straightforwardly to the case of the matching to the reference measure. By [9, Theorem 5 & Lemma 4] this yields the upper bound in (1.2) for measures with bounded support. Arguing then as in [24, Theorem 4] one can conclude the proof of the upper bound in (1.2) for measures with finite $q-$moment. Notice that this is a stronger statement than [9,

Theorem 6], where more integrability is required, or [24, Theorem 4], where a Riemann integrability condition is imposed on λ_a. Regarding the lower bound in (1.2), it can be proven exactly as in [9, Theorem 8], avoiding once more the Riemann integrability issue from [24, Theorem 4].

In the statement of Theorem 1.1, the cases $d \in \{1, 2\}$ are excluded because of a different asymptotic scaling, see [12, 20, 5, 4]. In the range $p \geq d/2$, the upper bound in (1.2) cannot hold without structural assumptions on λ, see e.g. [28]. See however [18] for some results for Gaussian or Gaussian-like densities (as well as a more general lower bound).

1.1 Main result

The natural question raised by Theorem 1.1 is the existence of the limit in (1.2). It would be of course a consequence of the equality $\underline{\mathsf{c}}(p, d) = \overline{\mathsf{c}}(p, d)$. This is explicitly conjectured in [24, Remark 2], albeit for a smaller constant, see Section 1.3 – but also in the bipartite case in [9]. In this paper we settle this conjecture in the case $p = 2$.

Theorem 1.2. *For every $d \geq 3$, it holds*

$$\underline{\mathsf{c}}(2, d) = \overline{\mathsf{c}}(2, d). \tag{1.3}$$

So far the only results in this direction were obtained in the case $d = p = 2$ [20, 5, 10, 4], where the constant $\overline{\mathsf{c}}(2, 2) = 1/(4\pi)$ can be explicitly computed, and the concave one-dimensional case $d = 1$, $p \in (0, 1/2)$, [52] where the special structure of the solution on the line can be exploited.

Remark 1.3. This problem is more subtle than it may first appear: for example, in [25], it is claimed that the lim inf and lim sup in (1.2) always coincide and the limit equals the right hand side, for $p = 1$ in any dimension $d \geq 3$. However, an error is contained in [25, Lemma 4.2] which cannot be fixed, also according to [9]. Moreover, the validity of (1.3) strongly relies on the stochastic properties of the Poisson point process. Indeed, the analog statement would not be true for the random grid, i.e. replacing μ by $\sum_{x \in \mathbb{Z}^d + Z} \delta_x$ where Z is a uniformly distributed random variable on $(0, 1)^d$. In light of Assumption 4.1, we see that (1.3) fails for the random grid because it is not hyperuniform when tested with with cubes (but it is when tested w.r.t. balls).

We also point out that if one is only interested in the equality between the lim inf and the lim sup in (1.2), when λ_a is constant (and possibly λ has a singular part with respect to the Lebesgue measure) it is still possible to conclude without knowing that $\underline{\mathsf{c}}(p, d) = \overline{\mathsf{c}}(p, d)$ by [9, Theorem 2]. See also [32].

1.2 Comments on the proof technique

The proof of Theorem 1.2 relies on novel tools from the stability theory of optimal transportation, combined with functional analytic tools and some ideas from quantitative stochastic homogenization. Since some fundamental features may be obscured by technical aspects, in this section we provide an overview of the argument in the simplified setting provided by the PDE ansatz from [20], which has been a stimulus for much of the recent progress in the field.

Deterministic bound

The starting point of the heuristics consists in linearizing the Monge-Ampère equation for the optimal transport map, into a Poisson equation, essentially replacing the Wasserstein distance with a Sobolev norm of negative order. Given two measures μ, λ on a bounded domain $\Omega \subseteq \mathbb{R}^d$ (e.g., a cube), we approximate (writing here $\approx$ for a heuristic equivalence, $\lesssim$ for an inequality)

$$\mathsf{W}^2_\Omega(\mu, \lambda) \approx \int_\Omega |\nabla f|^2,$$

where f solves the following Poisson equation, with null Neumann boundary conditions,

$$\begin{cases} \Delta f = \mu - \lambda & \text{in } \Omega, \\ \nabla f \cdot \nu_\Omega = 0 & \text{on } \partial\Omega, \end{cases}$$

where ν_Ω denotes the normal to the boundary of Ω. A similar ansatz applied to Wb^2_Ω yields the approximation

$$\mathsf{W}^2_\Omega(\mu, \lambda) \approx \int_\Omega |\nabla u|^2,$$

where u solves the Poisson equation with same datum, but null Dirichlet boundary conditions:

$$\begin{cases} \Delta u = \mu - \lambda & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{1.4}$$

Despite being a rather drastic approximation, the main problem remains: how to compare the "Dirichlet" energy, $\int_\Omega |\nabla u|^2$, with the "Neumann" one, $\int_\Omega |\nabla f|^2$? To this aim, we first recall the following representation (akin to the Benamou-Brenier formula) for the Neumann energy:

$$\int_\Omega |\nabla f|^2 = \min \left\{ \int_\Omega |b|^2 \ : \ \operatorname{div} b = \mu - \lambda, \ b \cdot \nu_\Omega = 0 \right\}. \tag{1.5}$$

Similarly, for the Dirichlet energy we have:

$$\int_\Omega |\nabla u|^2 = \min \left\{ \int_\Omega |b|^2 \ : \ \operatorname{div} b = \mu - \lambda \right\}. \tag{1.6}$$

In particular, since $b = \nabla f$ is an admissible vector field for (1.6),

$$\int_\Omega |\nabla u|^2 \leq \int_\Omega |\nabla f|^2.$$

This reflects a similar property for the Wasserstein distance. Hence one only needs to provide a converse inequality, up to an error term that is negligible in the limit when the difference $\mu - \lambda$ is small, in a suitable weak sense.

We see immediately that if u solves (1.4), to obtain a competitor for (1.5), the gradient vector field $b_0 := \nabla u$, only needs to be corrected for its flux at the boundary. We then introduce a smooth cut-off function $\eta : \Omega \to [0, 1]$, that is identically 1 on a neighborhood of $\partial\Omega$ and such that $\eta(x) = 0$ for any $x \in \Omega$ whose distance from $\partial\Omega$ is larger than a parameter r (to be chosen suitably small). In particular, we may assume that $|\nabla\eta| \lesssim r^{-1}$ and $|\nabla^2\eta| \lesssim r^{-2}$. Setting $\xi := 1 - \eta$, the corrected vector field $b_1 := b_0 - \eta b_0 = \xi b_0$ satisfies the no-flux condition on $\partial\Omega$ – it is actually null on $\partial\Omega$ – but its divergence reads

$$\operatorname{div} b_1 - (\mu - \lambda) = -\eta \operatorname{div} b_0 - \nabla\eta \cdot b_0 = -\eta(\mu - \lambda) - \nabla\eta \cdot \nabla u,$$

hence one needs to introduce a further correction. At this stage, we observe the following elementary identity:

$$\nabla\eta \cdot \nabla u = \operatorname{div}(u\nabla\eta) - u\Delta\eta. \tag{1.7}$$

This is a key observation for our analysis as it allows, via integration by parts, to bound weak norms of our error terms by u rather than ∇u, see in particular (1.12) below. Moreover, this suggests as a further correction, to use $b_2 := b_1 + u\nabla\eta$, which still has no flux at the boundary and

$$\operatorname{div} b_2 - (\mu - \lambda) = -\eta(\mu - \lambda) - u\Delta\xi.$$

To complete our construction, we solve the Poisson equation

$$\begin{cases} \Delta g = \eta(\mu - \lambda) + u\Delta\xi & \text{in } \Omega \\ \nabla g \cdot \nu_\Omega = 0 & \text{on } \partial\Omega, \end{cases} \tag{1.8}$$

and set $b_3 := b_2 + \nabla g$. Notice that using integration by parts, (1.4) and the fact that ξ vanishes on a neighborhood of $\partial\Omega$, we find

$$\int_\Omega \eta(\mu - \lambda) + u\Delta\xi = \int_\Omega \eta(\mu - \lambda) + \xi(\mu - \lambda) = 0$$

so that (1.8) is indeed solvable. Since b_3 is admissible for (1.5), we have

$$\int_\Omega |\nabla f|^2 \le \int_\Omega |b_3|^2.$$

Using the triangle inequality and standard energy estimates for elliptic equations we eventually find that for any $\varepsilon \in (0,1)$,

$$\begin{aligned} \int_\Omega |\nabla f|^2 - \int_\Omega |\nabla u|^2 \lesssim \varepsilon \int_\Omega |\nabla u|^2 \\ + \frac{1}{\varepsilon}\left[\|\eta(\mu - \lambda)\|^2_{H^{-1,2}(\Omega)} + \left(r^{-2} + \operatorname{diam}(\Omega)^2 r^{-4}\right) \int_\Omega |u|^2 \right]. \end{aligned} \tag{1.9}$$

This is the PDE counterpart of our main deterministic bound for Wasserstein distances, (3.46) in Theorem 3.7. Indeed, the first and last terms in the square brackets have a formal correspondence, by replacing the negative Sobolev norm with the Wasserstein distance and the solution to the Poisson problem with the Kantorovich potential.

Application to the random setting

Inequality (1.9) however becomes useful only if one can show that the terms in square brackets become negligible, which is indeed the case in our random setting. Let us assume therefore that $\Omega = (0,L)^d$, μ is a Poisson point process and λ is the Lebesgue measure renormalized to have the same total mass of μ on $(0,L)^d$. Inserting the known upper (and lower) bounds for the matching problem in the PDE heuristics, we expect

$$\frac{1}{L^d}\mathbb{E}\left[\int_{(0,L)^d} |\nabla u|^2\right] \approx \frac{1}{L^d}\mathbb{E}\left[\mathsf{Wb}^2_{(0,L)^d}(\mu,\lambda)\right] \approx 1, \tag{1.10}$$

which can be roughly summarized as $|\nabla u| \approx 1$, corresponding to fact that the transport happens typically on a distance of order ≈ 1.

Regarding the cut-off function η, we choose $r = \delta L$ for some given $\delta \in (0,1)$. Then, we clearly see that a competition arises between the terms in the square brackets in (1.9). The first term should become negligible when $\delta \to 0$, as seen again by the heuristics,

$$\frac{1}{L^d}\|\eta(\mu - \lambda)\|^2_{H^{-1,2}((0,L)^d)} \approx \frac{1}{L^d}\mathsf{W}^2_{(0,L)^d}(\eta\mu, \eta\lambda) \approx \delta.$$

Indeed, we expect that also the transport between the weighted measures happens at a distance of order ≈ 1, but the volume of the support is now $\approx \delta L^d$. On the other side, the second term evidently grows as $\delta \to 0$:

$$\frac{1}{L^d}\left(r^{-2} + L^2 r^{-4}\right)\int_{(0,L)^d} |u|^2 \approx \left(\delta^{-2} + \delta^{-4}\right)\cdot \frac{1}{L^{d+2}}\int_{(0,L)^d} |u|^2.$$

We notice that Poincaré inequality on $(0,L)^d$ yields the upper bound

$$\frac{1}{L^{d+2}}\int_{(0,L)^d}|u|^2 \lesssim \frac{L^2}{L^{d+2}}\int_{(0,L)^d}|\nabla u|^2 \approx 1, \tag{1.11}$$

which is however not sufficient to conclude. Instead, we are able to argue that

$$\mathbb{E}\left[\int_\Omega |u|^2\right] \ll L^{d+2} \tag{1.12}$$

where $\ll$ means that the left-hand side is asymptotically of lower order. Thus, we may let first $L\to\infty$ and then $\delta\to 0$ to conclude. The intuition behind (1.12) is that ∇u is highly oscillating, hence Poincaré inequality fails to correctly capture cancellations and thus the order of u. A similar phenomenon happens in quantitative stochastic homogenization, see for instance [6, 7], and in the PDE heuristics one sees how to borrow ideas from that field. Namely, we introduce an intermediate scale $1 \ll L_0 \ll L$ and partition $(0,L)^d$ into $m \approx (L/L_0)^d$ cubes, each of side length L_0, and (roughly) define a competitor $\tilde{u}$ on $(0,L)^d$ for the Dirichlet energy by gluing the solutions to (1.4) on each sub-cube. Decomposing the average expected energy into the sum of the contributions on each sub-cube, by stationarity of the Poisson point process, we find that it is equal to the expected average energy of the solution u_0 to the Poisson problem (1.4) on a single cube $(0,L_0)^d$:

$$\frac{1}{L^d}\mathbb{E}\left[\int_{(0,L)^d}|\nabla\tilde{u}|^2\right] = \frac{1}{L_0^d}\mathbb{E}\left[\int_{(0,L_0)^d}|\nabla u_0|^2\right]. \tag{1.13}$$

Using super-additivity arguments, it can be shown that the right-hand side of (1.13) converges as $L_0\to\infty$. In particular for L_0 large (which also implies L large) it must be almost equal to the left-hand side of (1.10). This implies that the vector field $\nabla\tilde{u}$ is almost minimizing (1.6), i.e.,

$$\frac{1}{L^d}\mathbb{E}\left[\int_{(0,L)^d}|\nabla\tilde{u}|^2\right] - \frac{1}{L^d}\mathbb{E}\left[\int_{(0,L)^d}|\nabla u|^2\right] = \frac{1}{L_0^d}\mathbb{E}\left[\int_{(0,L_0)^d}|\nabla u_0|^2\right] - \frac{1}{L^d}\mathbb{E}\left[\int_{(0,L)^d}|\nabla u|^2\right] \ll 1.$$

By strong convexity of the Dirichlet energy, we have the stability (in)equality

$$\mathbb{E}\left[\int_{(0,L)^d}|\nabla u-\nabla\tilde{u}|^2\right] = \mathbb{E}\left[\int_{(0,L)^d}|\nabla\tilde{u}|^2\right] - \mathbb{E}\left[\int_{(0,L)^d}|\nabla u|^2\right] \ll L^d. \tag{1.14}$$

Coupling this with Poincaré inequality and $\mathbb{E}\left[\int_{(0,L)^d}|\tilde{u}|^2\right] \lesssim L_0^2L^d$, by (1.11) applied in each sub-cube of side length L_0, we obtain

$$\begin{aligned}\mathbb{E}\left[\int_{(0,L)^d}|u|^2\right] &\lesssim \mathbb{E}\left[\int_{(0,L)^d}|\tilde{u}|^2\right] + \mathbb{E}\left[\int_{(0,L)^d}|u-\tilde{u}|^2\right]\\ &\lesssim L_0^2L^d + L^2\mathbb{E}\left[\int_{(0,L)^d}|\nabla u-\nabla\tilde{u}|^2\right] \ll L^{d+2},\end{aligned}$$

as claimed in (1.12).

Further comments

The argument above fairly describes the main structure of the proofs of our results. However, we need to take into account that the Wasserstein distance is not actually approximated, at least up to length scales of order ≈ 1, by the negative Sobolev norm, hence much of technical effort is to fit these ideas, that may appear simple in the PDE heuristics, into the actual optimal transport setting. A crucial point we highlight here is for example the need for an analogue of the stability inequality

(1.14), which in the PDE setting uses strong convexity of the energy, while the optimal transport cost is usually only weakly convex – it is a linear programming problem. For this, we rely in Theorem 3.9, on some recent results in the literature of stability for optimal transport problem for the standard quadratic cost [22, 23] together with a localization argument. Let us point out that in our application, the uniform monotonicity of the gradient of the Kantorovich functional as stated in [23, 36, 41, 21] is not enough and we need the seemingly stronger uniform convexity of the Kantorovich functional itself as presented in [22] (following the proof from [23]).

Let us point out that contrary to (1.14) which gives directly control on the gradients, the stability inequality (3.55) in the case of the transport cost gives only control on the potentials. This forces us to rely on the div-curl type identity (1.7) which acts as a replacement for the Caccioppoli inequality. Indeed, in the PDE heuristics it would be actually possible to avoid (1.7) thanks to this stronger stability property. As will be apparent in the proofs, see in particular (3.46) in Theorem 3.7, it seems that the lack of (1.7) is the main obstacle to the extension of our results from $p = 2$ to $p > 1$ arbitrary.

Finally, we notice that our arguments apply to general stationary random measures that satisfy suitable concentration bounds. Precise assumptions are given in Assumption 4.1. We thus actually show Theorem 4.3, which contains Theorem 1.2 as a particular case. Another notable example covered by Theorem 4.3, is the Brownian interlacement occupation measure – see [38, 26, 49] in any dimension $d \geq 5$.

1.3 Further questions and conjectures

Our work makes substantial progress towards the full resolution of the conjectured validity of the limit in (1.2). We list here some questions and open problems that may be further addressed:

i) A result for the random Euclidean bipartite matching problem, thus settling the conjecture from [9, Theorem 2], seems also quite natural, and the only missing ingredient seems to be a stability inequality akin to Theorem 3.9, when both measures are singular – only the case when one of the two is sufficiently regular can be treated at the moment.

ii) In [24], the constant $\underline{\mathfrak{c}}(d, p)$ is defined in terms of an asymmetric boundary Wasserstein distance, which is smaller than the one we consider here. It would be interesting to prove that both constants coincide. In particular it would also imply that they coincide with the constant from [34]. In light of [34, Theorem 1.1] this would be a major step towards proving that all these problems generate the same coupling between the Lebesgue measure and the Poisson point process in the limit $L \to \infty$.

iii) While we drew inspiration from quantitative stochastic homogenization, we only proved here a qualitative convergence result. As its name suggests, the main and far more ambitious goal of quantitative stochastic homogenization is to obtain quantitative (and hopefully optimal) convergence rate. It would be interesting to investigate if such bounds could be obtained also for the transport problem. It is in fact conjectured in [19] that the rate of convergence should be of the order of L^{2-d}, see also [29] for the current rate of convergence known in the case $p = d = 2$.

iv) As already mentioned, our convergence result applies as well to the optimal transport of the Brownian interlacement occupation measure, which in [38] has been shown to provide an upper bound to the rate of convergence of the occupation measure of a single Brownian path $(B_t)_{t\geq 0}$ in the flat torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$. We conjecture that the boundary variant may provide also an asymptotic lower bound, hence settling a question left open in [38]. We remark that by using the decomposition arguments from [4, Section 6] our main result yields that similar limits exist in the case of transport of i.i.d. points on compact manifolds. Asymptotic upper

and lower bounds for the occupation measure of general diffusion processes have been studied recently, see e.g. [56, 54, 57, 58, 60, 55, 59], although to our knowledge limit results are available only in lower dimensions – corresponding to the cases where the scaling behaviors of the costs are somehow exceptional.

1.4 Structure of the paper

The exposition is structured as follows. Section 2 provides the necessary background on general notation, measure, metric and Sobolev spaces. In Section 3, after introducing the optimal transport theory for distance costs and its boundary variant, we establish the main deterministic results, in particular Theorem 3.7. Section 4 follows by providing Theorem 4.3, which is a generalized version of Theorem 1.2 – we also show that the Brownian interlacement occupation measure satisfies its assumptions.

2 Notation and basic facts

Given a set A, we write χ_A for its indicator function. We write $x \cdot y$ for the standard scalar product in $\mathbb{R}^d$ and $|x| = \sqrt{x \cdot x}$ for the Euclidean norm. We always endow $\mathbb{R}^d$ with the Euclidean distance. We introduce the following notation for the "power" $p > 0$ of a vector $x \in \mathbb{R}^d$, given by $x^{(p)} := |x|^{p-1}x$, with the convention that $0^{(p)} = 0$. The notation $A \lesssim B$ means that there exists a constant $C > 0$, such that $A \leq CB$, where C depends on the dimension d and p. 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 $B \lesssim A$.

2.1 Measure spaces

Given a finite measure space $(E, \mathcal{E}, \lambda)$ and a function f on E, taking values in $\mathbb{R}^d$, we write for $p \in [1, \infty)$, $\|f\|_{L^p(\lambda)} := \left(\int_E |f|^p d\lambda\right)^{1/p}$ for the Lebesgue norm, and $\|f\|_\infty := \sup_{x \in E} |f(x)|$ for the uniform norm. For a real valued $f \in L^2(\lambda)$, we write

$$\mathrm{Var}_\lambda(f) := \inf_{a \in \mathbb{R}} \int_E |f - a|^2 \, d\lambda,$$

which is attained at $a = \fint_E f d\lambda := \lambda(E)^{-1} \int_E f d\lambda$. Notice that we do not require λ to be a probability: however, it holds

$$\mathrm{Var}_{c\lambda}(f) = c \, \mathrm{Var}_\lambda(f) \quad \text{for every } f \in L^2(\lambda), \, c > 0. \tag{2.1}$$

The function $f \mapsto \sqrt{\mathrm{Var}_\lambda(f)}$ defines a semi-norm on $L^2(\lambda)$ that is null on constant functions. If $\tilde{\lambda}$ is a measure on E such that $\tilde{\lambda} \leq \lambda$, then it holds

$$\mathrm{Var}_{\tilde{\lambda}}(f) \leq \mathrm{Var}_\lambda(f) \quad \text{for every } f \in L^2(\lambda). \tag{2.2}$$

When $\lambda = P$ is a probability measure, we employ the standard probabilistic notation, e.g. $\mathbb{E}[\cdot] = \int_E \cdot dP$ for the expectation value and $\mathrm{Var}(\cdot) = \mathrm{Var}_P(\cdot)$ for the variance.

When $E \subseteq \mathbb{R}^d$ is measurable and bounded, and λ is the Lebesgue measure on E, we simply write $\|f\|_{L^p(E)} := \|f\|_{L^p(\lambda)}$, $|E| := \lambda(E)$, $\int_E f := \int_E f(x)dx$ for $f \in L^1(E)$ and $\fint_E f := |E|^{-1} \int_E f$. We also write

$$\mathrm{Var}_E(f) := \mathrm{Var}_\lambda(f) \quad \text{for every } f \in L^2(E),$$

and notice that, for a measurable $\tilde{E} \subseteq E$ it holds

$$\mathrm{Var}_{\tilde{E}}(f) \leq \mathrm{Var}_E(f) \quad \text{for every } f \in L^2(E).$$

For simplicity, we always identify any finite measure $\tilde{\lambda}$ on E that is absolutely continuous with respect to the Lebesgue measure with its density that we keep denoting with the same letter. For example, we write $\int_E f\lambda := \int_E f d\lambda$.

2.2 Metric spaces

Given a metric space $(\mathcal{X}, \mathsf{d})$ and $\Omega \subseteq \mathcal{X}$, we write $\overline{\Omega}$ for its closure,

$$\mathsf{d}(\Omega) = \sup_{x,x' \in \Omega} \mathsf{d}(x, x')$$

for its diameter (also written $\mathrm{diam}(\Omega)$ if d is understood). We write

$$\mathsf{d}(x, \Omega^c) := \inf_{z \in \Omega^c} \mathsf{d}(x, z),$$

for the distance function from Ω^c and for $\mathcal{Y} \subseteq \mathcal{X}$,

$$\mathsf{d}(\Omega, \mathcal{Y}) = \sup_{x \in \Omega} \mathsf{d}(x, \mathcal{Y}).$$

Notice that this quantity is not symmetric. It will be convenient to introduce the notation

$$\Omega^{(r)} := \{x \in \Omega \,:\, \mathsf{d}(x, \Omega^c) > r\}, \quad \text{for } r \geq 0. \tag{2.3}$$

For a function $f : \mathcal{X} \to \mathbb{R}$, we write $\sup_\Omega f := \sup_{x \in \Omega} f(x)$, $\inf_\Omega f := \inf_{x \in \Omega} f(x)$, $\mathrm{osc}_\Omega(f) := \sup_\Omega f - \inf_\Omega f$ for its oscillation and

$$\mathrm{Lip}_\Omega(f) := \sup_{x \neq x' \in \Omega} \frac{|f(x) - f(x')|}{|x - x'|}$$

for its Lipschitz constant. We have

$$\mathrm{osc}_\Omega(f) \leq \mathrm{diam}(\Omega)\, \mathrm{Lip}_\Omega(f),$$

and, for every finite measure λ supported on Ω,

$$\mathrm{Var}_\lambda(f) \leq \frac{\lambda(\Omega)}{4} \mathrm{osc}_\Omega(f)^2. \tag{2.4}$$

When $\mathcal{X} \subseteq \mathbb{R}^d$ (endowed with the Euclidean distance) we often consider $\Omega \subseteq \mathcal{X}$ to be a bounded domain with Lipschitz boundary. In our main results we require – but always specify – that Ω is convex. Examples include the cases of a rectangle $R = \prod_{i=1}^d (a_i, b_i)$ with $a_i < b_i$ for $i = 1, \ldots, d$ or a cube Q, i.e. the side lengths $b_i - a_i$ for $i = 1, \ldots, d$ all coincide. In particular, for $\Omega \subseteq \mathbb{R}^d$, (2.4) reads

$$\mathrm{Var}_\Omega(f) \leq \frac{|\Omega|}{4} \mathrm{osc}_\Omega(f)^2. \tag{2.5}$$

2.3 Sobolev spaces

Given a bounded connected domain $\Omega \subseteq \mathbb{R}^d$ with Lipschitz boundary, the gradient of a function $f : \Omega \to \mathbb{R}$ is denoted by ∇f, the divergence of a vector field $b : \Omega \to \mathbb{R}^d$ is denoted by $\mathrm{div}\, b$, and the outward unit normal to the boundary of $\Omega \subseteq \mathbb{R}^d$ is denoted by ν_Ω. We recall that a Lipschitz function is differentiable at Lebesgue a.e. $x \in \Omega$ and that

$$\|\nabla f\|_{L^\infty(\Omega)} \leq \mathrm{Lip}_\Omega(f),$$

with equality if Ω is convex.

Given a function $\eta : \Omega \to [0, \infty)$, which we assume for simplicity bounded, weighted Sobolev spaces $H^{1,p}(\eta)$ are defined as the completion of smooth (or Lipschitz) functions with respect to the Sobolev norm

$$\|f\|_{H^{1,p}(\eta)} := \left(\|f\|^p_{L^p(\eta)} + \|\nabla f\|^p_{L^p(\eta)} \right)^{1/p}.$$

Since convergence in $H^{1,p}(\eta)$ entails convergence in $L^p(\eta)$ it follows that $H^{1,p}(\eta) \subseteq L^p(\eta)$. The subspace $H_0^{1,p}(\eta)$ of functions that have zero trace at the boundary is defined as the closure of smooth functions compactly supported in Ω. When, $\eta = \chi_\Omega$, we write $H^{1,p}(\Omega) := H^{1,p}(\eta)$ and similarly $H_0^{1,p}(\Omega) := H_0^{1,p}(\eta)$.

A key tool in Sobolev spaces analysis is the validity of a Poincaré-Wirtinger inequality, which in its weighted form reads as follows: there exists a constant $c < \infty$ such that for every Lipschitz function f and null weighted mean $\int_\Omega f\eta = 0$,

$$\|f\|_{L^p(\eta)} \leq c \, \|\nabla f\|_{L^p(\eta)} \, . \tag{2.6}$$

We write throughout $c_P(\eta, p)$ for the smallest constant such that (2.6) holds. Let us notice that, if η' is a weight equivalent to η in the sense that, for some constant $c_0 \in (0, \infty)$, $c_0^{-1}\eta(x) \leq \eta'(x) \leq c_0\eta(x)$ for every $x \in \Omega$, then it holds

$$\frac{1}{2c_0^{1/p}} c_P(\eta, p) \leq c_P(\eta', p) \leq 2c_0^{1/p} c_P(\eta). \tag{2.7}$$

Indeed, for $p = 2$ this follows (even without the factors $1/2$ and 2) from the fact that the left-hand side in (2.6) can be equivalently rewritten as $\mathrm{Var}_\eta(f)$, dropping the null mean requirement, and the homogeneity properties (2.1) (2.2) of the variance easily allow for a change of measure. For general p, one first notices that $c_P(\eta, p) < \infty$ is actually equivalent to the validity of an inequality of the type

$$\inf_{a \in \mathbb{R}} \|f - a\|_{L^p(\eta)} \leq \tilde{c} \, \|\nabla f\|_{L^p(\eta)} \quad \text{for every } f \in H^{1,p}(\eta), \tag{2.8}$$

for some constant $\tilde{c} \in (c_P(\eta, p)/2, c_P(\eta, p))$, and then argues as in the case of $p = 2$ using similar properties for the "p-variance" appearing in the left-hand side. Indeed, (2.6) implies (2.8) with $\tilde{c} = c_P(\eta, p)$, by taking $a = \fint_\Omega f\eta$. Conversely, assuming for simplicity that $\int_\Omega \eta = 1$, we notice that, for every $a \in \mathbb{R}$,

$$\left| \int_\Omega f\eta - a \right| \leq \|f - a\|_{L^1(\eta)} \leq \|f - a\|_{L^p(\eta)} \, ,$$

so that if (2.8) holds we obtain, by the triangle inequality,

$$\begin{aligned} \left\| f - \int_\Omega f\eta \right\|_{L^p(\eta)} &\leq \inf_{a \in \mathbb{R}} \left\{ \|f - a\|_{L^p(\eta)} + \left| \int_\Omega f\eta - a \right| \right\} \leq 2 \inf_{a \in \mathbb{R}} \|f - a\|_{L^p(\eta)} \\ &\leq 2c_P(\eta, p) \, \|\nabla f\|_{L^p(\eta)} \, . \end{aligned}$$

In case of a constant weight $\eta = \chi_\Omega$, we write $c_P(\Omega, p) := c_P(\chi_\Omega, p)$ and we recall that since we assume that $\Omega \subseteq \mathbb{R}^d$ is a bounded connected domain with Lipschitz boundary, it always holds $c_P(\Omega, p) < \infty$. In particular, in the case of $\Omega = Q_L$ a cube of side length L, we have $c_P(Q_L, p) \lesssim L$ and more generally c_P scales linearly with respect to dilations. In [46, Theorem 1.2] it is also proved that, for any family of domains $\{\Omega_i\}_{i \in I}$, the constants $c_P(\Omega_i, p)$ are uniformly bounded, provided that the following three conditions hold:

a) the diameters $\{\mathrm{diam}(\Omega_i)\}_{i \in I}$ are uniformly bounded,

b) the sets $\{\Omega_i\}_{i \in I}$ are uniformly connected, i.e., there exists $r > 0$ such that all the sets $\Omega_i^{(r)}$ are connected,

c) a uniform interior cone condition holds, i.e., there exists an open cone C (with finite height) such that, for every $i \in I$, $x \in \partial\Omega_i$, one can find a cone C_x with vertex at x and congruent to C with $C_x \subset \Omega_i$.

The uniform cone condition holds in particular if the boundaries $\{\partial\Omega_i\}_{i\in I}$ are uniformly Lipschitz.

Back to the weighted case, in [45, Theorem 1], it is proved that

$$c_P(\eta, p) \lesssim \inf_{0\le\tau\le\|\eta\|_\infty} \left(\frac{\|\eta\|_\infty |\Omega|}{\tau |\{\eta > \tau\}|}\right)^{1/p} \|t \mapsto c_P(\{\eta > t\}, p)\|_{L^\infty([0,\tau])}, \tag{2.9}$$

where the sets $\{\eta > t\} = \{x \in \Omega : \eta(x) > t\}$ are assumed to be connected domains with Lipschitz boundary.

Remark 2.1. As a simple application of this inequality and the uniformity discussed above, let us consider a weight $\eta : \Omega \to [0, 1]$ such that $|\{\eta > 1/2\}| > 0$ and the sets $(\{\eta > t\})_{t\in[0,1/2]}$ have uniformly Lipschitz boundaries, hence they satisfy a uniform cone property, and they are uniformly connected. Then, $\|t \mapsto c_P(\{\eta > t\}, p)\|_{L^\infty([0,1/2])}$ is finite and we obtain by (2.9) with $\tau = 1/2$ that $c_P(\eta, p) < \infty$. The same argument extend to any family of weights $(\eta_i)_{i\in I}$ such that

$$\sup_{i\in I} \|\eta_i\|_\infty < \infty, \quad \inf_{i\in I} |\{\eta_i > 1/2\}| > 0,$$

and the sets $(\{\eta_i > t\})_{t\in[0,1/2],i\in I}$ satisfy a uniform cone property and are uniformly connected. In such a case, we deduce

$$\sup_{i\in I} c_P(\eta_i, p) < \infty.$$

This may also follow by the continuity of the Poincaré constant with respect to Hausdorff convergence for sets satisfying a uniform cone condition, see [17, Corollary 7.4.2] for the case $p = 2$.

Sobolev spaces of negative order are defined in terms of the following dual "norm":

$$\|f\|_{H^{-1,p}(\eta)} = \sup_{\|\nabla\phi\|_{L^{p'}(\eta)}\le 1} \int_\Omega f\phi\eta \in [0, \infty] \tag{2.10}$$

for $p \in (1, \infty)$ where $p' = p/(p-1)$. By definition, the supremum is upon $\phi \in H^{1,p'}(\eta)$ or equivalently upon Lipschitz functions. We notice in particular that if $\|\cdot\|_{H^{-1,p}(\eta)} < \infty$ then $\int_\Omega f\eta = 0$. In this case we may restrict the supremum to functions ϕ having also average zero.

Using Hölder's inequality and (2.6), we obtain immediately that

$$\|f\|_{H^{-1,p}(\eta)} \lesssim c_P(\eta, p')\|f\|_{L^p(\eta)}, \tag{2.11}$$

provided that $\int_\Omega f\eta = 0$. We define the operator

$$\mathrm{div}_\eta(b) = \mathrm{div}(b) - \eta^{-1}\nabla\eta \cdot b \tag{2.12}$$

which is understood in the weak sense, i.e.,

$$\int_\Omega \nabla\phi \cdot b\eta = -\int_\Omega \mathrm{div}_\eta(b)\phi\eta \quad \text{for every Lipschitz function } \phi. \tag{2.13}$$

It is then classical that provided the Poincaré inequality (2.6) holds for p', then we have the dual representation

$$\|f\|_{H^{-1,p}(\eta)} = \inf\left\{\|b\|_{L^p(\eta)} : -\mathrm{div}_\eta(b) = f\right\}. \tag{2.14}$$

Indeed, inequality $\le$ is true as a simple consequence of Hölder inequality. For the reverse inequality, thanks to (2.6) we see that the supremum in the definition (2.10) is attained for some function ϕ. Using a first variation argument we have $\phi = \lambda h$ where $\lambda = \|\nabla h\|^{-1}_{L^{p'}(\eta)}$ and h solves (in the weak sense)

$$-\mathrm{div}_\eta(|\nabla h|^{p'-2}\nabla h) = f.$$

Using (2.13) we then see that setting $b = |\nabla h|^{p'-2}\nabla h$, we have $-\operatorname{div}_\eta b = f$ and

$$\|f\|_{H^{-1,p}(\eta)} = \|\nabla h\|_{L^{p'}(\eta)}^{p'-1} = \|b\|_{L^p(\eta)}$$

proving the other inequality.

As a consequence of (2.14), we have the following sub-additivity property: given a (finite) partition $\Omega = \bigcup_k \Omega_k$ into domains such that $c_P(\chi_{\Omega_k}\eta, p') < \infty$ for every k, then it holds

$$\|f\|_{H^{-1,p}(\eta)}^p \le (1+\varepsilon)\sum_k \left\| f - \int_{\Omega_k} f\eta_k \right\|_{H^{-1,p}(\chi_{\Omega_k}\eta)}^p + \frac{C}{\varepsilon^{p-1}} \left\| \sum_k \chi_{\Omega_k} \int_{\Omega_k} f\eta_k \right\|_{H^{-1,p}(\eta)}^p, \tag{2.15}$$

where $\varepsilon \in (0,1)$ is arbitrary and $\eta_k := \eta / \int_{\Omega_k} \eta$ (set equal to 0 if η is null on Ω_k). To see this, we first use the triangle inequality for $\|\cdot\|_{H^{-1,p}(\eta)}$,

$$\|f\|_{H^{-1,p}(\eta)}^p \le (1+\varepsilon) \left\| f - \sum_k \chi_{\Omega_k} \int_{\Omega_k} f\eta_k \right\|_{H^{-1,p}(\eta)}^p + \frac{C}{\varepsilon^{p-1}} \left\| \sum_k \chi_{\Omega_k} \int_{\Omega_k} f\eta_k \right\|_{H^{-1,p}(\eta)}^p.$$

For the first term we then use the formulation (2.14) with $b = \sum_k b_k \chi_{\Omega_k}$, where b_k is admissible for (2.14) with $f_k := \left(f - \fint_{\Omega_k} f\eta_k\right)\chi_{\Omega_k} \in L^p(\eta_k)$ and then minimize over b_k.

We end this section with two results that hold in the case of uniform weight $\eta = \chi_\Omega$. The first one provides a bound for the $H^{-1,2}$ norm of a function of the form $f = \nabla\eta\cdot\nabla u$ and is based on the identity (1.7) discussed in Section 1.2. Unfortunately this result seems to be very specific to the case $p = 2$.

Lemma 2.2. *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary, let $u \in H^{1,2}(\Omega)$ and $\eta : \Omega \to \mathbb{R}$ be C^2 smooth and such that $\nabla\eta\cdot\nu_\Omega = 0$ on $\partial\Omega$. Then, it holds*

$$\left\| \nabla\eta\cdot\nabla u - \fint_\Omega \nabla\eta\cdot\nabla u \right\|_{H^{-1,2}(\Omega)} \lesssim (\|\nabla\eta\|_\infty + \operatorname{diam}(\Omega)\,\|\Delta\eta\|_\infty)\,\operatorname{Var}_{\Omega_\eta}(u)^{1/2}. \tag{2.16}$$

where $\Omega_\eta := \Omega \cap \operatorname{supp}|\nabla\eta|$. Here, the implicit constant depends on Ω only and is invariant under dilations.

Proof. We use (1.7) and rewrite, for any smooth function ϕ with zero average on Ω,

$$\begin{aligned}\int_\Omega \phi\nabla\eta\cdot\nabla u &= \int_\Omega \phi\left[\operatorname{div}(u\nabla\eta) - u\Delta\eta\right] \\ &= -\int_\Omega u\nabla\phi\cdot\nabla\eta - \int_\Omega \phi u\Delta\eta\end{aligned}$$

where in the second identity there is no boundary term because $\nabla\eta\cdot\nu_\Omega = 0$. We bound from above

$$\int_\Omega u\nabla\phi\cdot\nabla\eta \le \|\nabla\eta\|_\infty\,\|u\|_{L^2(\Omega_\eta)}\,\|\nabla\phi\|_{L^2(\Omega)}$$

and, using Poincaré inequality (2.6), we find

$$\begin{aligned}\int_\Omega \phi u\Delta\eta &\le \|\Delta\eta\|_\infty\,\|u\|_{L^2(\Omega_\eta)}\,\|\phi\|_{L^2(\Omega)} \\ &\le \operatorname{diam}(\Omega)\,\|\Delta\eta\|_\infty\,\|u\|_{L^2(\Omega_\eta)}\,\|\nabla\phi\|_{L^2(\Omega)}.\end{aligned}$$

Summing these two bounds and taking the supremum over ϕ with $\|\nabla\phi\|_{L^2(\Omega)} \le 1$ yields (2.16) with $\|u\|_{L^2(\Omega_\eta)}^2$ instead of $\operatorname{Var}_{\Omega_\eta}(u)$. To conclude it is sufficient to notice that the left-hand side remains unchanged if one replaces u with $u - a$ for any constant $a \in \mathbb{R}$, and minimize upon a. □

For the second result, which is a particular (but not totally standard) case of Gagliardo-Nirenberg interpolation inequality, we recall first the case $p > d$ of Sobolev inequality, which reads

$$\sup_{x\neq x'\in\Omega} \frac{|f(x)-f(x')|}{|x-x'|^{1-d/p}} \lesssim \|\nabla f\|_{L^p(\Omega)} \quad \text{for every } f \in H^{1,p}_0(\Omega), \tag{2.17}$$

where the implicit constant depends on p and Ω, but is invariant with respect to dilations of Ω.

Lemma 2.3. *For every $p \geq 1$ and every Lipschitz function $f : \mathbb{R}^d \to \mathbb{R}$ that is constant on Ω^c, it holds*

$$\mathrm{osc}_\Omega(f) \lesssim \mathrm{diam}(\Omega)^{\frac{p}{p+d}} \,\mathrm{Lip}_\Omega(f)^{\frac{d}{d+p}} \,\|\nabla f\|^{\frac{p}{p+d}}_{L^p(\Omega)},$$

where the implicit constant depends on p and Ω but is invariant with respect to dilations of Ω.

Proof. Without loss of generality, we can assume that $f(x) = 0$ if $x \in \Omega^c$. Moreover, by scaling we may also assume that $\mathrm{diam}(\Omega) = 1$. Write f_ε for the convolution of f with a standard convolution kernel and split

$$\|f\|_\infty \lesssim \|f - f_\varepsilon\|_\infty + \|f_\varepsilon\|_\infty.$$

We have

$$\|f - f_\varepsilon\|_\infty \lesssim \varepsilon \|\nabla f\|_\infty \leq \varepsilon \,\mathrm{Lip}_\Omega(f)$$

and

$$\|f_\varepsilon\|_\infty \lesssim \varepsilon^{-d/p} \,\|f\|_{L^p(\mathbb{R}^d)} \leq c_P(\Omega, p)\varepsilon^{-d/p} \,\|\nabla f\|_{L^p(\Omega)},$$

where the second inequality follows from Poincaré inequality in $H^{1,p}_0(\Omega)$ (one does not need to subtract the mean). An optimization over ε yields the claimed interpolation inequality. □

3 Deterministic results

In this section, we collect all the deterministic tools related to optimal transport theory that we need in the proof of our main result. After recalling several basic facts that can be found in any of the standard references on the subject [2, 53, 43], we proceed towards the proof of our main result for this section, Theorem 3.7. It provides a non-trivial upper bound for W^2 in terms of Wb^2 and suitable error terms. A key tool that we use is a stability result for the quadratic optimal transport problem, see (1.14). Although we are only ultimately able to deal with the quadratic case, whenever the proofs allow, we provide more general results that readers may find useful.

3.1 Generalities

Let $\mathcal{X}$ denote a compact metric space and $\mathsf{d} : \mathcal{X} \times \mathcal{X} \to [0, \infty)$ be a pseudo-distance, i.e. like a standard distance with only the exception that $\mathsf{d}(x, y) = 0$ does not necessarily entail $x = y$. We also assume that d is (lower-)semicontinuous with respect to the topology of $\mathcal{X}$. Given two finite positive Borel measures μ, λ on $\mathcal{X}$, with $\mu(\mathcal{X}) = \lambda(\mathcal{X})$, the optimal transport cost of order $p \geq 1$ between μ and λ is defined as the quantity

$$\mathsf{Wd}^p(\mu, \lambda) = \inf_{\pi\in C(\mu,\lambda)} \int_{\mathcal{X}\times\mathcal{X}} \mathsf{d}(x, y)^p d\pi(x, y)$$

where $C(\mu, \lambda)$ denotes the set of couplings between μ and λ i.e., positive Borel measures π on $\mathcal{X} \times \mathcal{X}$ such that the marginal measure of x is λ and that of y is μ. In order to highlight the space $\mathcal{X}$ we may write $\mathsf{Wd}^p_{\mathcal{X}}$ instead of Wd^p. Moreover, if $\mathcal{X} \subseteq \mathbb{R}^d$ and $\mathsf{d}(x, y) = |y - x|$ is the Euclidean distance, we simply write $\mathsf{W}^p_{\mathcal{X}}$ for $\mathsf{Wd}^p_{\mathcal{X}}$.

The measures μ, λ are not necessarily probabilities, however for every positive constant $a \geq 0$, it holds

$$\mathsf{Wd}^p(a\mu, a\lambda) = a\mathsf{Wd}^p(\mu, \lambda),$$

which allows in principle to often reduce to the case $\mu(\mathcal{X}) = \lambda(\mathcal{X}) = 1$, whenever $\mu(\mathcal{X}) \neq 0$. For $\alpha \geq 1$, by Hölder inequality,

$$(\mathsf{Wd}^p(\mu, \lambda))^\alpha \leq \mu(\mathcal{X})^{\alpha-1}\mathsf{Wd}^{\alpha p}(\mu, \lambda). \tag{3.1}$$

Given two finite or countable collections of measures $(\mu_k)_k$, $(\lambda_k)_k$ such that $\mu_k(\mathcal{X}) = \lambda_k(\mathcal{X})$ for every k, the following subadditivity inequality holds:

$$\mathsf{Wd}^p\left(\sum_k \mu_k, \sum_k \lambda_k\right) \leq \sum_k \mathsf{Wd}^p\left(\mu_k, \lambda_k\right). \tag{3.2}$$

To prove it, given any collection of couplings $\pi_k \in C_\Omega(\mu_k, \lambda_k)$, simply notice that the measure $\sum_k \pi_k$ provides a coupling between $\sum_k \mu_k$ and $\sum_k \lambda_k$.

Clearly, if $\mathsf{d}_1(x, y) \leq \mathsf{d}_2(x, y)$ for every $x, y \in \mathcal{X}$, where d_1, d_2 denote lower-semicontinuous pseudo-distances, it holds

$$\mathsf{Wd}_1^p(\mu, \lambda) \leq \mathsf{Wd}_2^p(\mu, \lambda).$$

The trivial distance $\mathsf{d}(x, y) = 1_{\{x\neq y\}}$ yields (for any $p \geq 1$) that $\mathsf{Wd}^p(\mu, \lambda)$ equals the total variation distance, $\|\mu - \lambda\|_{\mathrm{TV}}$. Writing

$$\mathsf{d}(x, y)^p \leq 1_{\{x\neq y\}}\mathsf{d}(\mathcal{X})^p,$$

where we recall that $\mathsf{d}(\mathcal{X})$ is the diameter of $\mathcal{X}$ with respect to d, we obtain the bound

$$\mathsf{Wd}^p(\mu, \lambda) \leq \|\mu - \lambda\|_{\mathrm{TV}}\, \mathsf{d}(\mathcal{X})^p. \tag{3.3}$$

The general theory of optimal transportation yields the following dual representation:

$$\mathsf{Wd}^p(\mu, \lambda) = \sup\left\{\int_{\mathcal{X}} u d\lambda - \int_{\mathcal{X}} v d\mu \;:\; u(x) - v(y) \leq \mathsf{d}(x, y)^p \quad \forall x, y \in \mathcal{X}\right\}, \tag{3.4}$$

where the sup runs among continuous and bounded functions u, v on $\mathcal{X}$. It is slightly convenient to restrict v on a closed set $\mathcal{Y} \subseteq \mathcal{X}$, where μ is supported. Of course, one can always let $\mathcal{Y} = \mathcal{X}$, but a smaller $\mathcal{Y}$ (possibly finite) may lead to more elementary considerations. Duality can be then formulated in a less symmetric way, in terms of v (from now on defined on $\mathcal{Y}$) and its d^p-transform, given by

$$\mathsf{Q}_{\mathsf{d}^p}(v)(x) := \inf_{y\in\mathcal{Y}}\left\{v(y) + \mathsf{d}(x, y)^p\right\}.$$

When both d and p are clearly understood, we simply write $\mathsf{Q}(v) = \mathsf{Q}_{\mathsf{d}^p}(v)$ for such transform. It holds therefore

$$\mathsf{Wd}^p(\mu, \lambda) = \sup_v\left\{\int_{\mathcal{X}} \mathsf{Q}(v) d\lambda - \int_{\mathcal{Y}} v d\mu\right\}. \tag{3.5}$$

Indeed, if u is such that $u(x) - v(y) \leq \mathsf{d}(x, y)^p$, then $u(x) \leq \mathsf{Q}(v)(x)$, hence by replacing u in the right-hand side of (3.4) with $\mathsf{Q}(v)$ one increases it. Let us notice that if $\mathsf{d}(x, x') = 0$ then by the triangle inequality $\mathsf{d}(x, y) = \mathsf{d}(x', y)$ for every y, hence

$$\mathsf{d}(x, x') = 0 \quad \Rightarrow \quad \mathsf{Q}(v)(x) = \mathsf{Q}(v)(x'). \tag{3.6}$$

The transform is also monotone increasing:

$$v \leq \tilde{v} \quad \Rightarrow \quad \mathsf{Q}(v) \leq \mathsf{Q}(\tilde{v}).$$

If $a \in \mathbb{R}$, then

$$\mathsf{Q}(a + v)(x) = a + \mathsf{Q}(v)(x), \tag{3.7}$$

and setting $v(y) = 0$ for every $y \in \mathcal{Y}$, we find

$$\mathsf{Q}(0)(x) = \inf_{y \in \mathcal{Y}} \mathsf{d}(x, y)^p := \mathsf{d}(x, \mathcal{Y})^p.$$

By monotonicity, we then obtain

$$\sup_{x \in \mathcal{X}} \mathsf{Q}(v)(x) \leq \sup_{y \in \mathcal{Y}} v(y) + \mathsf{d}(\mathcal{X}, \mathcal{Y})^p \tag{3.8}$$

where we recall the notation

$$\mathsf{d}(\mathcal{X}, \mathcal{Y}) := \sup_{x \in \mathcal{X}} \mathsf{d}(x, \mathcal{Y}).$$

This upper bound, coupled with the lower bound

$$\inf_{x \in \mathcal{X}} \mathsf{Q}(v)(x) = \inf_{y \in \mathcal{Y}} \left\{ v(y) + \inf_{x \in \mathcal{X}} \mathsf{d}(x, y)^p \right\} = \inf_{y \in \mathcal{Y}} v(y),$$

yields the bound on the oscillation

$$\operatorname{osc}_{\mathcal{X}} (\mathsf{Q}(v)) \leq \operatorname{osc}_{\mathcal{Y}} (v) + \mathsf{d}(\mathcal{X}, \mathcal{Y})^p. \tag{3.9}$$

Finally, if the family $(\mathsf{d}(\cdot, y)^p)_{y \in \mathcal{Y}}$ is uniformly continuous (with respect to the topology on $\mathcal{X}$), then $\mathsf{Q}(v)$ is also continuous. In particular,

$$\operatorname{Lip}_{\mathcal{X}} (\mathsf{Q}(v)) \leq \sup_{y \in \mathcal{Y}} \operatorname{Lip}_{\mathcal{X}} (\mathsf{d}(\cdot, y)^p). \tag{3.10}$$

The roles played by u and v in (3.4) can be reversed by defining a dual d^p-transform, for $u : \mathcal{X} \to \mathbb{R}$, given by

$$\hat{\mathsf{Q}}_{\mathsf{d}^p}(u)(y) := \sup_{x \in \mathcal{X}} \{u(x) - \mathsf{d}(x, y)^p\}.$$

Again, if d and p are understood, we simply write $\hat{\mathsf{Q}} = \hat{\mathsf{Q}}_{\mathsf{d}^p}$. The dual transform is also monotone increasing and the oscillation bound holds:

$$\operatorname{osc}_{\mathcal{Y}} \left(\hat{\mathsf{Q}}_{\mathsf{d}^p}(u) \right) \leq \operatorname{osc}_{\mathcal{X}} (u) + \mathsf{d}(\mathcal{X}, \mathcal{Y})^p. \tag{3.11}$$

Moreover, it always holds

$$\mathsf{Q}\left(\hat{\mathsf{Q}}(u)\right) \geq u \quad \text{and} \quad \hat{\mathsf{Q}}\left(\mathsf{Q}(v)\right) \leq v$$

hence $\mathsf{Q} \circ \hat{\mathsf{Q}} \circ \mathsf{Q} = \mathsf{Q}$. Combining these two transforms, we see in particular that in (3.5) we can always assume that $v = \hat{\mathsf{Q}}(u)$ for some u, and denote in what follows $\mathcal{S}(\mathsf{d}^p, \mathcal{Y})$ the set of such functions. For $v \in \mathcal{S}(\mathsf{d}, \mathcal{Y})$, its d^p-sub-differential is defined as the set

$$\partial v(y) := \{x \in \mathcal{X} \,:\, \mathsf{Q}(v)(x) - v(y) = \mathsf{d}(x, y)^p\}.$$

Similarly, we define

$$\partial \mathsf{Q}(v)(x) := \{y \in \mathcal{Y} \,:\, \mathsf{Q}(v)(x) - v(y) = \mathsf{d}(x, y)^p\},$$

so that $x \in \partial v(y)$ if and only if $y \in \partial \mathsf{Q}(v)(x)$.

Most of the above facts actually hold for a general lower semicontinuous cost function $\mathsf{c}(x,y)$ instead of d^p. However, we additionally have that the p-th root of Wd^p enjoys the triangle inequality, as a consequence of the triangle inequality of d. For our purposes, it is more useful to state and prove the following inequality for Wd^p: for every $\bar{\lambda}$ with $\bar{\lambda}(\mathcal{X}) = \mu(\mathcal{X})$ and $\varepsilon \in (0,1)$, it holds

$$\mathsf{Wd}^p(\mu,\lambda) \le (1+\varepsilon)\mathsf{Wd}^p(\mu,\bar{\lambda}) + \frac{c}{\varepsilon^{p-1}}\mathsf{Wd}^p(\bar{\lambda},\lambda), \tag{3.12}$$

where $c = c(p) < \infty$. Although this follows at once from the triangle inequality for Wd, we give a proof that uses the dual formulation, as it entails an intermediate inequality (3.14), also useful for our purposes. Keeping d and p fixed, for any $t > 0$ we define $\mathsf{d}_t(x,y) := t^{\frac{1}{p}-1}\mathsf{d}(x,y)$. This scaling simply reflects on the transport cost as follows:

$$\mathsf{Wd}_t^p(\mu,\lambda) = t^{1-p}\mathsf{Wd}^p(\mu,\lambda),$$

but yields a modified transform

$$\mathsf{Q}_t(v)(x) := \mathsf{Q}_{\mathsf{d}_t^p}(v)(x) := \inf_{y\in\mathcal{Y}}\left\{v(y) + \frac{\mathsf{d}(x,y)^p}{t^{p-1}}\right\}, \tag{3.13}$$

for which the following semigroup inequality holds:

$$\mathsf{Q}_t\left(\mathsf{Q}_s(v)\right) \ge \mathsf{Q}_{s+t}(v),$$

where $\mathsf{Q}_s(v)$ is restricted to $\mathcal{Y}$. Indeed, it holds

$$\begin{aligned}
\mathsf{Q}_t\left(\mathsf{Q}_s(v)\right)(x) &= \inf_{y\in\mathcal{Y}}\left\{\inf_{z\in\mathcal{Y}}\left\{v(z) + \mathsf{d}_s(y,z)^p\right\} + \mathsf{d}_t(x,y)^p\right\} \\
&= \inf_{z\in\mathcal{Y}}\left\{v(z) + \inf_{y\in\mathcal{Y}}\left\{\mathsf{d}_t(x,y)^p + \mathsf{d}_s(y,z)^p\right\}\right\} \\
&\ge \inf_{z\in\mathcal{Y}}\left\{v(z) + \mathsf{d}_{s+t}(x,z)^p\right\},
\end{aligned}$$

where the second inequality follows from the triangle inequality for d and Hölder inequality:

$$\mathsf{d}(x,z) \le t^{1-1/p}\mathsf{d}_t(x,y) + s^{1-1/p}\mathsf{d}_s(y,z) \le (t+s)^{1-1/p}\left(\mathsf{d}_t(x,y)^p + \mathsf{d}_s(y,z)^p\right)^{1/p}.$$

Given $v : \mathcal{Y} \to \mathbb{R}$, we then write

$$\begin{aligned}
\int_{\mathcal{X}}\mathsf{Q}_{s+t}(v)d\lambda - \int_{\mathcal{Y}}v d\mu &= \int_{\mathcal{X}}\mathsf{Q}_{s+t}(v)d\lambda - \int_{\mathcal{X}}\mathsf{Q}_s(v)d\bar{\lambda} + \int_{\mathcal{X}}\mathsf{Q}_s(v)d\bar{\lambda} - \int_{\mathcal{Y}}v d\mu \\
&\le \left[\int_{\mathcal{X}}\mathsf{Q}_t\left(\mathsf{Q}_s(v)\right)d\lambda - \int_{\mathcal{X}}\mathsf{Q}_s(v)d\bar{\lambda}\right] + \left[\int_{\mathcal{X}}\mathsf{Q}_s(v)d\bar{\lambda} - \int_{\mathcal{Y}}v d\mu\right] \\
&\le \mathsf{Wd}_t^p\left(\bar{\lambda},\lambda\right) + \left[\int_{\mathcal{X}}\mathsf{Q}_s(v)d\bar{\lambda} - \int_{\mathcal{Y}}v d\mu\right].
\end{aligned}$$

The choices $s = 1$ and $t = \varepsilon$ lead to the inequality

$$\left[\int_{\mathcal{X}}\mathsf{Q}_1(v)d\lambda - \int_{\mathcal{Y}}v d\mu\right] \ge \left[\int_{\mathcal{X}}\mathsf{Q}_{1+\varepsilon}(v)d\lambda - \int_{\mathcal{Y}}v d\mu\right] - \varepsilon^{1-p}\mathsf{Wd}^p\left(\bar{\lambda},\lambda\right), \tag{3.14}$$

that we collect for later use. Back to the general case, further bounding from above we find

$$\begin{aligned}
\int_{\mathcal{X}}\mathsf{Q}_{s+t}(v)d\lambda - \int_{\mathcal{Y}}v d\mu &\le \mathsf{Wd}_t^p\left(\bar{\lambda},\lambda\right) + \left[\int_{\mathcal{X}}\mathsf{Q}_s(v)d\bar{\lambda} - \int_{\mathcal{Y}}v d\mu\right] \\
&\le t^{1-p}\mathsf{Wd}^p\left(\bar{\lambda},\lambda\right) + s^{1-p}\mathsf{Wd}^p\left(\mu,\bar{\lambda}\right).
\end{aligned}$$

Finally, taking the supremum over v gives

$$(s+t)^{1-p}\mathsf{Wd}^p(\mu,\lambda) \le t^{1-p}\mathsf{Wd}^p\left(\bar{\lambda},\lambda\right) + s^{1-p}\mathsf{Wd}^p\left(\mu,\bar{\lambda}\right).$$

Choosing $s = \varepsilon/c$ and $t = 1-s$ yields (3.12), provided that

$$(1-\varepsilon/c)^{1-p} \le 1+\varepsilon$$

which holds if $c = c(p) < \infty$ is sufficiently large.

As a consequence of (3.2) and (3.12) we obtain that, Wd^p is (approximately) sub-additive: for any finite or countable Borel partition $\mathcal{X} = \bigcup_k \Omega_k$, it holds

$$\mathsf{Wd}^p(\mu,\lambda) \le (1+\varepsilon)\sum_k \mathsf{Wd}^p\left(\chi_{\Omega_k}\mu, \frac{\mu(\Omega_k)}{\lambda(\Omega_k)}\chi_{\Omega_k}\lambda\right) + \frac{c}{\varepsilon^{p-1}}\mathsf{Wd}^p\left(\sum_k \frac{\mu(\Omega_k)}{\lambda(\Omega_k)}\chi_{\Omega_k}\lambda, \lambda\right), \tag{3.15}$$

where $c = c(p) < \infty$. Indeed, it is sufficient to apply (3.12) with

$$\bar{\lambda} := \sum_k \frac{\lambda(\Omega_k)}{\mu(\Omega_k)}\chi_{\Omega_k}\lambda, \tag{3.16}$$

and subsequently (3.2) with

$$\mu_k = \chi_{\Omega_k}\mu \quad \text{and} \quad \lambda_k := \frac{\mu(\Omega_k)}{\lambda(\Omega_k)}\chi_{\Omega_k}\lambda.$$

3.2 Boundary cost

Let us assume that $(\mathcal{X},\mathsf{d})$ is an (actual) compact metric space. Given $p \ge 1$, we define for any $\Omega \subseteq \mathcal{X}$ the following "boundary" pseudo-distance, for $x, y \in \mathcal{X}$:

$$\mathsf{b}_\Omega(x,y) := \min\left\{\mathsf{d}(x,y), (\mathsf{d}(x,\Omega^c)^p + \mathsf{d}(y,\Omega^c)^p)^{1/p}\right\}.$$

We remark that the notation b_Ω is slightly ambiguous, as it does not make explicit the dependence on d and also p, which will be anyway made clear by considering more often the quantity b_Ω^p. It is not difficult to check that b_Ω defines a pseudo-distance such that $\mathsf{b}_\Omega = \mathsf{b}_{\Omega^\circ}$, where Ω° denotes the interior part of Ω. Moreover, $\mathsf{b}_\Omega \le \mathsf{d}$ and

$$\mathsf{b}_\Omega(x,y) = \mathsf{d}(y,\Omega^c) \quad \text{if } x \in \overline{\Omega^c},$$

hence $\mathsf{b}_\Omega(x,x') = 0$ if $x, x' \in \overline{\Omega^c}$. Given measures μ, λ on $\mathcal{X}$ with $\mu(\mathcal{X}) = \lambda(\mathcal{X})$, we write accordingly to general the notation introduced above, with b_Ω in place of d,

$$\mathsf{Wb}_\Omega^p(\mu,\lambda) = \inf_{\pi\in C(\mu,\lambda)} \int_{\mathcal{X}\times\mathcal{X}} \mathsf{b}_\Omega(x,y)^p d\pi(x,y)$$

for the primal problem and

$$\mathsf{Wb}_\Omega^p(\mu,\lambda) = \sup_{v\in\mathcal{S}(\mathsf{b}_\Omega^p,\mathcal{Y})}\left\{\int_{\mathcal{X}} \mathsf{Q}_{\mathsf{b}_\Omega^p}(v)d\lambda - \int_{\mathcal{Y}} v d\mu\right\} \tag{3.17}$$

for the dual formulation. Let us notice that if $\mu(\Omega) = \lambda(\Omega)$, it holds

$$\mathsf{Wb}_\Omega^p(\mu,\lambda) = \mathsf{Wb}_\Omega^p(\chi_\Omega\mu, \chi_\Omega\lambda),$$

for one can transport without cost the respective masses in Ω^c. In all what follows, we always consider $\mathsf{Wb}^p_\Omega(\mu,\lambda)$ for measures such that $\mu(\Omega)=\lambda(\Omega)$. On the dual side, from (3.6) it follows that $\mathsf{Q}_{\mathsf{b}^p_\Omega}(v)$ is constant on $\overline{\Omega^c}$, and in particular

$$\mathsf{Q}_{\mathsf{b}^p_\Omega}(v)(x)=\inf_{y\in\mathcal{Y}}\{v(y)+\mathsf{d}(y,\Omega^c)^p\}\quad\text{for every } x\in\overline{\Omega^c}.$$

We denote in what follows with $\mathcal{S}_0(\mathsf{b}^p_\Omega,\mathcal{Y})$ the subset of $\mathcal{S}(\mathsf{b}^p_\Omega,\mathcal{Y})$ such that $\mathsf{Q}_{\mathsf{b}^p_\Omega}(v)=0$ on $\overline{\Omega^c}$, i.e.,

$$\inf_{y\in\mathcal{Y}}\{v(y)+\mathsf{d}(y,\Omega^c)^p\}=0.$$

For $v\in\mathcal{S}_0(\mathsf{b}^p_\Omega,\mathcal{Y})$ it also holds $v(y)=0$ for every $y\in\mathcal{Y}\cap\overline{\Omega^c}$. Indeed, we have immediately that

$$v(y)\geq-\mathsf{d}(y,\Omega^c)^p\quad\text{for every } y\in\mathcal{Y},\tag{3.18}$$

and in particular $v(y)\geq 0$ for $y\in\mathcal{Y}\cap\overline{\Omega^c}$. Moreover, for a general $x\in\mathcal{X}$, we find

$$\begin{aligned}\mathsf{Q}_{\mathsf{b}^p_\Omega}(v)(x)&=\inf_{y\in\mathcal{Y}}\{v(y)+\mathsf{b}_\Omega(x,y)^p\}\\&\leq\inf_{y\in\mathcal{Y}}\{v(y)+\mathsf{d}(y,\Omega^c)^p\}+\mathsf{d}(x,\Omega^c)^p=\mathsf{d}(x,\Omega^c)^p.\end{aligned}\tag{3.19}$$

This inequality entails, for $y\in\mathcal{Y}\cap\overline{\Omega^c}$, that

$$\begin{aligned}v(y)&=\sup_{x\in\mathcal{X}}\left\{\mathsf{Q}_{\mathsf{b}^p_\Omega}(v)(x)-\mathsf{b}_\Omega(x,y)^p\right\}\\&\leq\sup_{x\in\mathcal{X}}\{\mathsf{d}(x,\Omega^c)^p-\mathsf{d}(x,\Omega^c)^p\}=0,\end{aligned}$$

hence $v(y)=0$ for $y\in\mathcal{Y}\cap\overline{\Omega^c}$.

Let us notice that, by (3.7) and the analogue property for $\hat{\mathsf{Q}}$, one can always restrict the supremum in (3.17) to $v\in\mathcal{S}_0(\mathsf{b}^p_\Omega,\mathcal{Y})$. Furthermore, since $x\mapsto\mathsf{b}_\Omega(x,y)$ is 1-Lipschitz with respect to d (for every y), and

$$\mathsf{b}_\Omega(\mathcal{X})=\sup_{x,x'\in\mathcal{X}}\mathsf{b}_\Omega(x,x')\leq\mathsf{d}(\Omega)$$

we obtain by (3.10) and (3.9) that $\mathsf{Q}_{\mathsf{b}^p_\Omega}(v)$ is Lipschitz (with respect to d) with

$$\mathrm{Lip}_{\mathcal{X}}\left(\mathsf{Q}_{\mathsf{b}^p_\Omega}(v)\right)\leq p\mathsf{d}(\Omega)^{p-1},\quad\text{and}\quad\mathrm{osc}_{\mathcal{X}}\left(\mathsf{Q}_{\mathsf{b}^p_\Omega}(v)\right)\leq\mathrm{osc}_{\mathcal{Y}}(v)+\mathsf{d}(\Omega,\mathcal{Y})^p.\tag{3.20}$$

Similarly,

$$\mathrm{Lip}_{\mathcal{Y}}\left(\hat{\mathsf{Q}}_{\mathsf{b}^p_\Omega}(u)\right)\leq p\mathsf{d}(\Omega)^{p-1},\quad\text{and}\quad\mathrm{osc}_{\mathcal{Y}}\left(\hat{\mathsf{Q}}_{\mathsf{b}^p_\Omega}(u)\right)\leq\mathrm{osc}_{\mathcal{X}}(u)+\mathsf{d}(\Omega,\mathcal{Y})^p.$$

In particular, any $v\in\mathcal{S}(\mathsf{b}^p_\Omega,\mathcal{Y})$ is Lipschitz with

$$\mathrm{Lip}_{\mathcal{Y}}(v)\leq p\mathsf{d}(\Omega)^{p-1},\quad\text{hence}\quad\mathrm{osc}_{\mathcal{Y}}(v)\leq p\mathsf{d}(\Omega)^p,\quad\mathrm{osc}_{\mathcal{X}}\left(\mathsf{Q}_{\mathsf{b}^p_\Omega}(v)\right)\leq p\mathsf{d}(\Omega)^p+\mathsf{d}(\Omega,\mathcal{Y})^p.\tag{3.21}$$

If moreover $v\in\mathcal{S}_0(\mathsf{b}^p_\Omega,\mathcal{Y})$, then

$$\|v\|_\infty\leq p\mathsf{d}(\Omega)^p,\quad\text{and}\quad\left\|\mathsf{Q}_{\mathsf{b}^p_\Omega}(v)\right\|_\infty\leq p\mathsf{d}(\Omega)^p+\mathsf{d}(\Omega,\mathcal{Y})^p.\tag{3.22}$$

The quantity $\mathsf{Wb}^p_\Omega(\mu,\lambda)$ enjoys the following approximate super-additivity property: for any finite Borel partition $\Omega=\bigcup_k\Omega_k$, it holds

$$\mathsf{Wb}^p_\Omega(\mu,\lambda)\geq(1+\varepsilon)^{1-p}\sum_k\mathsf{Wb}^p_{\Omega_k}\left(\mu,\frac{\mu(\Omega_k)}{\lambda(\Omega_k)}\lambda\right)-\varepsilon^{1-p}\mathsf{Wd}^p_\Omega\left(\sum_k\frac{\mu(\Omega_k)}{\lambda(\Omega_k)}\chi_{\Omega_k}\lambda,\lambda\right),\tag{3.23}$$

for any $\varepsilon\in(0,1)$. To prove it, we rely on the following construction.

Lemma 3.1 (gluing)**.** *Consider a finite Borel partition $\Omega = \bigcup_k \Omega_k$ such that, for every $x \in \Omega_k$, $y \in \Omega_h$ with $h \neq k$, it holds*

$$\mathsf{d}(x,y) \geq \mathsf{d}(x,\Omega_k^c) + \mathsf{d}(y,\Omega_h^c). \tag{3.24}$$

For every k, let $v_k \in \mathcal{S}_0(\mathsf{b}^p_{\Omega_k}, \mathcal{Y})$ and define $v : \mathcal{Y} \to \mathbb{R}$, $v(y) := v_k(y)$ for $y \in \mathcal{Y} \cap \overline{\Omega_k}$ and $v(y) = 0$ for $y \in \mathcal{Y} \cap \overline{\Omega^c}$. Then, v is well-defined, with

$$\|v\|_\infty \leq p \max_k \mathsf{d}(\Omega_k)^p, \quad \left\|\mathsf{Q}_{\mathsf{b}^p_\Omega}(v)\right\|_\infty \leq p \max_k \mathsf{d}(\Omega_k)^p + \mathsf{d}(\Omega,\mathcal{Y})^p \tag{3.25}$$

and it holds, for every k,

$$\mathsf{Q}_{\mathsf{b}^p_\Omega}(v) \geq \mathsf{Q}_{\mathsf{b}^p_{\Omega_k}}(v_k) \quad \text{on } \Omega_k. \tag{3.26}$$

Remark 3.2. Condition (3.24) is always satisfied if $(\mathcal{X},\mathsf{d})$ is a length space.

Proof. We notice that if $y \in \mathcal{Y}$ belongs both to $\overline{\Omega_k}$ and $\overline{\Omega_h}$ with $h \neq k$, it follows that $y \in \overline{\Omega_k^c}$ and $y \in \overline{\Omega_h^c}$, hence $v_k(y) = v_h(y) = 0$ and thus v is well-defined. Inequalities (3.25) are immediate from (3.22). To prove (3.26), fix k and $x \in \overline{\Omega_k}$. If $y \in \mathcal{Y} \cap \overline{\Omega_k}$, then,

$$\mathsf{d}(x,\Omega^c) \geq \mathsf{d}(x,\Omega_k^c) \quad \text{and} \quad \mathsf{d}(y,\Omega^c) \geq \mathsf{d}(y,\Omega_k^c),$$

hence

$$\mathsf{b}_\Omega(x,y)^p \geq \mathsf{b}_{\Omega_k}(x,y)^p$$

and therefore

$$\inf_{y \in \mathcal{Y} \cap \overline{\Omega_k}} \{v(y) + \mathsf{b}_\Omega(x,y)^p\} \geq \inf_{y \in \mathcal{Y} \cap \overline{\Omega_k}} \{v_k(y) + \mathsf{b}_{\Omega_k}(x,y)^p\} \geq \mathsf{Q}_{\mathsf{b}^p_{\Omega_k}}(v_k)(x).$$

Using that from (3.19) with Ω_k instead of Ω and v_k instead of v,

$$\mathsf{d}(x,\Omega_k^c)^p \geq \mathsf{Q}_{\mathsf{b}^p_{\Omega_k}}(v_k)(x),$$

in order to prove (3.26), it is thus enough to argue that

$$\inf_{y \in \mathcal{Y} \cap \overline{\Omega_k^c}} \{v(y) + \mathsf{b}_\Omega(x,y)^p\} \geq \mathsf{d}(x,\Omega_k^c)^p.$$

Given $y \in \mathcal{Y} \cap \overline{\Omega_k^c}$, we distinguish two cases. Assume first that $y \in \overline{\Omega_h}$ for some $h \neq k$, then by (3.18) with Ω_h instead of Ω and v_h instead of v,

$$v(y) = v_h(y) \geq -\mathsf{d}(y,\Omega_h^c)^p.$$

Moreover,

$$\begin{aligned}
\mathsf{b}_\Omega(x,y)^p &= \min\{\mathsf{d}(x,y)^p, \mathsf{d}(x,\Omega^c)^p + \mathsf{d}(y,\Omega^c)^p\} \\
&\overset{(3.24)}{\geq} \min\left\{\left(\mathsf{d}(x,\Omega_k^c) + \mathsf{d}(y,\Omega_h^c)\right)^p, \mathsf{d}(x,\Omega_k^c)^p + \mathsf{d}(y,\Omega_h^c)^p\right\} \\
&\geq \mathsf{d}(x,\Omega_k^c)^p + \mathsf{d}(y,\Omega_h^c)^p.
\end{aligned}$$

Therefore, we find in this first case that

$$v(y) + \mathsf{b}_\Omega(x,y)^p \geq \mathsf{d}(x,\Omega_k^c)^p. \tag{3.27}$$

If instead $y \in \overline{\Omega^c}$, so that $v(y) = 0$, and $\mathsf{b}_\Omega(x,y) = \mathsf{d}(x,\Omega^c) \geq \mathsf{d}(x,\Omega_k^c)$, we conclude that (3.27) holds as well. □

In order to establish (3.23), we need to apply the result above to the modified cost $\mathsf{b}_{\Omega,t}(x,y) = t^{\frac{1}{p}-1}\mathsf{b}_\Omega(x,y)$, for $t>0$. We notice that the properties and definitions introduced above easily extend from the case $t=1$ to the general case, with minor modifications in the bounds – to see this, it is sufficient to replace the original d with d_t.

Consider then $v_k \in \mathcal{S}_0(\mathsf{d}^p_{\Omega_k,(1+\varepsilon)}, \mathcal{Y})$ that is optimal in the dual formulation of the cost, so that

$$\int_{\Omega_k} \mathsf{Q}_{\mathsf{d}^p_{\Omega,(1+\varepsilon)}}(v_k)d\lambda - \int_{\mathcal{Y}\cap\Omega_k} v_k d\mu = (1+\varepsilon)^{p-1}\mathsf{Wb}^p_{\Omega_k}\left(\mu, \frac{\mu(\Omega_k)}{\lambda(\Omega_k)}\lambda\right).$$

By (3.14) with $\bar\lambda$ as in the approximate sub-additivity (3.16), $\mathsf{Q} = \mathsf{Q}_{\mathsf{b}^p_\Omega}$ and v as provided by Lemma 3.1, we find

$$\begin{aligned}
\left[\int_{\mathcal{X}} \mathsf{Q}_{\mathsf{b}^p_\Omega}(v)d\lambda - \int_{\mathcal{Y}} v d\mu\right] &\geq \left[\int_{\mathcal{X}} \mathsf{Q}_{\mathsf{b}^p_\Omega,(1+\varepsilon)}(v)d\lambda - \int_{\mathcal{Y}} v d\mu\right] - \varepsilon^{1-p}\mathsf{Wd}^p\left(\bar\lambda,\lambda\right)\\
&\geq \sum_k \left[\int_{\Omega_k} \mathsf{Q}_{\mathsf{b}^p_\Omega,(1+\varepsilon)}(v)d\lambda - \int_{\mathcal{Y}\cap\Omega_k} v d\mu\right] - \varepsilon^{1-p}\mathsf{Wd}^p\left(\bar\lambda,\lambda\right)\\
&\overset{(3.26)}{\geq} \sum_k \left[\int_{\Omega_k} \mathsf{Q}_{\mathsf{d}^p_{\Omega_k},(1+\varepsilon)}(v_k)d\lambda - \int_{\mathcal{Y}\cap\Omega_k} v_k d\mu\right] - \varepsilon^{1-p}\mathsf{Wd}^p\left(\bar\lambda,\lambda\right)\\
&= (1+\varepsilon)^{1-p}\sum_k \mathsf{Wb}^p_{\Omega_k}\left(\mu, \frac{\mu(\Omega_k)}{\lambda(\Omega_k)}\lambda\right) - \varepsilon^{1-p}\mathsf{Wd}^p\left(\bar\lambda,\lambda\right).
\end{aligned}$$

If we further bound from above the left-hand side

$$\int_{\mathcal{X}} \mathsf{Q}_{\mathsf{b}^p_\Omega}(v)d\lambda - \int_{\mathcal{Y}} v d\mu \leq \mathsf{Wb}^p_\Omega(\mu,\lambda),$$

inequality (3.23) then follows. For later use, let us notice that we may also bound from above the left-hand side replacing v with

$$\tilde v := \hat{\mathsf{Q}}_{\mathsf{b}^p_\Omega}\left(\mathsf{Q}_{\mathsf{b}^p_\Omega}(v)\right) \in \mathcal{S}\left(\mathsf{b}^p_\Omega, \mathcal{Y}\right) \tag{3.28}$$

for which $\mathsf{Q}_{\mathsf{b}^p_\Omega}(v) = \mathsf{Q}_{\mathsf{b}^p_\Omega}(\tilde v)$ (recall that $\mathsf{Q}\circ\hat{\mathsf{Q}}\circ\mathsf{Q} = \mathsf{Q}$). We find

$$\int_{\mathcal{X}} \mathsf{Q}_{\mathsf{b}^p_\Omega}(\tilde v)d\lambda - \int_{\mathcal{Y}} \tilde v d\mu \geq (1+\varepsilon)^{1-p}\sum_k \mathsf{Wb}^p_{\Omega_k}\left(\mu, \frac{\mu(\Omega_k)}{\lambda(\Omega_k)}\lambda\right) - \varepsilon^{1-p}\mathsf{Wd}^p\left(\sum_k \frac{\mu(\Omega_k)}{\lambda(\Omega_k)}\chi_{\Omega_k}\lambda, \lambda\right). \tag{3.29}$$

Let us also notice that by (3.25)

$$\left\|\mathsf{Q}_{\mathsf{b}^p_\Omega}(\tilde v)\right\|_\infty \leq p\max_k \mathsf{d}(\Omega_k)^p + \mathsf{d}(\Omega,\mathcal{Y})^p. \tag{3.30}$$

We end this section with a general result providing uniform upper bounds for the optimal transport plan, extending in particular [3, Lemma 4.4] (see also [14, 30]). As always in this section, we consider μ, λ to be measures on $\mathcal{X}$ such that $\mu(\Omega) = \lambda(\Omega)$ and μ supported on $\mathcal{Y}$.

Lemma 3.3. *Let $v \in \mathcal{S}(\mathsf{b}^p_\Omega, \mathcal{Y})$ and set*

$$\ell := \left(\operatorname{osc}_{\mathcal{Y}}(v) + \mathsf{d}(\mathcal{X},\mathcal{Y})^p\right)^{1/p}. \tag{3.31}$$

1. *For every $x \in \partial_{\mathsf{b}^p_\Omega} v(y)$, it holds*

$$\mathsf{b}_\Omega(x,y) \leq \ell \tag{3.32}$$

and therefore (recall the definition (2.3) of $\Omega^{(\ell)}$),

$$x \in \Omega^{(\ell)} \quad \text{or} \quad y \in \Omega^{(\ell)} \quad \Rightarrow \quad \mathsf{b}_\Omega(x,y) = \mathsf{d}(x,y). \tag{3.33}$$

2. *Assume that Ω is midpoint convex, i.e., for every $x, y \in \Omega$ there exists $z \in \Omega$ such that $\mathsf{d}(x,z) = \mathsf{d}(z,y) = \mathsf{d}(x,y)/2$, and that for some $d > 0$, $\lambda_0 > 0$, it holds*

$$\lambda(B_r(x)) \geq \lambda_0 r^d \quad \text{for every } r \leq \ell, \text{ and } \lambda\text{-a.e. } x \in \Omega. \tag{3.34}$$

Let $\pi \in C(\mu,\lambda)$ and $v \in \mathcal{S}(\mathsf{b}^p_\Omega, \mathcal{Y})$ be optimizers for $\mathsf{Wb}^p_\Omega(\mu,\lambda)$ respectively in the primal and dual formulation, then, for π-a.e. (x,y),

$$x \in \Omega^{(2\ell)} \quad \Rightarrow \quad \lambda_0 \mathsf{d}(x,y)^{p+d} \lesssim_p \mathsf{Wb}^p_\Omega(\mu,\lambda)\,. \tag{3.35}$$

Proof. We set for brevity $u := \mathsf{Q}_{\mathsf{b}^p_\Omega}(v)$. To prove *(1)*, we first recall that by definition of the b^p_Ω-sub-differential, we have for $x \in \partial_{\mathsf{b}^p_\Omega} v(y)$,

$$u(x) - v(y) = \mathsf{b}_\Omega(x,y)^p. \tag{3.36}$$

Thus

$$\begin{aligned}\mathsf{b}_\Omega(x,y)^p &\leq \sup_{x' \in \mathcal{X}} u(x') - \inf_{y' \in \mathcal{Y}} v(y') \leq \sup_{y' \in \mathcal{Y}} v(y') - \inf_{y' \in \mathcal{Y}} v(y') + \mathsf{b}_\Omega(\mathcal{X},\mathcal{Y})^p \\ &\leq \operatorname{osc}_{\mathcal{Y}}(v) + \mathsf{d}(\mathcal{X},\mathcal{Y})^p = \ell^p,\end{aligned}$$

having used (3.8). Hence, inequality (3.32) is proved and (3.33) follows from the definitions of $\Omega^{(\ell)}$ and b_Ω.

We now turn to *(2)*. We first recall that since both π and v are optimizers for their respective formulations, it follows that for π-a.e. (x,y) we have $x \in \partial_{\mathsf{b}^p_\Omega} v(y)$ or equivalently (3.36). The identity extends by continuity to every $(x,y) \in \operatorname{supp}\pi$. This is because inequality $\leq$ holds for every (x,y), and integrating with respect to π, we have

$$\int_{\mathcal{X}\times\mathcal{Y}} (u(x) - v(y) - \mathsf{b}_\Omega(x,y)^p)\, d\pi(x,y) = \mathsf{Wb}^p_\Omega(\mu,\lambda) - \mathsf{Wb}^p_\Omega(\mu,\lambda) = 0.$$

Let us notice that (3.36) and the validity of the inequality $\leq$ for general pairs (not in $\operatorname{supp}\pi$) yields the inequality

$$\mathsf{b}_\Omega(x,y)^p + \mathsf{b}_\Omega(x',y')^p \leq \mathsf{b}_\Omega(x,y')^p + \mathsf{b}_\Omega(x,y')^p, \quad \text{for every } (x,y),(x',y') \in \operatorname{supp}\pi. \tag{3.37}$$

Consider $(x,y) \in \operatorname{supp}\pi$ with $x \in \Omega^{(2\ell)}$, so that in particular $\mathsf{d}(x,y) \leq \ell$. Set for simplicity of notation $\alpha := \mathsf{d}(x,y)/2$ and let z be a midpoint between x and y, so that

$$\mathsf{d}(x,z) = \mathsf{d}(y,z) = \alpha \leq \ell/2,$$

and therefore $\mathsf{d}(z,\Omega^c) \geq 2\ell - \ell/2 \geq \ell$. Consider the ball $B_{\varepsilon\alpha}(z)$, where $\varepsilon = \varepsilon(p) \in (0,1)$ is to be specified below, sufficiently small. Notice that, for every $x' \in B_{\varepsilon\alpha}(z)$, it holds

$$\mathsf{d}(x',x) \leq \mathsf{d}(x',z) + \mathsf{d}(x,z) \leq (1+\varepsilon)\alpha \leq 2\alpha \leq \ell,$$

hence, by the triangle inequality, $x' \in \Omega^{(\ell)}$. By (3.37) applied with $(x',y') \in \operatorname{supp}\pi$ and $x' \in B_{\varepsilon\alpha}(z)$, it follows that

$$\begin{aligned}\mathsf{d}(x,y)^p + \mathsf{d}(x',y')^p &= \mathsf{b}_\Omega(x,y)^p + \mathsf{b}_\Omega(x',y')^p \\ &\leq \mathsf{b}_\Omega(x',y) + \mathsf{b}_\Omega(x,y')^p \\ &\leq \mathsf{d}(x',y)^p + \mathsf{d}(x,y')^p.\end{aligned} \tag{3.38}$$

Then, we use the triangle inequality to bound from above

$$\mathsf{d}(x',y)^p \leq (\mathsf{d}(x',z) + \mathsf{d}(z,y))^p \leq (1+\varepsilon)^p \alpha^p$$

and similarly, with an additional use of the elementary bound $(a+b)^p \le (1+\varepsilon)a^p + c\varepsilon^{-(p-1)}b^p$ where $c = c(p) < \infty$,

$$\begin{aligned}\mathsf{d}(x,y')^p &\le (\mathsf{d}(x,z) + \mathsf{d}(z,x') + \mathsf{d}(x',y'))^p \le ((1+\varepsilon)\alpha^p + \mathsf{d}(x',y'))^p \\ &\le (1+\varepsilon)^{p+1}\alpha^p + \frac{c}{\varepsilon^{p-1}}\mathsf{d}(x',y')^p.\end{aligned}$$

Using these bounds in (3.38) and recalling that $\mathsf{d}(x,y)^p = 2^p\alpha^p$, we find the inequality

$$2^p\alpha^p + \mathsf{d}(x',y')^p \le (2+\varepsilon)(1+\varepsilon)^p\alpha^p + \frac{c(p)}{\varepsilon^{p-1}}\mathsf{d}(x',y')^p.$$

Choosing $\varepsilon = \varepsilon(p) \in (0,1)$ sufficiently small, e.g. such that

$$2^p - (2+\varepsilon)(1+\varepsilon)^p \ge 2^{p-1} - 1,$$

we deduce that

$$\alpha^p \lesssim_p \mathsf{d}(x',y')^p.$$

Recalling that $\mathsf{d}(x',y') = \mathsf{b}_\Omega(x',y')$, and integrating with respect to $(x',y') \in \operatorname{supp}\pi$ such that $x' \in B_{\varepsilon\alpha}(z)$ – and using the assumption (3.34) – we deduce that

$$\lambda_0\alpha^{p+d} \lesssim_p \int_{\mathcal{X}\times\mathcal{X}} \mathsf{b}_\Omega(x',y')^p d\pi(x',y') = \mathsf{Wb}^p_\Omega(\mu,\lambda),$$

hence (3.35) follows. □

3.3 The Euclidean case

We specialize to the Euclidean setting, i.e., assume that $\mathcal{X} \subseteq \mathbb{R}^d$ is compact and convex (e.g. is a sufficiently large cube) and $\mathsf{d}(x,y) = |y-x|$ is the Euclidean distance. We momentarily keep p general, although our main result holds for $p = 2$ only. Given $\Omega \subseteq \mathcal{X}$ Borel, since the function b_Ω is Lipschitz continuous with Lipschitz constant bounded by $p\,\mathrm{diam}(\Omega)^{p-1}$, it follows that any transform $u = \mathsf{Q}_{\mathsf{b}^p_\Omega}(v)$ is Lipschitz and by Rademacher's theorem, Lebesgue-a.e. differentiable. The classical Brenier theorem [16] exploits this fact to characterize the optimal transport plan from an absolutely continuous measure in terms of a map of gradient type: in the following result we extend it to the "boundary" case.

Lemma 3.4. *Let $p > 1$, $p' = p/(p-1)$, λ, μ be measures with $\lambda(\Omega) = \mu(\Omega)$, λ absolutely continuous with respect to the Lebesgue measure and μ supported on $\mathcal{Y}$. Let $\pi \in C(\mu,\lambda)$ and $v \in \mathcal{S}(\mathsf{b}^p_\Omega, \mathcal{Y})$ be respectively optimizers for $\mathsf{Wb}^p_\Omega(\mu,\lambda)$ in the primal and dual formulation and set*

$$u := \mathsf{Q}_{\mathsf{b}^p_\Omega}(v).$$

Then,

$$\int_\Omega |\nabla u|^{p'}\lambda \le p^{p'}\mathsf{Wb}^p_\Omega(\mu,\lambda), \tag{3.39}$$

and with ℓ defined as in (3.31), for π-a.e. (x,y), it holds

$$x \in \Omega^{(\ell)} \Rightarrow y = x - \left(\frac{\nabla u(x)}{p}\right)^{(p'/p)}. \tag{3.40}$$

In (3.40) we used conveniently the notation $x^{(\alpha)} = |x|^{\alpha-1}x$.

Remark 3.5. As a consequence of (3.39) and Lemma 2.3 with $p' = p/(p-1)$ in place of p, we also collect the following bound:

$$\operatorname{osc}_{\mathcal{X}}(u)^{p'+d} \lesssim \operatorname{diam}(\Omega)^{d(p-1)+p'} \mathsf{Wb}^p_\Omega(\mu,\lambda), \tag{3.41}$$

having used that $\operatorname{Lip}_{\mathcal{X}}(u) \le p \operatorname{diam}(\Omega)^{p-1}$. The implicit constant depends on p, d and Ω but is invariant with respect to dilations.

Proof. Both u and $x \mapsto \mathsf{d}(x,\Omega^c)$ are Lipschitz continuous hence differentiable Lebesgue-a.e. in $\mathcal{X}$. Moreover,

$$\left\|\nabla \mathsf{d}(\cdot,\Omega^c)\right\|_\infty \le \operatorname{Lip}_{\mathcal{X}}(\mathsf{d}(\cdot,\Omega^c)) \le 1. \tag{3.42}$$

Let (x,y) be such that (3.36) holds and x is a point of differentiability for both functions. Then, we prove

$$|\nabla u(x)|^{\frac{p}{p-1}} \le p^{\frac{p}{p-1}} \mathsf{b}^p_\Omega(x,y), \tag{3.43}$$

so that integration with respect to π yields (3.39). We consider separately the case $\mathsf{b}_\Omega(x,y) = |x-y|$ and the case $\mathsf{b}_\Omega(x,y) = (\mathsf{d}(x,\Omega^c)^p + \mathsf{d}(y,\Omega^c)^p)^{1/p}$. In the first case, we have by definition of u, for every $x' \in \mathcal{X}$,

$$u(x') \le v(y) + \mathsf{b}_\Omega(x',y)^p \le v(y) + |x'-y|^p,$$

with equality at $x' = x$. Differentiating at $x' = x$ yields

$$\nabla u(x) = p|x-y|^{p-2}(x-y) = p(x-y)^{(p-1)} \tag{3.44}$$

so that

$$|\nabla u(x)|^{\frac{p}{p-1}} = p^{\frac{p}{p-1}}|x-y|^p = p^{\frac{p}{p-1}} \mathsf{b}_\Omega(x,y)^p.$$

If instead $\mathsf{b}_\Omega(x,y) = (\mathsf{d}(x,\Omega^c)^p + \mathsf{d}(y,\Omega^c)^p)^{1/p}$, we have, for every x',

$$u(x') \le v(y) + \mathsf{b}_\Omega(x',y)^p \le v(y) + \mathsf{d}(x',\Omega^c)^p + \mathsf{d}(y,\Omega^c)^p,$$

with equality for $x' = x$. Differentiating again at $x' = x$ yields

$$\nabla u(x) = p\mathsf{d}(x,\Omega^c)^{p-1}\nabla \mathsf{d}(x,\Omega^c).$$

Using (3.42), we find

$$|\nabla u(x)|^{\frac{p}{p-1}} \le p^{\frac{p}{p-1}} \mathsf{d}(x,\Omega^c)^p \le p^{\frac{p}{p-1}} \mathsf{b}_\Omega(x,y)^p,$$

which concludes the proof of (3.43). Finally, to prove (3.40), given $x \in \Omega^{(\ell)}$, by (3.33) we have $|x-y| = \mathsf{b}_\Omega(x,y)$ and therefore identity (3.44) holds. Composing both sides with the function

$$\mathbb{R}^d \ni z \mapsto z^{(p'/p)} = z^{\left(\frac{1}{p-1}\right)} = |z|^{\frac{2-p}{p-1}} z$$

yields (3.40). □

Next, we recall a general upper bound for the optimal transport cost in terms of a negative Sobolev norm, see e.g. [42], but also [5, Proposition 2.3], [35, Theorem 2], and [31] in the context of random optimal transport problems.

Lemma 3.6. *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary, $\eta : \Omega \to (0,\infty)$ be bounded and such that a weighted Poincaré inequality (2.6) holds. Given finite measures $\mu = f\eta$, $\lambda = g\eta$, both absolutely continuous with respect to η, and such that $f, g \in L^p(\eta)$ and for some $g_0 > 0$ one has $g(x) \ge g_0 > 0$ for every $x \in \Omega$, it holds*

$$\mathsf{W}^p_\Omega(\mu,\lambda) \lesssim \frac{1}{g_0^{p-1}} \|f-g\|^p_{H^{-1,p}(\eta)}.$$

Using (2.6) in (3.45) we further deduce the inequality

$$\mathsf{W}^p_\Omega(\mu,\lambda) \lesssim \frac{c_P(\eta,p)^p}{g_0^{p-1}} \|f-g\|^p_{L^p(\eta)}\,. \tag{3.45}$$

Proof. By the triangle inequality and (3.2), we find, for every $\varepsilon \in (0,1)$,

$$\begin{aligned}\mathsf{W}^p_\Omega(\mu,\lambda) &\le (1+\varepsilon)\mathsf{W}^p_\Omega(\mu,(\mu+\lambda)/2) + \frac{c}{\varepsilon^{p-1}}\mathsf{W}^p_\Omega((\mu+\lambda)/2,\lambda)\\ &\le \frac{(1+\varepsilon)}{2}\mathsf{W}^p_\Omega(\mu,\lambda) + \frac{c}{\varepsilon^{p-1}}\mathsf{W}^p_\Omega((\mu+\lambda)/2,\lambda)\,,\end{aligned}$$

and choosing e.g. $\varepsilon = 1/4$ we deduce that

$$\mathsf{W}^p_\Omega(\mu,\lambda) \lesssim \mathsf{W}^p_\Omega((\mu+\lambda)/2,\lambda)\,.$$

Hence, we may assume that $f(x), g(x) \ge g_0/2$ for every $x \in \Omega$. Let now $b \in L^p(\eta)$ be such that $-\operatorname{div}_\eta b = f - g$ (recall the definition (2.12) of div_η). Setting $\rho_t = (1-t)\mu + t\lambda$ and $j_t = b\eta$ we see that the couple (ρ_t, j_t) is admissible for the Benamou-Brenier formulation of $\mathsf{W}^p_\Omega(\mu,\lambda)$, see e.g. [47] thus

$$\mathsf{W}^p_\Omega(\mu,\lambda) \le \int_\Omega \int_0^1 \frac{1}{\rho_t^{p-1}}|j_t|^p dt \lesssim \frac{1}{g_0^{p-1}}\int_\Omega |b|^p \eta.$$

Minimizing over all such b we conclude the proof by (2.14). □

We are now in a position to state and prove the main result of this section, which provides a quantitative upper bound for W^p in terms of Wb^p.

Theorem 3.7. *Let $d \ge 1$, $p > 1$, $p' = p/(p-1)$, $c, r > 0$, and*

1. *$\Omega \subseteq \mathbb{R}^d$ be convex and bounded, $\eta : \mathbb{R}^d \to [0,1]$ be C^2 smooth and such that $\eta(x) = 1$ if $\mathsf{d}(x,\Omega^c) \le r$, $\eta(x) = 0$ if $\mathsf{d}(x,\Omega^c) \ge 2r$, $\|\nabla\eta\|_\infty \le c/r$ and $\left\|\nabla^2\eta\right\|_\infty \le c/r^2$,*
2. *μ be a measure on Ω supported on $\mathcal{Y} \subseteq \overline{\Omega}$ with*

$$\frac{1}{2} \le \frac{\mu(\Omega)}{|\Omega|} \le 2, \quad \text{and set} \quad \lambda := \frac{\mu(\Omega)}{|\Omega|}\chi_\Omega,$$

3. *$v \in \mathcal{S}(\mathsf{b}^p_\Omega, \mathcal{Y})$ be an optimizer in the dual formulation for $\mathsf{Wb}^p_\Omega(\mu,\lambda)$, and set*

$$u := \mathsf{Q}_{\mathsf{b}^p_\Omega}(v), \qquad \ell := (\operatorname{osc}_{\mathcal{Y}}(v) + \mathsf{d}(\Omega,\mathcal{Y})^p)^{1/p}\,.$$

If $r \ge 4\ell$, then, for every $\varepsilon \in (0,1)$ it holds

$$\begin{aligned}\mathsf{W}^p_\Omega(\mu,\lambda) - \mathsf{Wb}^p_\Omega(\mu,\lambda) \lesssim \varepsilon\mathsf{Wb}^p_\Omega(\mu,\lambda) + \frac{1}{\varepsilon^{p-1}}\Bigg[&\mathsf{W}^p_\Omega\left(\frac{\int_\Omega \eta d\mu}{\int_\Omega \eta}\eta, \eta\mu\right)\\ &+ \operatorname{diam}(\Omega)^p\left(\left|\frac{\int_\Omega \eta d\mu}{\int_\Omega \eta} - 1\right|^p \int_\Omega \eta^p + r^{-2p}\left(\mathsf{Wb}^p_\Omega(\mu,\lambda)\right)^{\frac{d+2p}{d+p}}\right)\\ &+ \left\|(\nabla u)^{(p'/p)}\cdot\nabla\eta - \fint_\Omega (\nabla u)^{(p'/p)}\cdot\nabla\eta\right\|^p_{H^{-1,p}(\Omega)}\Bigg]\end{aligned} \tag{3.46}$$

where the implicit constant depends on c, d, p, Ω, (but not on r) and is invariant with respect to dilations of Ω.

Remark 3.8. Let us notice that in the case of $\Omega = (0, L)^d$ a cube of side length $L > 0$, one can easily find $\eta = \eta_{d,L,\delta}$ satisfying assumption *(1)* with $r := \delta L$, for any given $\delta \in (0, 1)$ and with $c = c(d) < \infty$ depending on d only. Indeed, it is sufficient to argue in the case $L = 1$ and $d = 1$, and define a suitable $\xi_\delta : \mathbb{R} \to [0, 1]$ even and such that $\xi_\delta(x) = 1$ for $|x| \leq 1/2 - 2\delta$, $\xi_\delta(x) = 0$ for $|x| \geq 1/2 - \delta$ and then set, for $x = (x_i)_{i=1}^d \in (0, L)^d$,

$$\eta_{d,L,\delta}(x) := 1 - \prod_{i=1}^{d} \xi_\delta(x_i/L - 1/2).$$

A more explicit choice for ξ_δ will be given in Section 4.

Proof. We split the proof into multiple steps.
Step 1 (Notation). Let $\pi \in C(\mu, \lambda)$ be an optimizer in the primal formulation of $\mathsf{Wb}^p_\Omega(\mu, \lambda)$. Write $(\pi(\cdot|x))_{x\in\Omega}$ for the regular conditional distribution of y given x, with respect to π, which is a Markov kernel, that we further use to define, for $x \in \Omega$, the function

$$\pi(\eta)(x) := \int_\Omega \eta(y)\pi(dy|x).$$

Since the marginal distribution of y with respect to π is μ, we notice that

$$\int_\Omega \pi(\eta)d\lambda = \int_{\Omega\times\Omega} \eta(y)d\pi(x, y) = \int_\Omega \eta d\mu.$$

To keep notation simple, we write

$$\kappa := \frac{\int_\Omega \pi(\eta)\lambda}{\int_\Omega \eta\lambda} = \frac{\int_\Omega \eta d\mu}{\int_\Omega \eta\lambda},$$

and

$$\xi(x) := 1 - \eta(x), \qquad \pi(\xi)(x) := 1 - \pi(\eta)(x) = \int_\Omega \xi(y)\pi(dy|x),$$

and notice that $\xi, \pi(\xi) \in [0, 1]$. Moreover, we notice that the measure $\xi(y)\pi(dx, dy)$ on $\Omega \times \Omega$ is a coupling between $\xi\mu$ (the marginal law of y) and $\pi(\xi)\lambda$, i.e.,

$$\xi(y)\pi(dx, dy) \in C(\pi(\xi)\lambda, \xi\mu). \tag{3.47}$$

Indeed, for every $E \subseteq \Omega$ Borel,

$$\int_{\Omega\times\Omega} \chi_E(x)\xi(y)d\pi(x, y) = \int_E \int_\Omega \xi(y)\pi(dy|x)d\lambda(x) = \int_E \pi(\xi)d\lambda.$$

We introduce two auxiliary measures on Ω:

$$\mu_1 := \eta\mu + \pi(\xi)\lambda, \quad \text{and} \quad \mu_2 := \kappa\eta\lambda + \pi(\xi)\lambda,$$

such that μ_2 is absolutely continuous with respect to the Lesbesgue measure. Let us notice that both measures have total mass $\mu(\Omega)$.
Step 2 (Reduction to three bounds). We claim that the thesis follows from two applications of (3.12) and the following three inequalities:

$$\mathsf{W}^p_\Omega(\mu, \mu_1) \leq \mathsf{Wb}^p_\Omega(\mu, \lambda), \tag{3.48}$$
$$\mathsf{W}^p_\Omega(\mu_1, \mu_2) \leq \mathsf{W}^p_\Omega(\kappa\eta\lambda, \eta\mu) \tag{3.49}$$

and

$$\begin{aligned}\mathsf{W}^p_\Omega(\mu_2,\lambda) \lesssim \operatorname{diam}(\Omega)^p \left[|\kappa-1|^p \int_\Omega \eta^p + r^{-2p}\left(\mathsf{Wb}^p_\Omega(\mu,\lambda)\right)^{\frac{d+2p}{d+p}}\right] \\ + \left\|(\nabla u)^{(p'/p)}\cdot\nabla\eta - \fint_\Omega (\nabla u)^{(p'/p)}\cdot\nabla\eta\right\|^p_{H^{-1,p}(\Omega)}.\end{aligned} \tag{3.50}$$

Indeed, for every $\varepsilon\in(0,1)$, it holds

$$\begin{aligned}\mathsf{W}^p_\Omega(\mu,\lambda) &\le (1+\varepsilon)\mathsf{W}^p_\Omega(\mu,\mu_1) + \frac{c_1}{\varepsilon^{p-1}}\mathsf{W}^p_\Omega(\mu_1,\lambda)\\ &\le (1+\varepsilon)\mathsf{W}^p_\Omega(\mu_1,\mu) + \frac{c_2}{\varepsilon^{p-1}}\left[\mathsf{W}^p_\Omega(\mu_1,\mu_2)+\mathsf{W}^p_\Omega(\mu_2,\lambda)\right],\end{aligned}$$

where the second inequality follows from (3.12) with $\varepsilon=1/2$, and the constant $c_1,c_2<\infty$ depend on p only and may be different in both lines (and are unrelated to the constant c in the assumptions). Inserting (3.48), (3.49), and (3.50) yields (3.46).

We then establish separately the validity of the three claimed inequalities.

Step 3 (Proof of (3.48)). Writing $\mu=\eta\mu+\xi\mu$, by (3.2) we have

$$\begin{aligned}\mathsf{W}^p_\Omega(\mu,\mu_1) &= \mathsf{W}^p_\Omega(\eta\mu+\xi\mu,\eta\mu+\pi(\xi)\lambda)\\ &\le \mathsf{W}^p_\Omega(\xi\mu,\pi(\xi)\lambda).\end{aligned}$$

To further bound from above, we use the coupling (3.47) so that

$$\mathsf{W}^p_\Omega(\xi\mu,\pi(\xi)\lambda) \le \int_{\Omega\times\Omega}|y-x|^p\xi(y)d\pi(x,y).$$

The assumption $r\ge 4\ell$ entails in particular that $\xi(y)=0$ if $y\notin\Omega^{(\ell)}$, therefore by (3.33) we deduce that in the integrand above we can replace $|y-x|$ with $\mathsf{b}_\Omega(x,y)$. Therefore,

$$\begin{aligned}\int_{\Omega\times\Omega}|y-x|^p\xi(y)d\pi(x,y) &= \int_{\Omega\times\Omega}\mathsf{b}_\Omega(x,y)^p\xi(y)d\pi(x,y)\\ &\le \int_{\Omega\times\Omega}\mathsf{b}_\Omega(x,y)^p d\pi(x,y)\\ &= \mathsf{Wb}^p_\Omega(\mu,\lambda)\end{aligned}$$

by optimality of π. Thus, (3.48) is proved.

Step 4 (Proof of (3.49)). This inequality is even simpler than the previous one, because, again by (3.2) we have that

$$\begin{aligned}\mathsf{W}^p_\Omega(\mu_2,\mu_1) &= \mathsf{W}^p_\Omega(\kappa\eta\lambda+\pi(\xi)\lambda,\eta\mu+\pi(\xi)\lambda)\\ &\le \mathsf{W}^p_\Omega(\kappa\eta\lambda,\eta\mu).\end{aligned}$$

Step 5 (Proof of (3.50)). We apply Lemma 3.6 on Ω with μ_2 in place of μ (there) and $g_0=1/2$:

$$\mathsf{W}^p_\Omega(\mu_2,\lambda) \lesssim \|\mu_2-\lambda\|^p_{H^{-1,p}(\Omega)}.$$

Writing $\lambda=\pi(\eta)\lambda+\pi(\xi)\lambda$, we find that

$$\mu_2-\lambda = \kappa\eta\lambda-\pi(\eta)\lambda = (\kappa\eta-\pi(\eta))\frac{\mu(\Omega)}{|\Omega|}.$$

Since $\mu(\Omega)/|\Omega|\le 2$, we need only to show that

$$\begin{aligned}\|\kappa\eta-\pi(\eta)\|^p_{H^{-1,p}(\Omega))} \lesssim \operatorname{diam}(\Omega)^p\left[|\kappa-1|^p\int_\Omega\eta^p + r^{-4}\left(\mathsf{Wb}^p_\Omega(\mu,\lambda)\right)^{\frac{d+2p}{d+p}}\right] \\ + \left\|(\nabla u)^{(p'/p)}\cdot\nabla\eta - \fint_\Omega(\nabla u)^{(p'/p)}\cdot\nabla\eta\right\|^p_{H^{-1,p}(\Omega)}.\end{aligned} \tag{3.51}$$

To this aim, recall that the function

$$\mathbb{R}^d \ni x \mapsto u(x) = \mathsf{Q}_{\mathsf{b}^p_\Omega}(v)(x)$$

is Lipschitz continuous, hence differentiable Lebesgue a.e. on Ω. We use this fact to introduce the auxiliary function, defined for Lebesgue a.e. $x \in \Omega$,

$$\rho(x) := \pi(\xi)(x) - \xi(x) + \left(\frac{\nabla u(x)}{p}\right)^{(p'/p)} \cdot \nabla\xi(x) = \eta(x) - \pi(\eta)(x) - \left(\frac{\nabla u(x)}{p}\right)^{(p'/p)} \cdot \nabla\eta(x)$$

where the second identity follows from the definition of ξ. We collect in particular the identity

$$\pi(\eta) = \eta - \left(\frac{\nabla u}{p}\right)^{(p'/p)} \cdot \nabla\eta - \rho.$$

Hence, by the triangle inequality, we find

$$\begin{aligned}
\|\kappa\eta - \pi(\eta)\|^p_{H^{-1,p}(\Omega)} &= \left\|(\kappa-1)\eta + \left(\frac{\nabla u}{p}\right)^{(p'/p)} \cdot \nabla\eta + \rho\right\|^p_{H^{-1,p}(\Omega)} \\
&\lesssim \left\|(\kappa-1)\eta - \fint_\Omega (\kappa-1)\eta\right\|^p_{H^{-1,p}(\Omega)} \\
&\quad + \left\|\nabla u^{(p'/p)} \cdot \nabla\eta - \fint_\Omega \nabla u^{(p'/p)} \cdot \nabla\eta\right\|^p_{H^{-1,p}(\Omega)} \\
&\quad + \left\|\rho - \fint_\Omega \rho\right\|^p_{H^{-1,p}(\Omega)}
\end{aligned}$$

and we bound separately the three terms. For the first one, we use (2.11), so that

$$\left\|(\kappa-1)\eta - \fint_\Omega (\kappa-1)\eta\right\|^p_{H^{-1,p}(\Omega)} \lesssim \operatorname{diam}(\Omega)^p |\kappa-1|^p \int_\Omega \eta^p$$

which we recognize as a contribution in (3.51). The second term already appears in (3.51), hence we focus on the third term.

We use again (2.11), so that

$$\left\|\rho - \fint_\Omega \rho\right\|^p_{H^{-1,p}(\Omega)} \lesssim \operatorname{diam}(\Omega)^p \int_\Omega \rho^p.$$

We see that the proof of (3.51), hence (3.50) and therefore of the entire result is completed if we argue that

$$\int_\Omega \rho^p \lesssim r^{-2p} \left(\mathsf{Wb}^p_\Omega(\mu,\lambda)\right)^{\frac{d+2p}{d+p}}. \tag{3.52}$$

Step 6 (Proof of (3.52)). We notice that $\rho(x) = 0$ for Lebesgue a.e. $x \notin \Omega^{(r/2)}$. Indeed, by construction we have $\xi(x) = 0$ and $\nabla\xi(x) = 0$ for such x's, and furthermore since $\xi(y) \le \chi_{\Omega^{(r)}}(y)$, we have that

$$\begin{aligned}
\int_{\Omega\setminus\Omega^{(r/2)}} \pi(\xi)(x) d\lambda(x) &= \int_{\Omega\times\Omega} \chi_{\Omega\setminus\Omega^{(r/2)}}(x)\xi(y) d\pi(x,y) \\
&\le \int_{\Omega\times\Omega} \chi_{\Omega\setminus\Omega^{(r/2)}}(x)\chi_{\Omega^{(r)}}(y) d\pi(x,y) = 0,
\end{aligned}$$

because for π-a.e. (x,y), if $y \in \Omega^{(r)}$ by (3.33) we have

$$|x-y| \le \ell \le \frac{r}{4},$$

and therefore, by triangle inequality,

$$\mathsf{d}(x,\Omega^c) \geq \mathsf{d}(y,\Omega^c) - |x-y| \geq r - \frac{r}{4} > \frac{r}{2}.$$

Moreover, we have again by (3.40) that for π-a.e. (x,y), if $x \in \Omega^{(r/2)}$, the conditional distribution $\pi(\cdot|x)$ reduces to a Dirac measure at $y = x - \left(\frac{\nabla u(x)}{p}\right)^{(p'/p)}$, hence

$$\pi(\xi)(x) = \xi\left(x - \left(\frac{\nabla u(x)}{p}\right)^{(p'/p)}\right).$$

Therefore, for such x's we recognize that $\rho(x)$ is the remainder of a first order expansion of ξ centered at x and evaluated at y and therefore we can bound

$$|\rho(x)| \lesssim \left\|\nabla^2\xi\right\|_\infty |y-x|^2 \lesssim |y-x|^2 r^{-2}$$

having used that $\nabla^2\xi = -\nabla^2\eta$. Finally, since $r/2 \geq 2\ell$, using (3.35) we can bound from above

$$|y-x|^p \lesssim \left(\mathsf{Wb}^p_\Omega(\mu,\lambda)\right)^{p/(d+p)}.$$

Thus,

$$\begin{aligned}
\int_\Omega \rho^p d\lambda &= \int_{\Omega^{(r/2)}} \rho^p d\lambda \\
&\lesssim r^{-2p}\left(\mathsf{Wb}^p_\Omega(\mu,\lambda)\right)^{p/(d+p)} \int_{\Omega^{(r/2)}\times\Omega} |y-x|^p d\pi(x,y) \\
&\lesssim r^{-2p}\left(\mathsf{Wb}^p_\Omega(\mu,\lambda)\right)^{p/(d+p)} \int_{\Omega^{(r/2)}\times\Omega} \mathsf{b}^p_\Omega(x,y) d\pi(x,y) \\
&\leq r^{-2p}\left(\mathsf{Wb}^p_\Omega(\mu,\lambda)\right)^{\frac{d+2p}{d+p}},
\end{aligned}$$

and the proof of (3.52) is completed. □

3.4 Stability and the quadratic case

Moving towards the end of this section, we specialize to the case $p = 2$, so that $p'/p = 1$ and we are in a position to apply Lemma 2.2 and bound from above the last term in (3.46) by

$$\left\|\nabla u\cdot\nabla\eta - \fint_\Omega \nabla u\cdot\nabla\eta\right\|^2_{H^{-1,2}(\Omega)} \lesssim \left(r^{-2} + \operatorname{diam}(\Omega)^2 r^{-4}\right)\operatorname{Var}_{\Omega^{(r)}}(u). \tag{3.53}$$

As already stated in the introduction, the problem is now to bound from above $\operatorname{Var}_{\Omega^{(r)}}(u)$. For this we rely on recent progress on the uniform convexity properties of the Kantorovich functional from [23] (in the form presented in [22]). While a similar stability result probably holds also for Wb^2_Ω, see e.g. [41, 21], we use instead that thanks to Lemma 3.3, the optimal transport plan for $\mathsf{Wb}^2_\Omega(\mu,\lambda)$, is also optimal for the usual quadratic cost when restricted to $\Omega^{(r)}$ so that [22] applies.

Theorem 3.9. *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded convex set and μ be a measure supported on $\mathcal{Y} \subseteq \overline{\Omega}$ with*

$$\frac{1}{2} \leq \frac{\mu(\Omega)}{|\Omega|} \leq 2, \quad \text{and set} \quad \lambda := \frac{\mu(\Omega)}{|\Omega|}\chi_\Omega. \tag{3.54}$$

Then, for every $v \in \mathcal{S}(\mathsf{b}^2_\Omega,\mathcal{Y})$ optimizer for the dual formulation of $\mathsf{Wb}^2_\Omega(\mu,\lambda)$, and every $\tilde{v} \in \mathcal{S}(\mathsf{b}^2_\Omega,\mathcal{Y})$, it holds

$$\operatorname{Var}_{\Omega^{(3\ell)}}(u-\tilde{u}) \lesssim \operatorname{diam}(\Omega)^2\left[\mathsf{Wb}^2_\Omega(\mu,\lambda) - \left(\int_\Omega \tilde{u}d\lambda - \int_{\mathcal{Y}} \tilde{v}d\mu\right)\right] \tag{3.55}$$

where $u := \mathsf{Q}_{\mathsf{b}^2_\Omega}(v)$, $\tilde{u} := \mathsf{Q}_{\mathsf{b}^2_\Omega}(\tilde{v})$ *and*

$$\ell := \left(\max\left\{\mathrm{osc}_{\mathcal{Y}}(v), \mathrm{osc}_{\mathcal{Y}}(\tilde{v})\right\} + \mathsf{d}(\mathcal{X},\mathcal{Y})^2\right)^{1/2}. \tag{3.56}$$

Proof. Up to multiplying μ and λ by $|\Omega|/\mu(\Omega)$ we may assume that $\mu(\Omega) = |\Omega|$ and thus $\lambda = \chi_\Omega$. We start by noting that since $v, \tilde{v} \in \mathcal{S}(\mathsf{b}^2_\Omega, \mathcal{Y})$, we may apply Lemma 3.3 to both functions and get that for $v' \in \{v, \tilde{v}\}$ and $u' := \mathsf{Q}_{\mathsf{b}^2_\Omega}(v')$ we have for $x \in \Omega^{(\ell)}$

$$u'(x) = \inf_{y\in\mathcal{Y}\cap\Omega} v'(y) + \mathsf{b}^2_\Omega(x,y) = \inf_{y\in\mathcal{Y}\cap\Omega^{(\ell)}} v'(y) + |x-y|^2$$

so that u' is d^2-concave in $\Omega^{(\ell)}$. Similarly, v' is also d^2-concave in $\Omega^{(\ell)}$ and $u' = \mathsf{Q}_{\mathsf{d}^2}(v')$ in $\Omega^{(2\ell)}$. Let then π be an optimizer for $\mathsf{Wb}^2_\Omega(\mu,\lambda)$ and define $\hat{\pi}$ by

$$\int \xi(x,y)d\hat{\pi}(x,y) = \int \xi(x,y)1_{\Omega^{(3\ell)}}(x)d\pi(x,y).$$

Notice that this is well-defined since the first marginal of π is absolutely continuous with respect to the Lebesgue measure. By definition the first marginal of $\hat{\pi}$ is $1_{\Omega^{(3\ell)}}$. We let $\hat{\mu}$ be its second marginal. Since for $(x,y) \in \mathrm{supp}\,\hat{\pi} \subset \mathrm{supp}\,\pi$ we have $x \in \partial_{\mathsf{b}^2_\Omega} v(y) \cap \Omega^{(3\ell)} \subset \partial v(y)$, we get that $\hat{\pi}$ is an optimizer for $\mathsf{W}^2(\hat{\mu}, 1_{\Omega^{(3\ell)}})$. Moreover using once more that on $\mathrm{supp}\,\hat{\pi}$ we have $x \in \partial v(y)$ we also have that v is an optimizer for the dual version of $\mathsf{W}^2(\hat{\mu}, 1_{\Omega^{(3\ell)}})$.

Thanks to [22, Corollary 1.31], with

$$\phi^0(x) := \frac{1}{2}(|x|^2 - u(x)), \quad \psi^0(y) := \frac{1}{2}(|y|^2 + v(y)), \qquad \text{and}$$
$$\phi^1(x) := \frac{1}{2}(|x|^2 - \tilde{u}(x)), \quad \psi^1(y) := \frac{1}{2}(|y|^2 + \tilde{v}(y))$$

we have

$$\begin{aligned}\mathrm{Var}_{\Omega^{(3\ell)}}(u-\tilde{u}) &= \mathrm{Var}_{\Omega^{(3\ell)}}(\phi^0 - \phi^1)\\ &\lesssim \mathrm{diam}(\Omega)^2\left[\int_{\Omega^{(3\ell)}}\phi^1 dx - \int_{\Omega^{(3\ell)}}\phi^0 dx + \langle\psi^1-\psi^0, (\nabla\phi^0)\#1_{\Omega^{(3\ell)}}\rangle\right].\end{aligned}$$

By (3.40), we have $(\nabla\phi^0)_\sharp 1_{\Omega^{(3\ell)}} = \hat{\mu}$ (where $\sharp$ denotes the push-forward) and thus

$$\begin{aligned}\int_{\Omega^{(3\ell)}}\phi^1 dx - \int_{\Omega^{(3\ell)}}\phi^0 dx + \langle\psi^1-\psi^0, (\nabla\phi^0)_\sharp 1_{\Omega^{(3\ell)}}\rangle &= \int\left[\phi^1(x)+\phi^0(x)+\psi^1(y)-\psi^0(y)\right]d\hat{\pi}(x,y)\\ &= \int\left[u(x)-\tilde{u}(x)+\tilde{v}(y)-v(y)\right]d\hat{\pi}(x,y).\end{aligned}$$

Since $\hat{\pi} \le \pi$ and since on $\mathrm{supp}\,\pi$,

$$u(x) - \tilde{u}(x) + \tilde{v}(y) - v(y) = \mathsf{b}^2_\Omega(x,y) - (\tilde{u}(x) - \tilde{v}(y)) \ge 0$$

we finally obtain

$$\begin{aligned}\mathrm{Var}_{\Omega^{(3\ell)}}(u-\tilde{u}) &\lesssim \mathrm{diam}(\Omega)^2\int\left[\mathsf{b}^2_\Omega(x,y) - (\tilde{u}(x)-\tilde{v}(y))\right]d\pi(x,y)\\ &= \mathrm{diam}(\Omega)^2\left[\mathsf{Wb}^2_\Omega(\mu,\lambda) - \left(\int_\Omega \tilde{u}d\lambda - \int_{\mathcal{Y}}\tilde{v}d\mu\right)\right].\end{aligned}$$

This concludes the proof of (3.55). □

We apply the above result to strengthen the super-additivity inequality (3.23).

Corollary 3.10. *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded convex set and μ be a measure supported on $\mathcal{Y} \subseteq \overline{\Omega}$ satisfying (3.54). Consider a finite partition $\Omega = \bigcup_k \Omega_k$ such that $\mu(\Omega_k) > 0$ for every k and let $v \in \mathcal{S}(\mathsf{b}^2_\Omega, \mathcal{Y})$ be an optimizer for the dual formulation of $\mathsf{Wb}^2_\Omega(\mu, \mu(\Omega)/|\Omega|)$. Then, for every $\varepsilon \in (0,1)$, it holds*

$$\begin{aligned}\mathrm{Var}_{\Omega^{(3\ell)}}\left(\mathsf{Q}_{\mathsf{b}^2_\Omega}(v)\right) \lesssim \mathrm{diam}(\Omega)^2 \Bigg[\mathsf{Wb}^2_\Omega\left(\mu, \frac{\mu(\Omega)}{|\Omega|}\right) - (1+\varepsilon)^{-1} \sum_k \mathsf{Wb}^2_{\Omega_k}\left(\mu, \frac{\mu(\Omega_k)}{|\Omega_k|}\right) \\ + \varepsilon^{-1} \mathsf{W}^2\left(\sum_k \frac{\mu(\Omega_k)}{|\Omega_k|}\chi_{\Omega_k}, \frac{\mu(\Omega)}{|\Omega|}\right)\Bigg] + |\Omega| \max_k \mathrm{diam}(\Omega_k)^4,\end{aligned} \tag{3.57}$$

with

$$\ell := 2 \max_k \mathrm{diam}(\Omega_k) \tag{3.58}$$

Proof. We apply Theorem 3.9 on $\mathcal{X} = \overline{\Omega}$, with $\tilde{v}$ as in (3.28) and consequently $\tilde{u} := \mathsf{Q}_{\mathsf{b}^2_\Omega}(\tilde{v})$ so that in particular (3.30) holds. Since $\mu(\Omega_k) > 0$ for every k, we deduce that $\mathcal{Y} \cap \Omega_k \neq \emptyset$ for every k, hence

$$\mathsf{d}(\Omega, \mathcal{Y}) \leq \max_k \mathrm{diam}(\Omega_k) \quad \text{and} \quad \max\left\{\mathrm{osc}_{\mathcal{Y}}(\tilde{v}), \mathrm{osc}_\Omega(\tilde{u})\right\} \leq 3 \max_k \mathrm{diam}(\Omega_k)^2. \tag{3.59}$$

Using (3.55) and (3.29) with $\lambda = \mu(\Omega)/|\Omega|$ (recall that we identify absolutely continuous measures with their densities) we find

$$\begin{aligned}\mathrm{Var}_{\Omega^{(3\ell)}}(u - \tilde{u}) &\lesssim \mathrm{diam}(\Omega)^2 \left[\mathsf{Wb}^2_\Omega(\mu, \lambda) - \left[\int_\Omega \mathsf{Q}_{\mathsf{b}^2_\Omega}(\tilde{v}) d\lambda - \int_{\mathcal{Y}} \tilde{v} d\mu\right]\right] \\ &\lesssim \mathrm{diam}(\Omega)^2 \Bigg[\mathsf{Wb}^2_\Omega(\mu, \lambda) - (1+\varepsilon)^{-1} \sum_k \mathsf{Wb}^2_{\Omega_k}\left(\mu, \frac{\mu(\Omega_k)}{\lambda(\Omega_k)}\lambda\right) \\ &\quad + \varepsilon^{-1} \mathsf{W}^2\left(\sum_k \frac{\mu(\Omega_k)}{\lambda(\Omega_k)}\chi_{\Omega_k}\lambda, \lambda\right)\Bigg].\end{aligned}$$

The thesis then follows from the triangle inequality for the variance and (2.5):

$$\begin{aligned}\mathrm{Var}_{\Omega^{(3\ell)}}(u) &\lesssim \mathrm{Var}_{\Omega^{(3\ell)}}(u - \tilde{u}) + \mathrm{Var}_{\Omega^{(3\ell)}}(\tilde{u}) \lesssim \mathrm{Var}_{\Omega^{(3\ell)}}(u - \tilde{u}) + |\Omega| \,\mathrm{osc}_\Omega(\tilde{u})^2 \\ &\overset{(3.59)}{\lesssim} \mathrm{Var}_{\Omega^{(3\ell)}}(u - \tilde{u}) + |\Omega| \max_k \mathrm{diam}(\Omega_k)^4.\end{aligned}$$

□

4 Application to random optimal transport

The aim of this section is to prove Theorem 1.2. We actually establish a more general version which applies to random Borel measures, under the following assumptions.

Assumption 4.1. *Let μ denote a random Borel measure on $\mathbb{R}^d$ such that*

i) (stationarity) for every $v \in \mathbb{R}^d$, it holds $\theta_v \mu = \mu$ in law, where $\theta_v \mu(A) = \mu(A + v)$ for every A Borel,

ii) (integrability) for every bounded Borel $A \subseteq \mathbb{R}^d$, $\mu(A)$ is integrable.

iii) (concentration) there exists $\alpha \in [0, d)$ *such that for every cube* $Q \subseteq \mathbb{R}^d$ *with* $\mathrm{diam}(Q) \geq 1$ *and bounded Borel* $\eta : Q \to [0, \infty)$*, it holds for every* $q \geq 1$*,*

$$\left\| \int_Q \eta d\mu - \mathbb{E}\left[\int_Q \eta d\mu \right] \right\|_{L^q(\mathbb{P})} \lesssim_q \mathrm{diam}(Q)^{(d+\alpha)/2} \|\eta\|_\infty . \tag{4.1}$$

Notice that, by stationarity and integrability, the function $A \mapsto \mathbb{E}[\mu(A)]$ is a translation invariant (σ-finite) measure, hence for some constant $c \in [0, \infty)$ it holds

$$\mathbb{E}[\mu(A)] = c|A|.$$

Without loss of generality, we assume in what follows that $c = 1$, so that $\mathbb{E}[\mu(A)] = |A|$ (if $c = 0$ the statements become trivial).

Notable examples of random measures μ satisfying Assumption 4.1 include:

a) The Poisson point process, that can be defined as the limit $\mu = \lim_{r\to\infty} \mu_r$, where

$$\mu_r := \sum_{i=1}^{N_r} \delta_{X_i},$$

where N_r denotes a Poisson random variable with $\mathbb{E}[N_r] = |\mathrm{B}(r)|$ and $(X_i)_{i=1}^\infty$ are i.i.d. (and independent from N_r) uniformly distributed on the ball $\mathrm{B}(r)$. In this case, (4.1) holds with $\alpha = 0$.

b) The Brownian interlacement occupation measure [38], that can be defined similarly as a limit $\mu_\infty = \lim_{r\to\infty} \mu_r$, where in this case

$$\mu_r := \sum_{i=1}^{N_r} \int_0^\infty \delta_{B^i_s + X_i} ds$$

and N_r denotes a Poisson random variable with $\mathbb{E}[N_r] = \mathrm{Cap}(\mathrm{B}(r))$ (the Newtonian capacity) $(X_i)_{i=1}^\infty$ are uniformly distributed on the sphere $\partial\mathrm{B}(r)$ and $(B^i)_{i=1}^\infty$ are independent standard Brownian motions on $\mathbb{R}^d$ – and all the listed random variables are independent. In this case (4.1) holds with $\alpha = 2$ as proved in Section 4.3 below.

Generalizing ideas and tools from [8, 31, 4, 32], in [38, Proposition A.2] the following existence result is established.

Proposition 4.2. *Let* μ *be a random measure on* $\mathbb{R}^d$ *satisfying i), ii), iii) in Assumption 4.1 with* $d > 2 + \alpha$*. Then, for every* $p \geq 1$*, the following limit exists:*

$$\lim_{L\to\infty} \frac{1}{L^d} \mathbb{E}\left[\mathsf{W}^p_{(0,L)^d}\left(\mu, \frac{\mu((0,L)^d)}{L^d}\right)\right] \in [0, \infty).$$

In particular, for every $p \geq 1$ it holds

$$\mathbb{E}\left[\mathsf{W}^p_{(0,L)^d}\left(\mu, \frac{\mu((0,L)^d)}{L^d}\right)\right] \lesssim L^d.$$

For Wb a similar statement can be obtained with a much simpler proof, see [18, Proposition A.3] in the case of the Poisson point process. For $p \geq 1$, the following limit exists:

$$\lim_{L\to\infty} \frac{1}{L^d} \mathbb{E}\left[\mathsf{Wb}^p_{(0,L)^d}\left(\mu, \frac{\mu((0,L)^d)}{L^d}\right)\right] \in [0, \infty).$$

Our main result is that if Assumption 4.1 holds then the two limits coincide when $p = 2$.

Theorem 4.3. *Let μ be a random measure on $\mathbb{R}^d$ satisfying Assumption 4.1 with*

$$d > 2 + \alpha.$$

Then, the following two limits (exist and) coincide:

$$\lim_{L\to\infty} \frac{1}{L^d}\mathbb{E}\left[\mathsf{W}^2_{(0,L)^d}\left(\mu, \frac{\mu((0,L)^d)}{L^d}\right)\right] = \lim_{L\to\infty} \frac{1}{L^d}\mathbb{E}\left[\mathsf{Wb}^2_{(0,L)^d}\left(\mu, \frac{\mu((0,L)^d)}{L^d}\right)\right]. \tag{4.2}$$

The proof is an application of the deterministic estimate (3.46) from Theorem 3.7 combined with (3.53) and (3.57) from Corollary 3.10 on a suitable event. As a stochastic ingredient, we will use the crucial fact that by convergence of the right-hand side term in (4.2), the first line on the right-hand side of (3.57) is small and thus $\mathrm{Var}_{\Omega^{(3\ell)}}\left(\mathsf{Q}_{\mathsf{b}^2_\Omega}(v)\right)$ is also small. We will also need the following two lemmas to bound from above the third right-hand side term in (3.57) and the second right-hand side term in (3.46). The first lemma reads as follows.

Lemma 4.4. *Let Assumption 4.1 hold with $d > 2 + \alpha$. For every $L_0 \geq 1$ and dyadic $m = 2^h \geq 2$, letting $L := mL_0$, it holds*

$$\frac{1}{L^d}\mathbb{E}\left[\mathsf{W}^p_{(0,L)^d}\left(\frac{\mu((0,L)^d)}{L^d}, \sum_{k=1}^{m^d} \frac{\mu(Q^k_{L_0})}{L_0^k}\chi_{Q^k_{L_0}}\right)\right] \lesssim L_0^{(2+\alpha-d)p/2}.$$

where $(0,L)^d = \bigcup_{k=1}^{m^d} Q^k_{L_0}$ is a partition into cubes of side length L_0.

Before we state the second lemma, we make an explicit choice for the cut-off η, following Remark 3.8. Let first

$$f(x) = 10x^3 - 15x^4 + 6x^5 = 1 + (x-1)^3(6x^2 + 3x + 1)$$

so that

$$f(0) = f'(0) = f''(0) = f'(1) = f''(1) = 0 \qquad \text{and} \qquad f(1) = 1. \tag{4.3}$$

For $\delta > 0$, we then consider $\xi_\delta : \mathbb{R} \to [0,1]$ even and defined on $[0,\infty)$ as

$$\xi_\delta(x) = \begin{cases} 1 & \text{for } x \in [0, 1/2 - 2\delta], \\ 1 - f\left(\frac{x-(\frac{1}{2}-2\delta)}{\delta}\right) & \text{for } x \in [1/2 - 2\delta, 1/2 - \delta], \\ 0 & \text{for } x \in [1/2 - \delta, \infty). \end{cases}$$

By (4.3) we see that $\xi_\delta \in C^2(\mathbb{R})$. Moreover, $\delta \left\|\xi'_\delta\right\|_\infty + \delta^2 \left\|\xi''_\delta\right\|_\infty \lesssim 1$. We then let, for $L > 0$,

$$\eta_{d,L,\delta}(x) = 1 - \prod_{i=1}^{d} \xi_\delta(x_i/L - 1/2), \tag{4.4}$$

so that $\left\|\nabla \eta_{d,L,\delta}\right\|_\infty \lesssim (\delta L)^{-1}$, $\left\|\nabla^2 \eta_{d,L,\delta}\right\|_\infty \lesssim (\delta L)^{-2}$, where the implicit constant depends on d only. Notice for later use that, writing $\Gamma = \left\{x \in (0,L)^d \,:\, \eta_{d,L,\delta}(x) = 0\right\}$, then Γ is a cube of side length $(1-2\delta)L$ and it holds for every $x \in (0,L)^d$,

$$\left(\frac{\mathsf{d}(x,\Gamma)}{\delta L}\right)^3 \lesssim \eta_{d,L,\delta}(x) \lesssim \left(\frac{\mathsf{d}(x,\Gamma)}{\delta L}\right)^3. \tag{4.5}$$

Lemma 4.5. *Let Assumption 4.1 hold with $d > 2+\alpha$. For every dyadic $\delta = 2^{-h_0} > 0$, there exists $c(\delta) < \infty$ such that, for every $L \geq 1$,*

$$\frac{1}{L^d}\mathbb{E}\left[\mathsf{W}^p_{(0,L)^d}\left(\eta\mu, \frac{\int_{(0,L)^d}\eta d\mu}{\int_{(0,L)^d}\eta}\eta\right)\right] \lesssim \delta + c(\delta)L^{(2+\alpha-d)p/2}. \tag{4.6}$$

where $\eta := \eta_{d,\delta,L}$ as in (4.4).

We show first how to deduce Theorem 4.3 using Lemma 4.4 and Lemma 4.5.

Proof of Theorem 4.3. We split the proof into several steps.
Step 1 (Notation). We write for brevity

$$\mathsf{w}(L) := \frac{1}{L^d}\mathbb{E}\left[\mathsf{W}^2_{Q_L}\left(\mu, \frac{\mu(Q_L)}{L^d}\right)\right], \quad \mathsf{wb}(L) := \frac{1}{L^d}\mathbb{E}\left[\mathsf{Wb}^2_{Q_L}\left(\mu, \frac{\mu(Q_L)}{L^d}\right)\right],$$

and let $\mathsf{w}(\infty) = \lim_{L\to\infty}\mathsf{w}(L)$, $\mathsf{wb}(\infty) = \lim_{L\to\infty}\mathsf{w}(L)$. Since $\mathsf{w}(L) \geq \mathsf{wb}(L)$ for every L, it is sufficient to establish the inequality $\mathsf{w}(\infty) \leq \mathsf{wb}(\infty)$. We introduce the quantity

$$\omega(L) := \sup_{L'\geq L}\left|\mathsf{wb}(\infty) - \mathsf{wb}(L')\right| \overset{L\to\infty}{\to} 0.$$

Without loss of generality we may assume that $L = 2^{\bar{h}}$ is dyadic. We write λ for the constant density $\lambda = \mu(Q_L)/L^d\chi_{Q_L}$ on $Q_L := (0,L)^d$. Let $\gamma \in (0,1)$ to be chosen below and set $h := \gamma\bar{h}$, $L_0 := 2^h = L^\gamma$ and $m := L/L_0 = 2^{\bar{h}-h} = 2^{(1-\gamma)\bar{h}} \in \mathbb{N}$. Notice that $L_0 \to \infty$ as $L \to \infty$. With this notation, Q_L is partitioned up to Lebesgue negligible (hence $\mathbb{P}$-a.s. μ negligible sets) into m^d disjoint sub-cubes $(Q^k_{L_0})_{k=1,\dots,m^d}$, each of side length L_0.

Furthermore, we introduce auxiliary parameters $\delta = 2^{-h_0}$ for $h_0 \in \mathbb{N}$, $\varepsilon \in (0,1)$ and let $\eta := \eta_{d,L,\delta}$ be as in (4.4), so that assumption *(1)* in Theorem 3.7 holds on $\Omega = Q_L$ with $r := \delta L$ and $c = c(d) < \infty$. We will let eventually $\delta \to 0$ and $\varepsilon \to 0$, but only after sending $L \to \infty$.
Step 2 (Reduction to a good event). We introduce the event

$$A := \left\{\mathsf{Wb}^2_{Q_L}(\mu,\lambda) \leq L^{d+\varepsilon}\right\} \cap \left\{\left|\mu(Q_L) - L^d\right| \leq L^{(1+\varepsilon)(d+\alpha)/2}\right\} \cap \left\{\mu(Q^k_{L_0}) > 0,\ \forall k = 1,\dots,m^d\right\}.$$

To estimate $\mathbb{P}(A^c)$ we use a union bound. First, by (4.1) we get for any $q > 0$

$$\mathbb{P}\left(\left|\mu(Q_L) - L^d\right| \geq L^{(1+\varepsilon)(d+\alpha)/2}\right) \lesssim_q L^{-q\varepsilon(d+\alpha)/2} \tag{4.7}$$

and

$$\sum_{k=1}^{m^d}\mathbb{P}\left(\mu(Q^k_{L_0}) = 0\right) \lesssim_q m^d L_0^{-q(d-\alpha)/2}. \tag{4.8}$$

For L large enough, in the event $\left\{\left|\mu(Q_L) - L^d\right| \leq L^{(1+\varepsilon)(d+\alpha)/2}\right\}$ (and in particular on A), we have

$$\frac{1}{2} \leq \frac{\mu(Q_L)}{L^d} \leq 2,$$

so that by Markov inequality and (3.1), we have for every $q' \geq 2$,

$$\begin{aligned}\mathbb{P}\left(\left\{\mathsf{Wb}^2_{Q_L}(\mu,\lambda) \geq L^{d+\varepsilon}\right\} \cap \left\{\left|\mu(Q_L) - L^d\right| \leq L^{(1+\varepsilon)(d+\alpha)/2}\right\}\right) &\lesssim_{q'} \mathbb{E}\left[\mathsf{Wb}^{q'}_{Q_L}(\mu,\lambda)\right] L^{-(d+\varepsilon)q'/2} \\ &\lesssim_{q'} L^{d-(d+\varepsilon)q'/2}.\end{aligned}$$

Combining the above with (4.7) and (4.8), this gives

$$P(A^c) \lesssim_{q,q'} L^{d-(d+\varepsilon)q'/2} + L^{-q\varepsilon(d+\alpha)/2} + m^d L_0^{-q(d-\alpha)/2}, \tag{4.9}$$

where the last term is then equal to

$$m^d L_0^{-q(d-\alpha)/2} = 2^{\bar{h}d - h(d+q(d-\alpha)/2)} = 2^{\bar{h}(d-\gamma(d+q(d-\alpha)/2))}.$$

On A^c, we trivially bound from above

$$\mathsf{W}^2_{Q_L}(\mu,\lambda) \lesssim_d \mu(Q_L)L^2$$

hence we have, by Cauchy-Schwarz inequality,

$$\begin{aligned}\frac{1}{L^d}\mathbb{E}\left[\chi_{A^c}\mathsf{W}^2_{Q_L}(\mu,\lambda)\right] &\lesssim L^{2-d}\mathbb{E}\left[\mu(Q_L)^2\right]^{1/2} P(A^c)^{1/2} \\ &\lesssim L^2\left(L^{d/2-(d+\varepsilon)q'/4} + L^{-q\varepsilon(d+\alpha)/4} + m^{d/2}L_0^{-q(d-\alpha)/4}\right) =: \mathsf{err}_\gamma(L).\end{aligned} \tag{4.10}$$

Notice that

$$L^2 m^{d/2} L_0^{-q(d-\alpha)/4} = 2^{2\bar{h}} 2^{\bar{h}\frac{d}{2} - h(d+q(d-\alpha)/4)} = 2^{\bar{h}\left[\left(2+\frac{d}{2}\right)-\gamma\left(d+q\frac{d-\alpha}{4}\right)\right]}.$$

We now choose for every $\gamma, \varepsilon \in (0,1)$ fixed, q' and q sufficiently large, such that

$$2 + d/2 - (d+\varepsilon)q'/4 < 0, \quad 2 - q\varepsilon(d+\alpha)/4 < 0, \quad \text{and} \quad \left(2+\frac{d}{2}\right) - \gamma\left(d + q\frac{d-\alpha}{4}\right) < 0.$$

With this choice we get $\lim_{L\to\infty} \mathsf{err}_\gamma(L) = 0$ and thus

$$\limsup_{L\to\infty} \frac{1}{L^d}\mathbb{E}\left[\chi_{A^c}\mathsf{W}^2_{Q_L}(\mu,\lambda)\right] = 0.$$

The thesis is reduced to

$$\lim_{L\to\infty} \frac{1}{L^d}\mathbb{E}\left[\chi_A \mathsf{W}^2_{Q_L}(\mu,\lambda)\right] \le \mathsf{wb}(\infty).$$

Let us notice for later use that, arguing similarly on each $Q^k_{L_0}$ and using (4.1) we get

$$\begin{aligned}\frac{1}{L_0^d}\mathbb{E}\left[\chi_{A^c}\mathsf{Wb}^2_{Q^k_{L_0}}\left(\mu, \frac{\mu(Q^k_{L_0})}{|L_0^d|}\right)\right] &\lesssim L_0^2 L_0^{-d}\mathbb{E}\left[\mu(Q^k_{L_0})^2\right]^{1/2} P(A^c)^{1/2} \\ &\overset{(4.9)}{\lesssim} L_0^2\left(L^{d/2-(d+\varepsilon)q'/4} + L^{-q\varepsilon(d+\alpha)/4} + m^{d/2}L_0^{-q(d-\alpha)/4}\right) \\ &\overset{L_0\le L}{\le} \mathsf{err}_\gamma(L).\end{aligned}$$

Hence,

$$\frac{1}{L^d}\sum_{k=1}^{m^d}\mathbb{E}\left[\chi_{A^c}\mathsf{Wb}^2_{Q^k_{L_0}}\left(\mu, \frac{\mu(Q^k_{L_0})}{|L_0^d|}\right)\right] = \frac{1}{m^d}\sum_{k=1}^{m^d}\frac{1}{L_0^d}\mathbb{E}\left[\chi_{A^c}\mathsf{Wb}^2_{Q^k_{L_0}}\left(\mu, \frac{\mu(Q^k_{L_0})}{|L_0^d|}\right)\right] \lesssim \mathsf{err}_\gamma(L).$$

Up to multiplying $\mathsf{err}_\gamma(L)$ by a small (universal) constant, we thus find

$$\frac{1}{L^d}\sum_{k=1}^{m^d}\mathbb{E}\left[\chi_A \mathsf{Wb}^2_{Q^k_{L_0}}\left(\mu, \frac{\mu(Q^k_{L_0})}{L_0^k}\right)\right] \ge \mathsf{wb}(L_0) - \mathsf{err}_\gamma(L). \tag{4.11}$$

Step 3 (Variance bound). On the event A, we apply Corollary 3.10 with $\mathcal{Y} = \mathrm{supp}(\mu) \cap \overline{\Omega}$ and obtain $v \in \mathcal{S}(\mathsf{b}^2_{Q_L}, \mathcal{Y})$ optimizer for the dual formulation of $\mathsf{Wb}^2_{Q_L}(\mu,\lambda)$. Writing $u := \mathsf{Q}_{\mathsf{b}^2_{Q_L}}(v)$, we have by (3.41) with $p = 2$ (hence $p' = 2$) and $\Omega = Q_L$, that

$$\mathrm{osc}_{Q_L}(u) \lesssim_d L^{1+\frac{d+\varepsilon}{d+2}},$$

and since $\mu(Q^k_{L_0}) > 0$ for every k,

$$\mathsf{d}(Q_L, \mathcal{Y}) \lesssim_d L_0.$$

By (3.11) we have

$$\ell = (\mathrm{osc}_{\mathcal{Y}}(v) + \mathsf{d}(Q_L, \mathcal{Y})^2)^{1/2} \le (\mathrm{osc}_{Q_L}(u) + 2\mathsf{d}(Q_L, \mathcal{Y})^2)^{1/2} \lesssim \max\left\{L^{(1+\frac{d+\varepsilon}{d+2})/2}, L_0\right\}.$$

Therefore, since $L_0 \ll L$ (which is guaranteed by our choice $L_0 = L^\gamma$ for $\gamma \in (0,1)$) we have, for every given $\varepsilon, \delta \in (0,1)$, that

$$r = \delta L \gg \ell. \tag{4.12}$$

By (3.57) and (2.2) on $\Omega = Q = \bigcup_k Q^k_{L_0}$, we obtain that for every $\varepsilon' \in (0,1)$,

$$\begin{aligned}\mathrm{Var}_{Q_L^{(r)}}(u) \lesssim L^2\Bigg[&\mathsf{Wb}^2_{Q_L}(\mu,\lambda) - (1+\varepsilon')^{-1}\sum_k \mathsf{Wb}^2_{Q^k_{L_0}}\left(\mu, \frac{\mu(Q^k_{L_0})}{L_0^k}\right)\\ &+ \frac{1}{\varepsilon'}\mathsf{W}^2_{Q_L}\left(\lambda, \sum_k \frac{\mu(Q^k_{L_0})}{L_0^k}\chi_{Q^k_{L_0}}\right)\Bigg] + L^d L_0^4.\end{aligned}$$

Taking expectation on A, we find

$$\begin{aligned}\frac{1}{L^d}\mathbb{E}\left[\chi_A \mathrm{Var}_{Q_L^{(r)}}(u)\right] &\lesssim L^2\left(\mathsf{wb}(L) - (1+\varepsilon')^{-1}(\mathsf{wb}(L_0) - \mathsf{err}_\gamma(L)) + \frac{1}{\varepsilon'}L_0^{2+\alpha-d}\right) + L_0^4\\ &\lesssim L^2\left(\omega(L_0) + \varepsilon' + \mathsf{err}_\gamma(L) + \frac{1}{\varepsilon'}L_0^{2+\alpha-d}\right) + L_0^4\end{aligned}$$

having used (4.11) and Lemma 4.4 with $p = 2$. The choice $\varepsilon' = L_0^{(2+\alpha-d)/2}$ finally yields

$$\frac{1}{L^d}\mathbb{E}\left[\chi_A \mathrm{Var}_{Q_L^{(r)}}(u)\right] \lesssim L^2\left(\omega(L_0) + \mathsf{err}_\gamma(L) + L_0^{(2+\alpha-d)/2}\right) + L_0^4. \tag{4.13}$$

Step 4 (Application of Theorem 3.7). In the event A, thanks to (4.12) we may apply Theorem 3.7 together with Lemma 2.2 on $\Omega = Q_L$ to obtain that for every $\varepsilon'' \in (0,1)$,

$$\begin{aligned}\mathsf{W}^2_{Q_L}(\mu,\lambda) - \mathsf{Wb}^2_{Q_L}(\mu,\lambda) \lesssim \varepsilon''\mathsf{Wb}^2_{Q_L}(\mu,\lambda) + \frac{1}{\varepsilon''}\Bigg[&\mathsf{W}^2_{Q_L}\left(\frac{\int_{Q_L}\eta d\mu}{\int_{Q_L}\eta}\eta, \eta\mu\right)\\ &+ L^2\left(\left|\frac{\int_{Q_L}\eta d\mu}{\int_{Q_L}\eta} - 1\right|^2 \int_{Q_L}\eta^2 + r^{-4}\left(\mathsf{Wb}^2_{Q_L}(\mu,\lambda)\right)^{\frac{d+4}{d+2}}\right)\\ &+ \left(r^{-2} + L^2 r^{-4}\right)\mathrm{Var}_{Q_L^{(r)}}(u)\Bigg].\end{aligned} \tag{4.14}$$

After taking expectation (on A) and dividing both sides by L^d, we estimate all the terms in the right-hand side, starting from the third line, which is bounded using (4.13):

$$\frac{1}{L^d}\left(r^{-2} + L^2 r^{-4}\right)\mathbb{E}\left[\chi_A \mathrm{Var}_{Q_L^{(r)}}(u)\right] \lesssim \delta^{-4}\left(\omega(L_0) + \mathsf{err}_\gamma(L) + L_0^{(2+\alpha-d)/2} + L^{-2}L_0^4\right). \tag{4.15}$$

For fixed δ the right-hand side goes to zero provided

$$\lim_{L\to\infty} L^{-2}L_0^4 = 0 \qquad \text{or equivalently} \qquad \gamma \in (0, 1/2). \tag{4.16}$$

For the second line in (4.14), we have two contributions: the term

$$L^{2-d}r^{-4}\left(\mathsf{Wb}^2_{Q_L}(\mu,\lambda)\right)^{\frac{d+4}{d+2}} \lesssim \delta^{-4}L^{\frac{2d+\varepsilon(d+4)}{d+2}-2}, \tag{4.17}$$

which is infinitesimal as $L \to \infty$, for ε small enough, and

$$L^{2-d}\left|\frac{\int_{Q_L}\eta d\mu}{\int_{Q_L}\eta}-1\right|^2\int_{Q_L}\eta^2 = L^{2-d}\left|\int_{Q_L}\eta d\mu-\int_{Q_L}\eta\right|^2\frac{\int_{Q_L}\eta^2}{\left(\int_{Q_L}\eta\right)^2}$$
$$\lesssim \delta^{-1}L^{2-2d}\left|\int_{Q_L}\eta d\mu-\int_{Q_L}\eta\right|^2$$

where the inequality follows by the properties of η. We further bound in expectation, using (4.1):

$$\delta^{-1}L^{2-2d}\mathbb{E}\left[\chi_A\left|\int_{Q_L}\eta d\mu-\int_{Q_L}\eta\right|^2\right] \lesssim \delta^{-1}L^{2-2d}\operatorname{Var}_{\mathbb{P}}\left(\int_{Q_L}\eta d\mu\right) \lesssim \delta^{-1}L^{2+\alpha-d} \tag{4.18}$$

that is infinitesimal as $L \to \infty$, by the assumption $d > 2+\alpha$. For the terms in the first line of the right-hand side of (4.14), we have first that

$$\frac{\varepsilon''}{L^d}\mathbb{E}\left[\chi_A \mathsf{Wb}^2_{Q_L}(\mu,\lambda)\right] \lesssim \varepsilon''\mathsf{wb}(L) \lesssim \varepsilon'', \tag{4.19}$$

hence we are left with only one contribution, i.e., the second term in the first line, that we bound in expectation using Lemma 4.5 with $p = 2$:

$$\frac{1}{L^d}\mathbb{E}\left[\mathsf{W}^2_{Q_L}\left(\frac{\int_{Q_L}\eta d\mu}{\int_{Q_L}\eta}\eta,\eta\mu\right)\right] \lesssim \delta + c(\delta)L^{2+\alpha-d}. \tag{4.20}$$

Plugging (4.15),(4.17),(4.18),(4.19) and (4.20) in (4.14) yields in combination with (4.10),

$$\frac{1}{L^d}\mathbb{E}\left[\mathsf{W}^2_{Q_L}(\mu,\lambda)\right]-\frac{1}{L^d}\mathbb{E}\left[\mathsf{Wb}^2_{Q_L}(\mu,\lambda)\right] \lesssim \varepsilon''+\frac{1}{\varepsilon''}\left[\delta+c(\delta)\left(o_L(1)+o_{L_0}(1)+\mathsf{err}_\gamma(L)+L^{-2}L_0^4\right)\right]. \tag{4.21}$$

Step 5 (Conclusion). We now choose $\gamma \in (0,1/2)$ arbitrarily. In particular, (4.16) holds. Sending first $L \to \infty$ we get from (4.21),

$$\mathsf{w}(\infty)-\mathsf{wb}(\infty) = \lim_{L\to\infty}\left(\frac{1}{L^d}\mathbb{E}\left[\mathsf{W}^2_{Q_L}(\mu,\lambda)\right]-\frac{1}{L^d}\mathbb{E}\left[\mathsf{Wb}^2_{Q_L}(\mu,\lambda)\right]\right) \lesssim \varepsilon''+\frac{\delta}{\varepsilon''}.$$

Sending first $\delta \to 0$ and then $\varepsilon'' \to 0$ (or directly choosing $\varepsilon'' = (\delta)^{1/2}$) we obtain

$$\mathsf{w}(\infty)-\mathsf{wb}(\infty) \le 0$$

which was the original claim. □

4.1 Proof of Lemma 4.4

We write for brevity $\kappa(\Omega) := \mu(\Omega)/|\Omega|$. Writing $Q_L := (0,L)^d$, we first reduce ourselves to the event

$$A = \left\{\kappa(Q_L) \ge \frac{1}{2}\right\}.$$

Indeed, by Markov inequality and (4.1), we easily find, for every $q \ge 1$,

$$\mathbb{P}(A^c) \lesssim_q L^{-q},$$

hence, by trivially estimating on A^c,

$$\mathsf{W}^p_{Q_L}\left(\kappa(Q_L), \sum_{k=1}^{m^d} \kappa(Q^k_{L_0})\chi_{Q^k_{L_0}}\right) \lesssim L^p \mu(Q_L),$$

we find, by Cauchy-Schwarz inequality,

$$\frac{1}{L^d}\mathbb{E}\left[\chi_{A^c}\mathsf{W}^p_{Q_L}\left(\kappa(Q_L), \sum_{k=1}^{m^d} \kappa(Q^k_{L_0})\right)\right] \lesssim_q L^{p-q/2} \lesssim L^{2-d+\alpha}$$

provided that we choose q large enough.

On the event A, we use instead Lemma 3.6 with uniform weight $\eta = \chi_{Q_L}$, so that $g_0 = \kappa(Q_L) \geq 1/2$, and we obtain

$$\frac{1}{L^d}\mathbb{E}\left[\chi_A \mathsf{W}^p_{Q_L}\left(\kappa(Q_L), \sum_{k=1}^{m^d} \kappa(Q^k_{L_0})\chi_{Q^k_{L_0}}\right)\right] \lesssim \frac{1}{L^d}\mathbb{E}\left[\left\|\kappa(Q_L) - \sum_{k=1}^{m^d} \kappa(Q^k_{L_0})\chi_{Q^k_{L_0}}\right\|^p_{H^{-1,p}(Q_L)}\right].$$

The thesis follows therefore by arguing that

$$\frac{1}{L^d}\mathbb{E}\left[\left\|\kappa(Q_L) - \sum_{k=1}^{m^d} \kappa(Q^k_{L_0})\chi_{Q^k_{L_0}}\right\|^p_{H^{-1,p}(Q_L)}\right] \lesssim L_0^{(2+\alpha-d)p/2}. \tag{4.22}$$

To this aim, we remark that estimating the $H^{-1,p}$ norm with the L^p norm as in (2.11) would result in a cruder bound, depending on m, hence we need to iterate over dyadic decompositions. For every cube Q of side length $\ell \geq 2$, partitioned into 2^d subcubes $(Q^i)_{i=1,\dots,2^d}$ each with side length $\ell/2 \geq 1$, we notice that it holds

$$\mathbb{E}\left[\left\|\kappa(Q) - \sum_{i=1}^{2^d} \kappa(Q^i)\chi_{Q^i}\right\|^p_{H^{-1,p}(Q)}\right] \lesssim \ell^{2+\alpha}.$$

Indeed, it is sufficient to apply (2.11) with $c_P(Q,p) \lesssim \ell$ to obtain

$$\begin{aligned}\mathbb{E}\left[\left\|\kappa(Q) - \sum_{i=1}^{2^d} \kappa(Q^i)\chi_{Q^i}\right\|^p_{H^{-1,p}(Q)}\right] &\lesssim \ell^p \mathbb{E}\left[\int_Q \sum_{i=1}^{2^d} \left|\kappa(Q) - \kappa(Q^i)\right|^p \chi_{Q^i}\right] \\ &\lesssim \ell^p\left(\ell^d \mathbb{E}\left[|\kappa(Q) - 1|^p\right] + \sum_{i=1}^{2^d}\left(\frac{\ell}{2}\right)^d \mathbb{E}\left[\left|\kappa(Q^i) - 1\right|^p\right]\right) \\ &\overset{(4.1)}{\lesssim} \ell^{d+p\left(1-\frac{d-\alpha}{2}\right)}.\end{aligned}$$

Starting from $Q = Q_L$, we use the sub-additivity property (2.15) of $H^{-1,p}$ (with uniform weight) by further iterating h times the decomposition into cubes of halved side lengths, so that at the i-th iteration the side lengths of the cubes are $\ell_i := L/2^i = L_0 2^{h-i}$ and moreover we can specify a suitable $\varepsilon = \varepsilon_i \in (0,1)$. At the h-th step, we obtain exactly the partition into the m^d sub-cubes $Q^k_{L_0}$,

and we deduce that

$$\begin{aligned}\frac{1}{L^d}\mathbb{E}\left[\left\|\kappa(Q_L)-\sum_{k=1}^{m^d}\kappa(Q_{L_0}^k)\chi_{Q_{L_0}^k}\right\|_{H^{-1,p}(Q_L)}^p\right] &\lesssim L^{-d}\sum_{i=1}^{h}2^{id}\left(L2^{-i}\right)^{d+p\left(1-\frac{d-\alpha}{2}\right)}\varepsilon_i^{-(p-1)}\prod_{j=1}^{i-1}(1+\varepsilon_j)\\ &=\sum_{i=1}^{h}\left(L2^{-i}\right)^{p\left(1-\frac{d-\alpha}{2}\right)}\varepsilon_i^{-(p-1)}\prod_{j=1}^{i-1}(1+\varepsilon_j)\\ &=L_0^{p\left(1-\frac{d-\alpha}{2}\right)}\sum_{i=1}^{h}\left(2^{h-i}\right)^{p\left(1-\frac{d-\alpha}{2}\right)}\varepsilon_i^{-(p-1)}\prod_{j=1}^{i-1}(1+\varepsilon_j).\end{aligned} \tag{4.23}$$

We specify $\varepsilon_i = (2^{h-i})^{-(d-2-\alpha)/2}$, so that the product terms are uniformly bounded from above:

$$\begin{aligned}\prod_{j=1}^{i-1}(1+\varepsilon_j) &\le \prod_{j=1}^{h}\left(1+(2^{h-j})^{-\frac{d-2-\alpha}{2}}\right)\le\exp\left(\sum_{j=1}^{h}2^{-\frac{h-j}{2}(d-2-\alpha)}\right)\\ &\lesssim 1.\end{aligned}$$

Therefore

$$\sum_{i=1}^{h}\left(2^{h-i}\right)^{p\left(1-\frac{d-\alpha}{2}\right)}\varepsilon_i^{-(p-1)}\prod_{j=1}^{i-1}(1+\varepsilon_j)\lesssim\sum_{i=1}^{h}\varepsilon_i\lesssim 1.$$

Combining this with (4.23) this concludes the proof of (4.22).

4.2 Proof of Lemma 4.5

In this case, we fix $m = 2^h$ such that $L_0 := L/m \in [1, 2)$. Moreover, we write for brevity

$$\kappa(\Omega) := \frac{\int_\Omega \eta d\mu}{\int_\Omega \eta},$$

with the convention that $\kappa(\Omega) = 0$ if $\int_\Omega \eta = 0$.

We preliminarily reduce ourselves to the event

$$A := \left\{\kappa(Q_L) \ge \frac{1}{4}\right\}.$$

By the properties of η, in particular the fact that η is identically 1 if $\mathsf{d}(x, Q_L^c) \le \delta L$ and 0 if $\mathsf{d}(x, Q_L^c) \ge 2\delta L$, we see that

$$\left\{\mu(\mathsf{d}(\cdot, Q_L^c) \le \delta L) \ge \delta L^d/2\right\} \subseteq A,$$

hence by Markov inequality and (4.1) we find, for every $q \ge 1$,

$$\mathbb{P}(A^c) \lesssim_{\delta,q} L^{-q}.$$

On the event A, we partition Q_L into m^d disjoint subcubes $Q_L = \bigcup_{k=1}^{m^d} Q_{L_0}^k$, each of side length L_0, and use (3.15) with $\eta\mu$ instead of μ, $\eta \int_{Q_L} \eta\mu / \int_{Q_L} \eta$ instead of λ, and $\varepsilon = 1/2$.We find

$$\mathsf{W}_{Q_L}^p(\eta\mu, \kappa(Q_L)\eta) \lesssim \sum_k^{m^d} \mathsf{W}_{Q_{L_0}^k}^p\left(\eta\mu, \kappa(Q_{L_0}^k)\eta\right) + \mathsf{W}_{Q_L}^p\left(\sum_{k=1}^{m^d}\kappa(Q_{L_0}^k)\chi_{Q_{L_0}^k}\eta, \kappa(Q_L)\eta\right). \tag{4.24}$$

Taking expectation (on A) and using that $L_0 \le 2$, we find

$$\mathbb{E}\left[\mathsf{W}^p_{Q^k_{L_0}}\left(\eta\mu, \kappa(Q^k_{L_0})\eta\right)\right] \lesssim \mathbb{E}\left[\int_{Q^k_{L_0}} \eta\mu\right] = \int_{Q^k_{L_0}} \eta$$

so that the first contribution in the right-hand side of (4.24) can be bounded as

$$\mathbb{E}\left[\sum_k^{m^d} \mathsf{W}^p_{Q^k_{L_0}}\left(\eta\mu, \kappa(Q^k_{L_0})\eta\right)\right] \lesssim \int_{Q_L} \eta \lesssim L^d\delta,$$

by the properties of η. After dividing by L^d, this yields a first contribution in (4.6). For the second term, we apply Lemma 3.6, so that

$$\mathbb{E}\left[\chi_A \mathsf{W}^p_{Q_L}\left(\sum_{k=1}^{m^d} \kappa(Q^k_{L_0})\chi_{Q^k_{L_0}}\eta, \kappa(Q_L)\eta\right)\right] \lesssim \mathbb{E}\left[\left\|\sum_{k=1}^{m^d} \kappa(Q^k_{L_0})\chi_{Q^k_{L_0}} - \kappa(Q_L)\right\|^p_{H^{-1,p}(\eta)}\right].$$

Let us notice that, as we argue on A, we can use the lower bound on $\kappa(Q_L)$ yielding $g_0 = 1/4$ in Lemma 3.6, and moreover $c_P(\eta, p) < \infty$ by Remark 2.1.

To complete the proof, we argue that

$$\frac{1}{L^d}\mathbb{E}\left[\left\|\sum_{k=1}^{m^d} \kappa(Q^k_{L_0})\chi_{Q^k_{L_0}} - \kappa(Q_L)\right\|^p_{H^{-1,p}(\eta)}\right] \lesssim \delta + c(\delta)L^{(2-d+\alpha)p/2}. \tag{4.25}$$

Let us notice that this can be regarded as a weighted analogue of (4.22), and indeed, as in that case, we iterate over dyadic decompositions. First, we notice that for any given cube Q of side length $\ell \ge 2$, partitioned into 2^d subcubes $(Q^i)_{i=1,\dots,2^d}$ each with side length $\ell/2 \ge 1$, it holds

$$\mathbb{E}\left[\left\|\kappa(Q) - \sum_{i=1}^{2^d} \kappa(Q^i)\chi_{Q^i}\right\|^p_{H^{-1,p}(\chi_Q\eta)}\right] \lesssim c_P(\chi_Q\eta, p)^p \ell^d \ell^{-\frac{p}{2}(d-\alpha)}.$$

Indeed, it is sufficient to apply (2.11) and obtain

$$\begin{aligned}\mathbb{E}\left[\left\|\kappa(Q) - \sum_{i=1}^{2^d} \kappa(Q^i)\chi_{Q^i}\right\|^p_{H^{-1,p}(Q)}\right] &\lesssim c_P(\chi_Q\eta, p)^p \mathbb{E}\left[\int_Q \sum_{i=1}^{2^d} \left|\kappa(Q) - \kappa(Q^i)\right|^p \chi_{Q^i}\eta\right] \\ &\lesssim c_P(\chi_Q\eta, p)^p \left(\mathbb{E}\left[|\kappa(Q) - 1|^p\right]\int_Q \eta + \sum_{i=1}^{2^d} \mathbb{E}\left[\left|\kappa(Q^i) - 1\right|^p\right]\int_{Q^i}\eta\right) \\ &\overset{(4.1)}{\lesssim} c_P(\chi_Q\eta, p)^p \ell^d \ell^{-\frac{p}{2}(d-\alpha)}.\end{aligned}$$

The crux of the argument is to obtain suitable bounds for the possibly different constants $c_P(\chi_Q\eta, p)$ that appear in the subsequent dyadic decompositions of Q_L. We notice that, if η is identically null on Q, we can trivially estimate

$$\mathbb{E}\left[\left\|\kappa(Q) - \sum_{i=1}^{2^d} \kappa(Q^i)\chi_{Q^i}\right\|^p_{H^{-1,p}(Q)}\right] = 0,$$

hence it is sufficient to consider only the contributions of the cubes Q where η is not identically null. For every $i = 1, \dots, h$, let us denote with $\mathcal{Q}_i$ the collection of cubes of side length $L2^{-i+1}$ that

are obtained by iterated dyadic decomposition of Q_L and which intersect the support of η. Clearly, we have $\sharp \mathcal{Q}_i \le 2^{id}$ and moreover we can always bound from above (using that η is defined as in (4.4))

$$\sup_{Q \in \mathcal{Q}_i} \frac{c_P(\chi_Q \eta, p)}{|Q|^{1/d}} \lesssim_\delta 1.$$

Let us recall that $\delta = 2^{-h_0}$. For $i \ge h_0$, we have $\sharp \mathcal{Q}_i \le \delta 2^{id}$ and we claim that

$$\sup_{Q \in \mathcal{Q}_i} \frac{c_P(\chi_Q \eta, p)}{|Q|^{1/d}} \lesssim 1, \tag{4.26}$$

i.e., the (rescaled) Poincaré constant becomes uniformly controlled with respect to δ. Let us postpone the proof of (4.26) and conclude the proof of (4.25). Using (4.26) and arguing as in (4.23), we collect the inequality

$$\mathbb{E}\left[\left\| \kappa(Q_L) - \sum_{k=1}^{m^d} \kappa(Q_{L_0}^k) \chi_{Q_{L_0}^k} \right\|_{H^{-1,p}(Q_L)}^p\right] \le c(\delta) \sum_{i=1}^{h_0-1} \left(L2^{-i}\right)^{p\left(1-\frac{d-\alpha}{2}\right)} \varepsilon_i^{-(p-1)} \prod_{j=1}^{i-1}(1+\varepsilon_j) + \delta \sum_{i=h_0}^{h} \left(L2^{-i}\right)^{p\left(1-\frac{d-\alpha}{2}\right)} \varepsilon_i^{-(p-1)} \prod_{j=1}^{i-1}(1+\varepsilon_j).$$

For $i = 1, \cdots, h_0 - 1$, we choose $\varepsilon_i = 1/2$ so that (recall that h_0 depends only on δ)

$$\sum_{i=1}^{h_0-1} \left(L2^{-i}\right)^{p\left(1-\frac{d-\alpha}{2}\right)} \varepsilon_i^{-(p-1)} \prod_{j=1}^{i-1}(1+\varepsilon_j) \lesssim_\delta L^{(2+\alpha-d)p/2}.$$

For $i = h_0, \cdots, h$, we choose instead $\varepsilon_i = (L/2^i)^{-(d-2-\alpha)/2}$, so that arguing as above we find

$$\sum_{i=h_0}^{h} \left(L2^{-i}\right)^{p\left(1-\frac{d-\alpha}{2}\right)} \varepsilon_i^{-(p-1)} \prod_{j=1}^{i-1}(1+\varepsilon_j) \lesssim 1.$$

This would prove (4.25) and we are only left with the proof of (4.26). For this we write $\Gamma := \{x \in Q_L : \eta(x) = 0\}$, that is a cube of side length $(1-2\delta)L$. Given a cube $Q \in \mathcal{Q}_i$ with side length $\ell = 2^{-i}L \le \delta L$, we distinguish between two cases. If $\partial Q \cap \Gamma = \emptyset$, then by construction (in particular using that $\delta = 2^{-h_0}$ is also dyadic) we have $d(Q, \Gamma) \gtrsim \ell$. In particular, by (4.5), for every $x, y \in Q$, we have

$$\frac{\eta(x)}{\eta(y)} \sim \left(\frac{d(x,\Gamma)}{d(y,\Gamma)}\right)^3 \lesssim \left(\frac{\ell + d(Q,\Gamma)}{d(Q,\Gamma)}\right)^3 \lesssim 1.$$

By (2.7), up to multiplication by a constant, translation and scaling, (4.26) then follows from the standard Poincaré inequality in $(0,1)^d$. If otherwise $\partial Q \cap \Gamma \neq \emptyset$, set for $k \in \{1, \cdots, d\}$, $\Gamma_k = \{0\}^k \times (0,1)^{d-k}$ and then for $x \in (0,1)^d$, $\eta_k(x) = d^3(x, \Gamma_k)$. By (2.7) and (4.5), we see that after suitable translation, scaling and multiplication by a constant, it is enough to prove that, for every $k \in \{1, \ldots, d\}$,

$$c_P(\eta_k \chi_{(0,1)^d}, p) \lesssim 1. \tag{4.27}$$

But this follows at once from Remark 2.1.

Remark 4.6. The proofs of Lemma 4.4 and Lemma 4.5 have a similar structure, and could be combined together, but we prefer to keep them separate to avoid adding technicalities. Moreover, in the range $d > 4 + \alpha$ one could rely upon easier bounds, estimating the Wasserstein distance in terms of the total variation (3.3) and arguing as in [9]. However this would leave out from the application to the transport of i.i.d. points in dimension $d \in \{3, 4\}$, and that of the Brownian interlacement occupation measure for $d \in \{5, 6\}$.

4.3 Bounds for the Brownian interlacement occupation measure

The aim of this section is to provide a proof of the validity of Assumption 4.1, with $\alpha = 2$, for the Brownian interlacement occupation measure. As already noticed, the validity of Assumption 4.1 with $\alpha = 0$ for a Poisson point process is rather straightforward: however, since it may give some intuition for the analogue derivations in the case of the Brownian interlacement occupation measure, we report here the argument leading to (4.1). Indeed, given a Poisson point process μ, a cube Q and $\eta : Q \to [0, \infty)$ bounded, we write

$$\int_Q \eta d\mu = \sum_{i=1}^{N(Q)} \eta(X_i),$$

as a sum of a Poisson number $N(Q)$, with mean $|Q|$, of i.i.d. variables $\eta(X_i)$ where and each X_i is uniformly distributed on Q. Hence, $\mathbb{E}\left[\int_Q \eta d\mu\right] = \int_Q \eta$ and we can write

$$\int_Q \eta d\mu - \mathbb{E}\left[\int_Q \eta d\mu\right] = \sum_{i=1}^{N(Q)} \left(\eta(X_i) - \fint_Q \eta\right) + (N(Q) - |Q|) \fint_Q \eta.$$

The variables $Y_i := \eta(X_i) - \fint_Q \eta$ are centered and uniformly bounded by $\|\eta\|_\infty$. Thus, by Rosenthal's inequality, for any $q \geq 2$ and $n \geq 1$,

$$\left\| \sum_{i=1}^{n} Y_i \right\|_{L^q(\mathbb{P})} \lesssim_q \left(n^{1/q} + n^{1/2}\right) \|\eta\|_\infty \lesssim n^{1/2} \|\eta\|_\infty .$$

Conditioning upon $N(Q) = n$ and using that for a Poisson variable $\mathbb{E}\left[N(Q)^{q/2}\right] \lesssim_q |Q|^{q/2}$, we then find

$$\left\| \sum_{i=1}^{N(Q)} Y_i \right\|_{L^q(\mathbb{P})} \lesssim_q |Q|^{1/2} \|\eta\|_\infty .$$

Finally, by the concentration properties of the Poisson law,

$$\left\| (N(Q) - |Q|) \fint_Q \eta \right\|_{L^q(\mathbb{P})} \lesssim \|(N(Q) - |Q|)\|_{L^q(\mathbb{P})} \|\eta\|_\infty \lesssim_q |Q|^{1/2} \|\eta\|_\infty$$

and (4.1) is settled.

Moving to the Brownian interlacement occupation measure case, we first collect some facts about classical potential theory and the interlacement process, see [38, 26, 49, 44]. We write $p_t(x, y)$ for the transition function of Brownian motion and

$$g(x, y) = \int_0^\infty p_t(x, y) dt = \frac{c_d}{|x - y|^{d-2}}$$

for the Brownian Green's kernel (where c_d is a constant that depends on the dimension). Given $x \in \mathbb{R}^d$, we write $\mathbb{P}_x, \mathbb{E}_x$ for the law and expectation of Brownian motion started in $x \in \mathbb{R}^d$, similarly $\mathbb{P}_\lambda, \mathbb{E}_\lambda$ for any finite measure λ, not necessarily a probability.

For a compact $K \subset \mathbb{R}^d$ we denote by e_K its equilibrium measure, which is uniquely defined by requiring that the associated potential

$$V(e_K)(x) = \int_{\mathbb{R}^d} g(x, y) de_K(y)$$

satisfies $V(e_K) = 1$ on K (and $V(e_k)(x)$ is infinitesimal as $|x| \to \infty$).

The mass $e_K(\mathbb{R}^d) = \mathrm{Cap}(K)$ is called the capacity of K and we denote by $\tilde{e}_K = \frac{1}{\mathrm{Cap}(K)} e_K$ the normalised measure, assuming that $\mathrm{Cap}(K) \in (0, \infty)$. Note that the characterising property of the e_K implies that

$$\Delta^{-1} e_K \begin{cases} = 1 & \text{on } K \\ \leq 1 & \text{on } \mathbb{R}^d \end{cases}.$$

Given a continuous curve $\omega = (\omega_t)_{t \geq 0}$, write $\tau_K(\omega)$ for the first hitting time of K. For any $x \in \mathbb{R}^d$ we have the fundamental identity

$$\mathbb{P}_y(\tau_K < \infty) = \int_{\mathbb{R}^d} g(x, y) de_K(x) \tag{4.28}$$

relating Brownian hitting probabilities, Green's function and the equilibrium measure. Denoting with $\mu^B_\infty = \int_0^\infty \delta_{B_s} ds$ the occupation measure of a Brownian path, i.e., any Borel set $A \subset \mathbb{R}^d$,

$$\mu^B_\infty(A) = \int_0^\infty \chi_A(B_s)\, ds$$

identity (4.28) implies

$$\begin{aligned} \mathbb{E}_{e_K}\left[\mu^B_\infty(A)\right] &= \int_{\mathbb{R}^d} \int_0^\infty \int_{\mathbb{R}^d} \chi_A(y) p_t(x, y) dy dt de_K(x) \\ &= \int_{\mathbb{R}^d \times \mathbb{R}^d} g(x, y) \chi_A(y) dy de_K(x) \\ &= \int_A \mathbb{P}_y[\tau_K < \infty] dy \leq |A|. \end{aligned}$$

Hence, we deduce the inequality between measures

$$\mathbb{E}_{e_K}\left[\mu^B_\infty(A)\right] \leq |A|,$$

with equality if $A \subseteq K$. Moving from sets to functions, we have for any $f \geq 0$ that

$$\mathbb{E}_{e_K}\left[\int_0^\infty f(B_s) ds\right] \leq \int_{\mathbb{R}^d} f. \tag{4.29}$$

Next, we recall that given any compact $K \subseteq \mathbb{R}^d$, the restriction of the Brownian interlacement occupation measure μ on K can be represented as

$$\mu \llcorner K = \sum_{i=1}^{N_K} \mu^{B^i}_\infty \llcorner K,$$

where N_K is a Poisson random variable with mean $\mathbb{E}[N_K] = \mathrm{Cap}(K)$ and $(B^i)_{i=1}^\infty$ are independent Brownian motions, independent of N_K, each with initial law $\tilde{e}_K$. In particular, we get for $A \subset K$,

$$\mathbb{E}[\mu(A)] = \mathrm{Cap}(K) \frac{|A|}{\mathrm{Cap}(K)} = |A|$$

and similarly, for any function η supported in K, it holds

$$\mathbb{E}\left[\int_K \eta d\mu\right] = \int_K \eta.$$

We are now in a position to prove the concentration inequality (4.1).

Lemma 4.7. *Let $d \geq 3$ and μ be the Brownian interlacement occupation measure on $\mathbb{R}^d$. Then, for any cube Q with* $\mathrm{diam}(Q) \geq 1$ *and any non-negative and bounded η and any $q \geq 2$, it holds*

$$\left\| \int_Q \eta d\mu - \mathbb{E}\left[\int_Q \eta d\mu \right] \right\|_{L^q(\mathbb{P})} \leq c \, \mathrm{diam}(Q)^{\frac{d+2}{2}} \, \|\eta\|_\infty ,$$

where $c = c(q, d) < \infty$.

Proof. By considering $\eta \chi_Q$ we can assume η to be supported on Q. We write

$$\int \eta d\mu = \sum_{i=1}^{N} Y_i$$

for iid random variables $Y_i = \int \eta(B_s^i) ds$ and an independent Poisson random variable N with parameter $\mathrm{Cap}(Q) = \mathrm{diam}(Q)^{d-2} \, \mathrm{Cap}(Q_1) \sim \mathrm{diam}(Q)^{d-2}$. Since $0 \leq \int \eta(B_s) ds \leq \|\eta\|_\infty \int \chi_Q(B_s) ds$ it follows from Lemma 3.2 in [38, Lemma 3.2] that Y_1 has finite moments of all orders, in fact that

$$\mathbb{E}[Y_1^q] \lesssim_q \|\eta\|_\infty^q \, \mathrm{diam}(Q)^{2q}.$$

Applying Rosenthal's inequality as we did for the Poisson point process case – see also [38, eq. (2.57)] – it follows that $\int \eta d\mu$ has moments of all orders and that

$$\begin{aligned} \left\| \int_Q \eta d\mu - \mathbb{E}\left[\int_Q \eta d\mu \right] \right\|_{L^q(\mathbb{P})} &\lesssim_q \mathrm{diam}(Q)^{\frac{d-2}{q}} \, \|Y_1\|_{L^q(\mathbb{P})} + \mathrm{diam}(Q)^{\frac{d-2}{2}} \, \|Y_1\|_{L^2(\mathbb{P})} \\ &\lesssim_q \|\eta\|_\infty \, \mathrm{diam}(Q)^{\frac{d+2}{2}}. \end{aligned}$$

□

Acknowledgments and funding

Part of this work was conceived at the CIRM 2988 conference "PDE & Probability in interaction: functional inequalities, optimal transport and particle systems". The authors warmly thank the conference organizers and CIRM staff for setting up a stimulating environment. Among the participants, they thank J.-C. Mourrat and A. Delalande for fruitful comments on their respective research works. The authors thank L. Ambrosio and O. Mischler for useful comments and remarks on a preliminary version of the work.

M.G. and D.T. acknowledge the project G24-202 "Variational methods for geometric and optimal matching problems" funded by Università Italo Francese.

M.H. is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2044 -390685587, *Mathematics Münster: Dynamics–Geometry–Structure* and by the DFG through the SPP 2265: *Random Geometric Systems*.

D.T. acknowledges the MUR Excellence Department Project awarded to the Department of Mathematics, University of Pisa, CUP I57G22000700001, the HPC Italian National Centre for HPC, Big Data and Quantum Computing - CUP I53C22000690001, the PRIN 2022 Italian grant 2022WHZ5XH - "understanding the LEarning process of QUantum Neural networks (LeQun)", CUP J53D23003890006, the INdAM-GNAMPA project 2024 "Tecniche analitiche e probabilistiche in informazione quantistica". Research also partly funded by PNRR - M4C2 - Investimento 1.3, Partenariato Esteso PE00000013 - "FAIR - Future Artificial Intelligence Research" - Spoke 1 "Human-centered AI", funded by the European Commission under the NextGeneration EU programme.

References

[1] M. Ajtai, J. Komlós, and G. Tusnády. "On optimal matchings". In: *Combinatorica* 4 (1984), pp. 259–264 (cit. on p. 2).

[2] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Second edition. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2008 (cit. on p. 13).

[3] L. Ambrosio, F. Glaudo, and D. Trevisan. "On the optimal map in the 2-dimensional random matching problem". In: *Discrete Contin. Dyn. Syst.* 39.12 (2019), pp. 7291–7308 (cit. on p. 20).

[4] L. Ambrosio, M. Goldman, and D. Trevisan. "On the quadratic random matching problem in two-dimensional domains". In: *Electron. J. Probab.* 27 (2022), pp. 1–35 (cit. on pp. 3, 7, 31).

[5] L. Ambrosio, F. Stra, and D. Trevisan. "A PDE approach to a 2-dimensional matching problem". In: *Probability Theory and Related Fields* 173 (2019), pp. 433–477 (cit. on pp. 3, 23).

[6] S. Armstrong, T. Kuusi, and J.-C. Mourrat. *Quantitative stochastic homogenization and large-scale regularity*. Vol. 352. Springer, 2019 (cit. on p. 6).

[7] S. Armstrong and C. K. Smart. "Quantitative stochastic homogenization of convex integral functionals". In: *Annales scientifiques de l'Ecole normale supérieure* 49.2 (2016), pp. 423–481 (cit. on p. 6).

[8] F. Barthe and C. Bordenave. "Combinatorial optimization over two random point sets." In: *Séminaire de probabilités XLV*. Ed. by A. L. C. Donati-Martin and A. Rouault. Springer, 2013, pp. 483–535 (cit. on p. 31).

[9] F. Barthe and C. Bordenave. "Combinatorial optimization over two random point sets". In: *Séminaire de Probabilités XLV*. Springer, 2013, pp. 483–535 (cit. on pp. 2, 3, 7, 40).

[10] D. Benedetto and E. Caglioti. "Euclidean random matching in 2d for non-constant densities". In: *Journal of Statistical Physics* 181.3 (2020), pp. 854–869 (cit. on p. 3).

[11] S. Bobkov and M. Ledoux. "A simple Fourier analytic proof of the AKT optimal matching theorem". In: *The Annals of Applied Probability* 31.6 (2021), pp. 2567–2584 (cit. on p. 2).

[12] S. Bobkov and M. Ledoux. "One-dimensional empirical measures, order statistics, and Kantorovich transport distances". In: *Mem. Amer. Math. Soc.* 261.1259 (2019), pp. v+126 (cit. on pp. 2, 3).

[13] S. Bobkov and M. Ledoux. "Transport inequalities on Euclidean spaces for non-Euclidean metrics". In: *Journal of Fourier Analysis and Applications* 26.4 (2020), p. 60 (cit. on p. 2).

[14] G. Bouchitté, C. Jimenez, and M. Rajesh. "A new L^∞ estimate in optimal mass transport". In: *Proceedings of the American Mathematical Society* 135.11 (2007), pp. 3525–3535 (cit. on p. 20).

[15] J. Boutet de Monvel and O. C. Martin. "Almost sure convergence of the minimum bipartite matching functional in Euclidean space". In: *Combinatorica* 22 (2002), pp. 523–530 (cit. on p. 2).

[16] Y. Brenier. "Polar factorization and monotone rearrangement of vector-valued functions". English. In: *Commun. Pure Appl. Math.* 44.4 (1991), pp. 375–417 (cit. on p. 22).

[17] D. Bucur, G. Buttazzo, et al. "Variational methods in shape optimization problems". In: *(No Title)* (2000) (cit. on p. 11).

[18] E. Caglioti, M. Goldman, F. Pieroni, and D. Trevisan. *Subadditivity and optimal matching of unbounded samples*. cvgmt preprint. 2024 (cit. on pp. 2, 3, 31).

[19] S. Caracciolo, C. Lucibello, G. Parisi, and G. Sicuro. "Scaling hypothesis for the Euclidean bipartite matching problem". In: *Physical Review E* 90.1 (2014) (cit. on p. 7).

[20] S. Caracciolo, C. Lucibello, G. Parisi, and G. Sicuro. "Scaling hypothesis for the Euclidean bipartite matching problem". In: *Physical Review E* 90.1 (2014), p. 012118 (cit. on p. 3).

[21] L. Chizat, A. Delalande, and T. Vaškevičius. "Sharper Exponential Convergence Rates for Sinkhorn's Algorithm in Continuous Settings". In: *arXiv preprint arXiv:2407.01202* (2024) (cit. on pp. 7, 28).

[22] A. Delalande. "Quantitative Stability in Quadratic Optimal Transport". PhD thesis. Université Paris-Saclay, 2022 (cit. on pp. 7, 28, 29).

[23] A. Delalande and Q. Merigot. "Quantitative stability of optimal transport maps under variations of the target measure". In: *Duke Mathematical Journal* 172.17 (2023), pp. 3321–3357 (cit. on pp. 7, 28).

[24] S. Dereich, M. Scheutzow, and R. Schottstedt. "Constructive quantization: Approximation by empirical measures". In: *Annales de l'IHP Probabilités et statistiques*. Vol. 49. 4. 2013, pp. 1183–1203 (cit. on pp. 1–3, 7).

[25] V. Dobrić and J. E. Yukich. "Asymptotics for transportation cost in high dimensions". In: *Journal of Theoretical Probability* 8 (1995), pp. 97–118 (cit. on pp. 2, 3).

[26] A. Drewitz, B. Ráth, and A. Sapozhnikov. *An introduction to random interlacements*. Springer, 2014 (cit. on pp. 7, 41).

[27] A. Figalli and N. Gigli. "A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions". In: *Journal de mathématiques pures et appliquées* 94.2 (2010), pp. 107–130 (cit. on p. 2).

[28] N. Fournier and A. Guillin. "On the rate of convergence in Wasserstein distance of the empirical measure". In: *Probab. Theory Related Fields* 162.3-4 (2015), pp. 707–738 (cit. on p. 3).

[29] M. Goldman, M. Huesmann, and F. Otto. "Almost sharp rates of convergence for the average cost and displacement in the optimal matching problem". In: *arXiv preprint arXiv:2312.07995* (2023) (cit. on p. 7).

[30] M. Goldman and F. Otto. "A variational proof of partial regularity for optimal transportation maps". In: *Annales scientifiques de l'Ecole Normale Supérieure* 53.5 (2020), pp. 1209–1233 (cit. on p. 20).

[31] M. Goldman and D. Trevisan. "Convergence of asymptotic costs for random Euclidean matching problems". In: *Probability and Mathematical Physics* 2.2 (2021), pp. 341–362 (cit. on pp. 23, 31).

[32] M. Goldman and D. Trevisan. "Optimal transport methods for combinatorial optimization over two random point sets". In: *Probab. Theory Related Fields* (2023), pp. 1–70 (cit. on pp. 2, 3, 31).

[33] S. Graf and H. Luschgy. *Foundations of quantization for probability distributions*. Springer, 2007 (cit. on p. 1).

[34] M. Huesmann and K.-T. Sturm. "Optimal transport from Lebesgue to Poisson". In: *The Annals of Probability* 41.4 (2013), pp. 2426–2478 (cit. on p. 7).

[35] M. Ledoux. "On optimal matching of Gaussian samples". In: *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* 457.Veroyatnost' i Statistika. 25 (2017), pp. 226–264 (cit. on pp. 2, 23).

[36] C. Letrouit and Q. Mérigot. "Gluing methods for quantitative stability of optimal transport maps". In: *arXiv preprint arXiv:2411.04908* (2024) (cit. on p. 7).

[37] L. Lovász and M. D. Plummer. *Matching theory*. Vol. 367. American Mathematical Soc., 2009 (cit. on p. 1).

[38] M. Mariani and D. Trevisan. "Wasserstein Asymptotics for Brownian Motion on the Flat Torus and Brownian Interlacements". In: *arXiv preprint arXiv:2307.10325* (2023) (cit. on pp. 7, 31, 41, 43).

[39] M. Mezard and A. Montanari. *Information, physics, and computation*. Oxford University Press, 2009 (cit. on p. 1).

[40] M. Mézard, G. Parisi, and M. A. Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*. Vol. 9. World Scientific Publishing Company, 1987 (cit. on p. 1).

[41] O. Mischler and D. Trevisan. "Quantitative stability in optimal transport for general power costs". In: *arXiv e-prints*, arXiv:2407.19337 (July 2024), arXiv:2407.19337. arXiv: 2407.19337 [math.FA] (cit. on pp. 7, 28).

[42] R. Peyre. "Comparison between W_2 distance and $\dot{H}^{-1}$ norm, and localization of Wasserstein distance". English. In: *ESAIM, Control Optim. Calc. Var.* 24.4 (2018), pp. 1489–1501 (cit. on p. 23).

[43] G. Peyré, M. Cuturi, et al. "Computational Optimal Transport: With Applications to Data Science". In: *Foundations and Trends® in Machine Learning* 11.5-6 (2019), pp. 355–607 (cit. on p. 13).

[44] S. Port. *Brownian motion and classical potential theory*. Elsevier, 2012 (cit. on p. 41).

[45] M. Rathmair. "On how Poincaré inequalities imply weighted ones". In: *Monatshefte für Mathematik* 188.4 (2019), pp. 753–763 (cit. on p. 11).

[46] D. Ruiz. "On the uniformity of the constant in the Poincaré inequality". In: *Advanced Nonlinear Studies* 12.4 (2012), pp. 889–903 (cit. on p. 10).

[47] F. Santambrogio. *Optimal transport for applied mathematicians*. Vol. 87. Progress in Nonlinear Differential Equations and their Applications. Calculus of variations, PDEs, and modeling. Birkhäuser/Springer, Cham, 2015, pp. xxvii+353 (cit. on p. 24).

[48] J. M. Steele. *Probability theory and combinatorial optimization*. SIAM, 1997 (cit. on p. 2).

[49] A.-S. Sznitman. "On scaling limits and Brownian interlacements". In: *Bulletin of the Brazilian Mathematical Society, New Series* 44 (2013), pp. 555–592 (cit. on pp. 7, 41).

[50] M. Talagrand. "Matching random samples in many dimensions". In: *The Annals of Applied Probability* (1992), pp. 846–856 (cit. on p. 2).

[51] M. Talagrand. *Upper and lower bounds for stochastic processes: decomposition theorems*. Vol. 60. Springer Nature, 2022 (cit. on pp. 1, 2).

[52] D. Trevisan and M. Goldman. "On the concave one-dimensional random assignment problem and Young integration theory". In: *Annali Scuola Normale Superiore - Classe di Scienze* (2022), pp. 18–18 (cit. on p. 3).

[53] C. Villani. *Optimal transport: old and new*. Vol. 338. Grundlehren der Mathematischen Wissenschaften. Springer-Verlag, Berlin, 2009, pp. xxii+973 (cit. on pp. 1, 13).

[54] F.-Y. Wang. “Convergence in Wasserstein distance for empirical measures of Dirichlet diffusion processes on manifolds”. In: *Journal of the European Mathematical Society* (2022) (cit. on p. 8).

[55] F.-Y. Wang. “Convergence in Wasserstein distance for empirical measures of semilinear SPDEs”. In: *The Annals of Applied Probability* 33.1 (2023), pp. 70–84 (cit. on p. 8).

[56] F.-Y. Wang. “Precise limit in Wasserstein distance for conditional empirical measures of Dirichlet diffusion processes”. In: *Journal of Functional Analysis* 280.11 (2021), p. 108998 (cit. on p. 8).

[57] F.-Y. Wang. “Wasserstein convergence rate for empirical measures on noncompact manifolds”. In: *Stochastic Processes and their Applications* 144 (2022), pp. 271–287 (cit. on p. 8).

[58] F.-Y. Wang and B. Wu. “Wasserstein Convergence for Empirical Measures of Subordinated Diffusions on Riemannian Manifolds”. In: *Potential Analysis* (2022), pp. 1–22 (cit. on p. 8).

[59] F.-Y. Wang, B. Wu, and J.-X. Zhu. “Sharp L^q-Convergence Rate in p-Wasserstein Distance for Empirical Measures of Diffusion Processes”. In: *arXiv preprint arXiv:2408.09116* (2024) (cit. on p. 8).

[60] F.-Y. Wang and J.-X. Zhu. “Limit theorems in Wasserstein distance for empirical measures of diffusion processes on Riemannian manifolds”. In: *Annales de l’Institut Henri Poincare (B) Probabilites et statistiques*. Vol. 59. 1. Institut Henri Poincaré. 2023, pp. 437–475 (cit. on p. 8).

[61] J. E. Yukich. *Probability theory of classical Euclidean optimization problems*. Lecture Notes in Mathematics. Springer, 2006 (cit. on p. 2).