hexagon:2611.01214v1

EQUALITY OF THE ORDINARY AND BOUNDARY ASYMPTOTIC COSTS FOR THE RANDOM MATCHING PROBLEM

MICHAEL GOLDMAN, MARTIN HUESMANN, DARIO TREVISAN, AND ARTHUR WACHTEL

ABSTRACT. For a stationary random measure on $\mathbb{R}^d$, $d \geq 3$, we show that the periodic, ordinary and boundary p-transport costs to Lebesgue measure on large cubes have the same limit per unit volume, for every $p \geq 1$. We assume the concentration hypothesis of Huesmann, Goldman and Trevisan, a centering condition on the optimal stationary displacement and, for $p = 1$, that the measure is atomic. This extends their quadratic result. The theorem applies to compound Poisson measures with exponentially integrable marks and, for $p > 1$ and $d \geq 5$, to Brownian interlacements. In particular, it gives the equality of the ordinary and boundary constants conjectured by Caglioti, Goldman, Pieroni and Trevisan, and, through an identity of Xu, of the Barthe–Bordenave constants for bipartite matching when $1 \leq p < d/2$.

1. INTRODUCTION

1.1. **Random matching constants.** Let $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ be independent uniform points in the unit cube of $\mathbb{R}^d$ and let $p \geq 1$. The random bipartite matching problem asks for the asymptotic behaviour of

$$\min_{\sigma} \sum_{i=1}^{n} |X_i - Y_{\sigma(i)}|^p,$$

the minimum being over all permutations of $\{1, \dots, n\}$. A closely related semi-discrete problem replaces one of the two samples by the uniform measure. For the linear cost the order of magnitude of the matching cost was determined by [1] in dimension two and by [24] in higher dimensions. In dimension $d \geq 3$ a typical point is matched at distance of order $n^{-1/d}$, so that the cost is of order $n^{1-p/d}$ (see [11] for every $p \geq 1$), and the question is whether $n^{p/d-1}$ times the cost converges.

In the language of transport on large cubes, this asks for the existence of a limiting cost per unit volume. Dobrić and Yukich [7] proved convergence for uniform samples when $p = 1$ and $d \geq 3$. Barthe and Bordenave [3, Theorem 2] proved almost sure convergence of the bipartite matching cost of uniform samples for $0 < p < d/2$, and Dereich et al. [6, Theorem 2] proved convergence of the expected transport cost between the empirical measure and the uniform distribution for $1 \leq p < d/2$. In both works the restriction $p < d/2$ comes from the fluctuations of the number of points: a discrepancy of order $L^{d/2}$ between the masses in a cube of side L, repaired by edges of length of order L, costs $L^{p+d/2}$, which is negligible with respect to the volume only when $p < d/2$. Goldman and Trevisan [11] removed this restriction and proved convergence of the expected cost for every $p \geq 1$ and $d \geq 3$, using subadditivity together with a PDE ansatz in the spirit of [2] and of the prediction of [5].

Subadditivity arguments naturally produce two constants. Ordinary transport on a cube is subadditive up to errors, which gives the upper constant. For the lower

Date: November 27, 2026.

2020 *Mathematics Subject Classification.* 60D05, 49Q22, 60G55, 90C27.

Key words and phrases. Random matching, bipartite matching, semi-discrete matching, optimal transport, Poisson point process, compound Poisson process, Brownian interlacements, boundary conditions, stationary allocation.

bound one needs a superadditive functional, in which mass may be exchanged with the boundary of the cube. Barthe and Bordenave [3] penalize each unmatched point by the pth power of its distance to the boundary and obtain a lower constant $\beta'_p(d)$ defined through Poisson samples [3, Lemma 27]; in [6] the uniform measure may leak to the boundary; and Caglioti et al. [4] use the boundary transport cost $W_Q^{b,p}$ of [8], in which mass may be sent to and received from the boundary at the price of the distance, and show that the corresponding semi-discrete constant exists and is positive [4, Proposition A.3]. All these authors expect the lower and upper constants to coincide. Barthe and Bordenave write that they strongly suspect $\beta_p(d) = \beta'_p(d)$ [3, Section 1]; Dereich, Scheutzow and Schottstedt conjecture the equality of the constants with and without leakage [6, Remark 2]; and Caglioti et al. [4, Remark 1.3] record the conjecture that the boundary and ordinary constants coincide for every p and d, noting that this is known only for $p = 2$. For $p = 2$ and $d \geq 3$ the equality is due to [14], who compare the ordinary and boundary costs for general stationary random measures with quadratic cost; for $p \neq 2$ the extension of their argument is obstructed by the lack of an analogue of the quadratic identity relating the optimal displacement to the gradient of a dual potential. Huesmann and Wachtel [16] recently identified the limit of the boundary cost per unit volume with the cost of the optimal stationary allocation for the Poisson process, for every $p \geq 1$, and state that the comparison with the ordinary cost remains open for $p \neq 2$ [16, Remark 1.4].

In this paper we prove the equality of the periodic, ordinary and boundary constants in expectation, for every $p \geq 1$ and $d \geq 3$, for semi-discrete matching and for bipartite matching. For semi-discrete matching this gives, through the bounds recalled in [14, Theorem 1.1], the existence of the limit of $n^{p/d}\mathbb{E}W_p^p(\frac{1}{n}\sum_i \delta_{X_i}, \lambda)$ for samples of a general law λ (Corollary 8.2), which is the question raised in [14, Section 1.1]. For bipartite matching it gives $\beta_p(d) = \beta'_p(d)$ when $1 \leq p < d/2$, and therefore, by [3, Theorem 2], an almost sure limit for samples with general densities (see Section 8, where we also discuss the constants of Dereich, Scheutzow and Schottstedt). All these results are consequences of a single theorem (Theorem 1.2) on transport towards a stationary random measure satisfying the concentration hypothesis of [14], applied to *compound* (independently marked) Poisson measures. The same theorem extends the quadratic result of [14] for the occupation measure of Brownian interlacements to every $p > 1$ in dimension $d \geq 5$.

1.2. **The setting and the three costs.** Let $d \geq 3$ and let ν be a random locally finite Borel measure on $\mathbb{R}^d$. We assume:

(A1) (stationarity) the law of ν is invariant under translations, and $\mathbb{E}\nu([0,1)^d) = 1$;
(A2) (concentration) there is $\alpha \in [0, d-2)$ such that for every $q \geq 1$, every cube Q with $\operatorname{diam} Q \geq 1$ and every bounded Borel $f \geq 0$ on Q,

$$\Big\| \int_Q f \, d\nu - \int_Q f \, dx \Big\|_{L^q(\mathbb{P})} \leq C_q \operatorname{diam}(Q)^{(d+\alpha)/2} \|f\|_\infty. \tag{1.1}$$

These are the hypotheses of [14, Assumption 4.1], under which the quadratic case of our main result is [14, Theorem 4.3]; the condition $\alpha < d-2$ is their condition $d > 2+\alpha$. Our main example is the *compound* (or marked) Poisson measure: let $\mathcal{P} = \{X_i\}$ be a unit-intensity Poisson point process on $\mathbb{R}^d$, and attach to its points independent marks W_i, independent of $\mathcal{P}$, with common law satisfying

$$W \geq 0, \qquad \mathbb{E}W = 1, \qquad \mathbb{E}e^{t_0 W} < \infty \quad \text{for some } t_0 > 0. \tag{H}$$

Then $\nu^W = \sum_i W_i \delta_{X_i}$ satisfies (A1)–(A2) with $\alpha = 0$ (Lemma 2.2). Points with mark zero do not contribute to ν^W; its atoms are the points X_i with $W_i > 0$. Unit marks give the Poisson process and exponential marks are related to bipartite matching (Section 8).

Another example is the occupation measure of Brownian interlacements, with $\alpha = 2$ (Section 8.3).

For each real $L > 0$ put

$$Q_L = (0, L)^d, \quad \nu_L = \nu|_{Q_L}, \quad \Sigma_L = \nu(Q_L), \quad a_L = \Sigma_L / L^d, \quad \lambda_L = a_L \mathbf{1}_{Q_L} dx. \tag{1.2}$$

The finite measures λ_L and ν_L have the common total mass Σ_L (possibly zero). For finite measures μ, λ of equal mass, $\Pi(\mu, \lambda)$ is the set of their couplings; no normalization by the mass is made. Let $\rho_L(x) = \operatorname{dist}(x, Q_L^c)$, let $d_L^{\mathrm{per}}(x, y) = \min_{k \in \mathbb{Z}^d} |x - y + Lk|$ be the distance of the flat torus $\mathbb{R}^d/(L\mathbb{Z})^d$, and let

$$c_{b,p,L}(x, y) = \min\{|x - y|^p, \rho_L(x)^p + \rho_L(y)^p\}.$$

The *periodic*, *ordinary* and *boundary* costs are

$$T_L = \inf_{\pi \in \Pi(\lambda_L, \nu_L)} \int d_L^{\mathrm{per}}(x, y)^p \, d\pi, \tag{1.3}$$

$$U_L = \inf_{\pi \in \Pi(\lambda_L, \nu_L)} \int |x - y|^p \, d\pi, \tag{1.4}$$

$$B_L = \inf_{\pi \in \Pi(\lambda_L, \nu_L)} \int c_{b,p,L}(x, y) \, d\pi. \tag{1.5}$$

On $\{\Sigma_L = 0\}$ all three costs vanish. In the boundary cost a pair (x, y) may be connected directly or through the boundary, at the price $\rho_L(x)^p + \rho_L(y)^p$; for balanced marginals this is the reservoir boundary cost of [4, (2.5)] and [16, (2.1)] (see Remark 2.5). The exponent $p \geq 1$ is fixed throughout and is suppressed from the notation. For $\nu = \nu^W$ we write T_L^W, U_L^W, B_L^W.

Lemma 1.1. *For every $p \geq 1$, $L > 0$ and $x, y \in Q_L$,*

$$c_{b,p,L}(x, y) \leq d_L^{\mathrm{per}}(x, y)^p \leq |x - y|^p. \tag{1.6}$$

Consequently, for every realization,

$$0 \leq B_L \leq T_L \leq U_L, \qquad 0 \leq U_L - T_L \leq U_L - B_L. \tag{1.7}$$

Proof. The second inequality in (1.6) follows by taking $k = 0$. For the first, choose a minimizing $k \in \{-1, 0, 1\}^d$. If $k = 0$, then $c_{b,p,L}(x, y) \leq |x - y|^p = d_L^{\mathrm{per}}(x, y)^p$. Otherwise choose a coordinate j with $k_j \neq 0$. The corresponding coordinate difference has length $x_j + L - y_j$ or $L - x_j + y_j$, so it is at least $\rho_L(x) + \rho_L(y)$. Hence, since $p \geq 1$, $d_L^{\mathrm{per}}(x, y)^p \geq (\rho_L(x) + \rho_L(y))^p \geq \rho_L(x)^p + \rho_L(y)^p$. Taking infima over the same set of couplings gives (1.7). □

1.3. **Main results.** A stationary coupling of Lebesgue measure and ν is a coupling whose joint law with ν is invariant under translations (Definition 3.1; a coupling using additional randomness can be averaged over it given ν, which preserves the cost); its specific cost is the expected cost per unit volume of source. Let

$$e_\nu = e_\nu(p, d) = \inf\{\text{specific } p\text{-cost of stationary couplings of } dx \text{ and } \nu\},$$

and let $\mathcal{I}$ be the σ-algebra of events invariant under all translations.

Theorem 1.2. *Let $d \geq 3$, $1 \leq p < \infty$, and let ν satisfy (A1)–(A2). If $p = 1$, assume moreover that ν is almost surely purely atomic. Then $e_\nu < \infty$ (proved in Corollary 4.2), and there is a unique optimal stationary coupling of dx and ν; it is induced by a translation-equivariant map T_ν, whose displacement has the form $x - T_\nu(\omega, x) = \beta_\nu(\tau_x \omega)$. Assume in addition the centering condition*

$$\mathbb{E}[\beta_\nu \mid \mathcal{I}] = 0. \tag{1.8}$$

Then, along all real $L \to \infty$,

$$\lim_{L\to\infty} L^{-d}\mathbb{E}[U_L - B_L] = 0, \tag{1.9}$$

and the three specific costs have the common limit

$$\lim_{L\to\infty} L^{-d}\mathbb{E}T_L = \lim_{L\to\infty} L^{-d}\mathbb{E}U_L = \lim_{L\to\infty} L^{-d}\mathbb{E}B_L = e_\nu(p,d). \tag{1.10}$$

Moreover $e_\nu > 0$ unless ν is almost surely Lebesgue measure.

The centering condition (1.8) holds in each of the following cases (Proposition 3.5): $p = 2$; or the law of ν is ergodic and invariant under the reflection $x \mapsto -x$. Under (A1)–(A2), (1.8) is equivalent to convergence of $L^{-d}\mathbb{E}U_L$ to e_ν (Proposition 7.3). It can fail even with $\alpha = 0$ (Remark 7.4). A randomly shifted lattice shows that (A1) and centering alone do not suffice (Remark 7.5). For $p = 2$, Theorem 1.2 thus contains [14, Theorem 4.3], of which it gives a different proof. For $p = 1$ and diffuse ν we do not know whether the optimal stationary coupling is unique.

Theorem 1.3. *Let $d \geq 3$, $1 \leq p < \infty$, and let the marks satisfy (H). Then Theorem 1.2 applies to $\nu = \nu^W$: with $e_W(p,d) = e_{\nu^W}(p,d)$, we have $0 < e_W(p,d) < \infty$, the optimal stationary coupling of dx and ν^W is unique and induced by a translation-equivariant map, and*

$$\lim_{L\to\infty} L^{-d}\mathbb{E}[U_L^W - B_L^W] = 0,$$
$$\lim_{L\to\infty} L^{-d}\mathbb{E}T_L^W = \lim_{L\to\infty} L^{-d}\mathbb{E}U_L^W = \lim_{L\to\infty} L^{-d}\mathbb{E}B_L^W = e_W(p,d).$$

Theorem 1.3 has two specializations, derived in Section 8.

- *Unit marks $W \equiv 1$.* Then ν is the Poisson process and $e_1(p,d)$ is the cost $e_*(p,d)$ of the optimal stationary allocation of [15]. Corollary 8.1 gives the conjecture recorded in [4] for every $p \geq 1$, $d \geq 3$, and Corollary 8.2 the limit for samples of a general law.
- *Exponential marks $W \sim \mathrm{Exp}(1)$.* By the averaging identity of [26, Theorem 2], the expected marked costs are exactly Poisson mixtures of the expected bipartite assignment costs. Theorem 8.3 gives the equality of the periodic, ordinary and boundary constants of bipartite matching for every $p \geq 1$, $d \geq 3$, and Corollary 8.7 gives $\beta_p(d) = \beta'_p(d)$ for $1 \leq p < d/2$, together with its consequence for general densities.

For the occupation measure of Brownian interlacements, Corollary 8.10 gives the equality of the three constants for every $p > 1$ in dimension $d \geq 5$; the case $p = 2$ is [14, Theorem 4.3].

1.4. **Strategy of the proof.** By Lemma 1.1 the main point is $\mathbb{E}U_L \leq \mathbb{E}B_L + o(L^d)$. For $p = 2$ this is [14, Theorem 4.3], whose proof combines a deterministic boundary repair [14, Theorem 3.7] with a control of the displacement flux through the Kantorovich potential: $g = \nabla u/2$, so that $\nabla u \cdot \nabla \eta = \mathrm{div}(u\nabla\eta) - u\Delta\eta$ can be estimated through oscillation bounds on u. No such representation is available for $p \neq 2$. We keep the boundary repair and replace the potential by two qualitative mechanisms, which are the new part of the argument.

1. Strong screening (Section 4). Let π_L be a boundary-optimal plan. We discard its strict shortcuts (pairs connected through the boundary), rescale the remaining direct part so that it is dominated by both marginals, tile space by translated copies of the cube using the *same* realization of ν, and average over the position of the grid. This stationarization follows [16, Section 3], who use it to identify the limit of the boundary cost with the optimal stationary cost e_ν. The resulting stationary subcouplings have vanishing defects and energies converging to e_ν. Since the environment is kept in the joint

law, compactness and uniqueness of the stationary optimizer give *strong* convergence to the graph of the optimal map T_ν in the original realization (Lemma 3.7). Transferred back to the cube by an exact unfolding identity, this shows that the barycentric displacement g_L of the direct edges of π_L is close in L^p to the stationary displacement $b_\nu = x - T_\nu$, that the shortcut energy is $o(L^d)$, and that the direct-edge costs are uniformly integrable (Theorem 4.6).

2. Ergodic flux cancellation (Section 5). The stationary displacement b_ν is a stationary field, and by the ergodic theorem and the centering condition (1.8) its averages over mesoscopic blocks are small in L^p. A two-scale Poincaré inequality turns this into smallness of $g_L \cdot \nabla \eta_L$ minus its mean in a negative Sobolev norm, for any C^2 macroscopic cutoff η_L (Proposition 5.2). The argument needs no rate and no gradient structure of g_L, and it includes $p = 1$.

3. Boundary repair (Sections 6 and 7). Following [14, Theorem 3.7], we keep the plan π_L in the bulk and replace it, in a boundary layer of width δL, by a transport to a multiple of Lebesgue measure. The remaining discrepancy is a density, whose transport cost is bounded by a negative Sobolev norm [14, Lemma 3.6]; it splits into a mass correction, the flux of step 2, the shortcut energy, and Taylor remainders controlled by uniform integrability. The layer cost is controlled by the dyadic estimate in Lemma 7.1.

The stationary transport theory we use is classical. For point processes, existence and uniqueness of the optimal stationary coupling and its representation by a map are due to [15, Theorem 1.1], and in a general framework to [13, Theorems 1.1–1.3]. Since we need full translation equivariance, real masses and the linear cost, we include in Proposition 3.4 a short proof: an equivariant n-point exchange shows that optimal stationary couplings are cyclically monotone, which gives the map property for $p > 1$ by the theorem of Gangbo and McCann, and for $p = 1$ when ν is atomic by a two-point argument. The limit of the boundary cost is [16, Theorem 2.2]; we give a short self-contained proof (Proposition 4.4). The finite-volume inputs taken from the literature are the ordinary cost bound of [14, Proposition 4.2] and their negative Sobolev estimate for $p > 1$ [14, Lemma 3.6]. For the boundary-layer estimate we give a dyadic argument adapted from [14, Lemma 4.5]. The applications in Section 8 also use the de-Poissonization of [4, Proposition A.1]. The proof is qualitative: it gives no rate of convergence, and it does not address $p = \infty$.

1.5. **Organization and conventions.** Section 2 collects the consequences of (A1)–(A2), the properties of the marked Poisson measure, and those of the boundary cost. Section 3 studies stationary couplings: compactness, existence, uniqueness and strong graph convergence. Section 4 stationarizes boundary-optimal plans, identifies the limit of the boundary cost, and proves the strong screening theorem. Section 5 proves the flux estimate, Section 6 the deterministic comparison, and Section 7 estimates the boundary layer and proves Theorems 1.2 and 1.3. Section 8 derives the consequences for semi-discrete and bipartite matching, and treats Brownian interlacements. Section A contains a proof of Xu's averaging identity.

Throughout, $p \geq 1$ and $d \geq 3$ are fixed. For a cost c and finite measures μ, λ of equal mass, $\mathrm{OT}_c(\mu, \lambda)$ denotes the unnormalized optimal transport cost, and $W_p^p = \mathrm{OT}_{|x-y|^p}$. In W_p^p, $W^{b,p}$, $W^{1,p'}$ and $W_\#^{-1,p}$ the letter W does not refer to the marks, and the periodic cost T_L should not be confused with the optimal map T_ν. Constants C, c depend only on p, d and the constants in (A2) (for ν^W, on the law of W) unless further dependence is indicated, as in $C(\delta)$. For $p \geq 1$ we let $p' = p/(p-1) \in (1, \infty]$.

2. Preliminaries

We collect the probabilistic properties of ν used later, and we write the boundary cost as a transport cost for a pseudometric, which gives a triangle inequality between measures of different masses.

2.1. Consequences of the assumptions.

Lemma 2.1. *Let ν satisfy (A1)–(A2). For every $q \geq 1$ and $L \geq 1$,*

$$\begin{aligned} \mathbb{E}|\Sigma_L - L^d|^q \leq C_q L^{q(d+\alpha)/2}, \qquad \mathbb{E}\Sigma_L^2 \leq CL^{2d}, \\ \mathbb{E}|a_L - 1|^q + \mathbb{P}(a_L \notin [\tfrac{1}{2}, 2]) \leq C_q L^{-q(d-\alpha)/2}, \end{aligned} \tag{2.1}$$

and $\mathbb{E}|\int f\,d\nu - \int f\,dx|^q \leq C_q L^{q(d+\alpha)/2}\|f\|_\infty^q$ for every bounded Borel $f \geq 0$ supported in $\overline{Q}_L$. Moreover

$$\sup_{L\geq 1} L^{-d}\mathbb{E}U_L < \infty, \qquad \textit{hence} \qquad \sup_{L\geq 1} L^{-d}\mathbb{E}B_L < \infty. \tag{2.2}$$

Proof. The moment bounds are (1.1) with $\operatorname{diam} Q_L = \sqrt{d}L$ (and $f = \mathbf{1}_{Q_L}$ for Σ_L). Then $\mathbb{E}|a_L - 1|^q = L^{-dq}\mathbb{E}|\Sigma_L - L^d|^q$, $\mathbb{P}(a_L \notin [\frac{1}{2}, 2]) \leq \mathbb{P}(|\Sigma_L - L^d| \geq L^d/2) \leq 2^q L^{-dq}\mathbb{E}|\Sigma_L - L^d|^q$ by Markov's inequality, and $\mathbb{E}\Sigma_L^2 \leq 2L^{2d} + 2\mathbb{E}|\Sigma_L - L^d|^2$. Since $\alpha < d - 2$, [14, Proposition 4.2] applies and gives $\mathbb{E}U_L \leq CL^d$ for large L and every $p \geq 1$; for bounded L we use $U_L \leq d^{p/2}L^p\Sigma_L$. Finally $B_L \leq U_L$. □

Since $\alpha < d$, the bad event $G_L^c = \{a_L \notin [\frac{1}{2}, 2]\}$ has probability $O(L^{-k})$ for every k; this replaces exponential bounds throughout.

2.2. The marked Poisson measure.

Lemma 2.2. *Assume (H), let $\nu = \nu^W$ and $m_W = \mathbb{E}[W^2 e^{t_0 W/2}] < \infty$. For a bounded measurable $f : \mathbb{R}^d \to \mathbb{R}$ with compact support, let $Z_f = \int f\,d\nu - \int f\,dx$. Then*

$$\log \mathbb{E}e^{tZ_f} = \int \left(\mathbb{E}e^{tWf(x)} - 1 - tf(x)\right) dx \leq \frac{m_W}{2}\, t^2 \int f^2 dx \qquad \textit{for } |t|\,\|f\|_\infty \leq t_0/2, \tag{2.3}$$

in particular $\mathbb{E}Z_f = 0$ and $\operatorname{Var} Z_f = \mathbb{E}[W^2]\int f^2 dx$. For every $q \geq 1$,

$$\|Z_f\|_{L^q(\mathbb{P})} \leq C_{q,W}\left(\|f\|_{L^2} + \|f\|_{L^\infty}\right), \tag{2.4}$$

and, for every $L \geq 1$,

$$\mathbb{E}|\Sigma_L - L^d|^q \leq C_{q,W} L^{dq/2}, \qquad \mathbb{P}(a_L \notin [1/2, 2]) \leq 2e^{-c_W L^d}. \tag{2.5}$$

Finally, the law of ν is invariant under translations and under the reflection $x \mapsto -x$, the restrictions of ν to disjoint Borel sets are independent, and the law of ν is mixing, hence ergodic, under translations.

Proof. The marked process $\{(X_i, W_i)\}$ is a Poisson process on $\mathbb{R}^d \times [0, \infty)$ with intensity $dx \otimes \operatorname{Law}(W)$ (marking theorem), and the exponential formula for Poisson processes [see 19] gives $\log \mathbb{E}e^{t\int f\,d\nu} = \int (\mathbb{E}e^{tWf(x)} - 1)\,dx$ whenever the right side is finite; it is finite for $|t|\|f\|_\infty \leq t_0$ by (H). Since $\mathbb{E}W = 1$, this is the identity in (2.3). For the inequality, $e^u - 1 - u \leq \frac{u^2}{2}e^{|u|}$ with $u = tWf(x)$ and $|u| \leq t_0 W/2$. Expanding at $t = 0$ gives the mean and the variance. The bound (2.3), applied to $\pm Z_f$, and Chernoff's inequality give

$$\mathbb{P}(|Z_f| > u) \leq 2\exp\Big(-\min\Big\{\frac{u^2}{2m_W\|f\|_2^2}, \frac{t_0 u}{4\|f\|_\infty}\Big\}\Big),$$

and integrating this tail gives (2.4). With $f = \mathbf{1}_{Q_L}$ and $u = L^d/2$ we get (2.5). The invariance under translations and reflection and the independence on disjoint sets are inherited from the marked Poisson process. If A, B are events determined by the restrictions of ν to bounded sets, then A and the event B_z obtained by translating

the configuration by z are independent for $|z|$ large, so $\mathbb{P}(A \cap B_z) \to \mathbb{P}(A)\mathbb{P}(B)$; by approximation this extends to all events, which is mixing. □

In particular, for every cube Q with $\operatorname{diam} Q \geq 1$ and every bounded measurable $f \geq 0$ on Q, (2.4) gives $\|Z_f\|_{L^q} \leq C_{q,W,d} \operatorname{diam}(Q)^{d/2} \|f\|_\infty$. Thus ν^W satisfies (A1)–(A2) with $\alpha = 0$, that is [14, Assumption 4.1] with $\alpha = 0$; this was observed for exponential marks in [26, Appendix A.2]. It is purely atomic, and its law is ergodic and reflection invariant.

2.3. **The boundary cost as a transport distance.** Let $X_L = Q_L \sqcup \{\partial\}$, where ∂ is an extra point, and define

$$d_{b,L}(x,y) = \min\{|x-y|, (\rho_L(x)^p + \rho_L(y)^p)^{1/p}\}, \quad d_{b,L}(x,\partial) = \rho_L(x), \quad d_{b,L}(\partial,\partial) = 0,$$

for $x, y \in Q_L$. Then $d_{b,L}^p = c_{b,p,L}$ on $Q_L \times Q_L$.

Lemma 2.3. *$d_{b,L}$ is a pseudometric on X_L.*

Proof. Write $\rho = \rho_L$ and $\|(s,t)\|_p = (s^p + t^p)^{1/p}$. Symmetry is clear. We check $d(x,z) \leq d(x,y) + d(y,z)$. If both terms on the right are direct, use $|x-z| \leq |x-y| + |y-z|$. If $d(x,y) = |x-y|$ and $d(y,z) = \|(\rho(y), \rho(z))\|_p$, then, since ρ is 1-Lipschitz, Minkowski's inequality gives $\|(\rho(x), \rho(z))\|_p \leq \|(\rho(y) + |x-y|, \rho(z))\|_p \leq |x-y| + \|(\rho(y), \rho(z))\|_p$; the symmetric case is identical. If both are shortcuts, then $\|(\rho(x), \rho(z))\|_p \leq \rho(x) + \rho(z) \leq d(x,y) + d(y,z)$. Triangles through ∂ follow from $\|(\rho(x), \rho(y))\|_p \leq \rho(x) + \rho(y)$ and from $\rho(x) \leq \min\{|x-y| + \rho(y), \|(\rho(x), \rho(y))\|_p\}$. □

For finite measures μ, λ on Q_L, of possibly different masses, the *boundary* (or reservoir) cost is

$$W_{Q_L}^{b,p}(\mu,\lambda) = \inf\Big\{\mathcal{W}(\mu + s\delta_\partial, \lambda + t\delta_\partial) : s,t \geq 0,\ \mu(Q_L) + s = \lambda(Q_L) + t\Big\}, \tag{2.6}$$

where $\mathcal{W}$ is the optimal transport cost on X_L with cost $d_{b,L}^p$.

Lemma 2.4. *Let μ, λ, κ be finite measures on Q_L.*

(i) If $\mu(Q_L) = \lambda(Q_L)$, then $W_{Q_L}^{b,p}(\mu,\lambda)$ is the transport cost with cost $c_{b,p,L}$; in particular $B_L = W_{Q_L}^{b,p}(\lambda_L, \nu_L)$.

(ii) $W_{Q_L}^{b,p}$ is symmetric, and for every $\varepsilon \in (0,1]$, $W_{Q_L}^{b,p}(\mu,\lambda) \leq (1+\varepsilon)^{p-1} W_{Q_L}^{b,p}(\mu,\kappa) + (1+\varepsilon^{-1})^{p-1} W_{Q_L}^{b,p}(\kappa,\lambda)$.

(iii) For every $a \geq 0$, $W_{Q_L}^{b,p}(a\,dx|_{Q_L}, dx|_{Q_L}) \leq C_{d,p} L^{d+p} |a-1|^p$.

Proof. (i) Let γ couple $\mu + s\delta_\partial$ and $\lambda + s\delta_\partial$. The mass sent from Q_L to ∂ and the mass sent from ∂ to Q_L are both equal to $\mu(Q_L) - \gamma(Q_L \times Q_L)$. Pairing them by a product coupling, and using $\rho(x)^p + \rho(y)^p \geq c_{b,p,L}(x,y)$, gives a coupling of μ and λ whose $c_{b,p,L}$-cost does not exceed the cost of γ. The reverse inequality is obtained with $s = t = 0$.

(ii) Symmetry follows from the symmetry of $d_{b,L}$ and of (2.6). Take admissible (s_1,t_1) for (μ,κ) and (s_2,t_2) for (κ,λ), with costs within ε' of the infima. Adding $s_2\delta_\partial$ to both marginals of the first problem and $t_1\delta_\partial$ to both marginals of the second does not increase their costs (the added mass stays at ∂), and produces the three measures $\mu + (s_1+s_2)\delta_\partial$, $\kappa + (t_1+s_2)\delta_\partial$, $\lambda + (t_1+t_2)\delta_\partial$ of equal mass. By Lemma 2.3 and the gluing lemma, $\mathcal{W}^{1/p}$ satisfies the triangle inequality, and the elementary inequality $(\alpha+\beta)^p \leq (1+\varepsilon)^{p-1}\alpha^p + (1+\varepsilon^{-1})^{p-1}\beta^p$ gives the claim after $\varepsilon' \downarrow 0$.

(iii) Let c be the center of Q_L, $\ell = a^{1/d}$ and $S(x) = c + \ell(x-c)$, so that $S_\#(a\,dx|_{Q_L}) = dx|_{S(Q_L)}$ and $|x - S(x)| \leq |\ell - 1|\sqrt{d}L/2 \leq |a-1|\sqrt{d}L/2$ for $a \leq 2$. If $a \geq 1$, then $Q_L \subset S(Q_L)$: send x to $S(x)$ when $S(x) \in Q_L$, and to ∂ otherwise, at cost $\rho_L(x)^p \leq |x - S(x)|^p$. If $a < 1$, then $S(Q_L) \subset Q_L$: send x to $S(x)$, and supply $dx|_{Q_L \setminus S(Q_L)}$ from ∂, at cost $\rho_L(y)^p \leq ((1-\ell)L/2)^p$ per unit mass. In both cases the cost is at most $C_{d,p} L^{d+p} |a-1|^p$

for $a \le 2$. For $a > 2$ send all of $a\,dx$ to ∂ and supply dx from ∂: the cost is at most $(a+1)L^d(L/2)^p \le C_p L^{d+p}|a-1|^p$. □

Remark 2.5. The cost (2.6) coincides with the boundary cost $W^{b,p}_{Q_L}$ of [4, (2.5)] and [16, (2.1)], defined as the infimum of $\int |x-y|^p\,d\pi$ over measures π on $\overline{Q}_L \times \overline{Q}_L$ whose marginals restricted to Q_L are μ and λ: an edge from $x \in Q_L$ to ∂Q_L costs at least $\rho_L(x)^p$, with equality at a nearest boundary point, an edge of $Q_L \times Q_L$ costs at least its $c_{b,p,L}$-cost, and edges inside ∂Q_L can be removed. Conversely, choosing measurably a nearest boundary point $\bar{x}$ of each $x \in Q_L$, an edge (x,∂) is realized as $(x,\bar{x})$, an edge (∂,y) as $(\bar{y},y)$, and a shortcut edge (x,y) as the two edges $(x,\bar{x})$ and $(\bar{y},y)$, at the same cost; this is admissible since the marginals on ∂Q_L are free.

Part (iii) is the estimate [4, Lemma 2.5, (2.7)], there stated for constant densities m, m' on a bounded set in the form $\lesssim \operatorname{diam}^{p+d} |m-m'|^p / \max(m,m')^{p-1}$; we include the short proof of the version needed here. A comparison of balanced and unbalanced boundary costs through (ii)–(iii) is used in [16, Proposition 4.1] and in [4, Proposition A.3, (A.4)].

3. Stationary couplings

This section concerns transport on the whole space. We define stationary couplings of Lebesgue measure and ν through their joint law with the environment (Section 3.1), prove a compactness statement for stationary subcouplings with small defects (Lemma 3.3), and show that the optimal stationary coupling exists, is unique and is induced by an equivariant map, for every $p \ge 1$ (Proposition 3.4). Combining these facts with convergence of the energies gives convergence to the graph of the optimal map in L^p (Lemma 3.7), which is the form used in Section 4. In this section we only use (A1), and, for $p = 1$, that ν is purely atomic. Finiteness of e_ν is proved in Corollary 4.2, by a finite-volume argument; in this section it is assumed where needed.

3.1. Invariant couplings and mass transport.

Let Ω be the space of locally finite Borel measures on $\mathbb{R}^d$ with the vague topology, a Polish space [17], and let $\mathbf{P}$ be the law of ν on Ω. For $z \in \mathbb{R}^d$ let $(\tau_z\omega)(A) = \omega(A+z)$ be the environment seen from z, and let

$$\Theta_z(\omega,x,y) = (\tau_z\omega, x-z, y-z) \quad \text{on } \Omega \times \mathbb{R}^d \times \mathbb{R}^d.$$

Let $m(d\omega,dx) = \mathbf{P}(d\omega)dx$ and $m'(d\omega,dy) = \mathbf{P}(d\omega)\omega(dy)$. Both are invariant under $(\omega,x) \mapsto (\tau_z\omega, x-z)$, by stationarity of $\mathbf{P}$ (for m', substitute $y' = y - z$ and use $(\tau_z\omega)(dy')$).

Definition 3.1. An *invariant subcoupling* is a measure $\mathbf{q}$ on $\Omega \times \mathbb{R}^d \times \mathbb{R}^d$, invariant under every Θ_z, whose projections $\mathbf{q}_{\omega,x}$ and $\mathbf{q}_{\omega,y}$ under $(\omega,x,y) \mapsto (\omega,x)$ and $(\omega,x,y) \mapsto (\omega,y)$ satisfy $\mathbf{q}_{\omega,x} \le m$ and $\mathbf{q}_{\omega,y} \le m'$. It is an *invariant* (or *stationary*) *coupling* if both are equalities. Its *energy* and its *defect* are

$$e(\mathbf{q}) = \int \mathbf{1}_{[0,1)^d}(x)|x-y|^p\,d\mathbf{q},$$

$$\theta(\mathbf{q}) = \max\{(m-\mathbf{q}_{\omega,x})(\Omega\times[0,1)^d),\ (m'-\mathbf{q}_{\omega,y})(\Omega\times[0,1)^d)\},$$

and $e_\nu = e_\nu(p,d) = \inf\{e(\mathbf{q}) : \mathbf{q} \text{ an invariant coupling}\}$.

Disintegrating with respect to $\mathbf{P}$ the finite measures $\mathbf{1}_{x\in(-N,N)^d}\mathbf{q}$, whose environment marginals are at most $(2N)^d\mathbf{P}$, and patching the kernels by their almost sure consistency in N, we write $\mathbf{q} = \mathbf{P}(d\omega)q^\omega(dx,dy)$ with q^ω locally finite; for an invariant coupling, q^ω is a coupling of dx and ω for $\mathbf{P}$-a.e. ω, and by uniqueness of disintegrations, for each z, $q^{\tau_z\omega}$ is the image of q^ω under $(x,y) \mapsto (x-z,y-z)$ for $\mathbf{P}$-a.e. ω. Conversely, a measurable family of couplings with this equivariance defines an invariant coupling. Since the first marginal of q^ω is dx, we can write $q^\omega(dx,dy) = dx\,q^\omega_x(dy)$ with q^ω_x a

probability measure on $\mathbb{R}^d$. We denote by $\mathcal{I}$ the σ-algebra of Borel sets $A \subset \Omega$ with $\tau_z A = A$ for every $z \in \mathbb{R}^d$.

Lemma 3.2. *Let $k \geq 1$ and let $\mathcal{M}$ be a measure on $\Omega \times (\mathbb{R}^d)^k$ invariant under the diagonal maps $(\omega, z_1, \dots, z_k) \mapsto (\tau_z\omega, z_1 - z, \dots, z_k - z)$. For every invariant measurable $F \geq 0$, the measures $A \mapsto \int \mathbf{1}_A(z_i) F \, d\mathcal{M}$ are translation invariant on $\mathbb{R}^d$, and $\int \mathbf{1}_{[0,1)^d}(z_i) F \, d\mathcal{M} = \int \mathbf{1}_{[0,1)^d}(z_j) F \, d\mathcal{M}$ for all i, j.*

Proof. Translation invariance is the invariance of $\mathcal{M}$ applied to $\mathbf{1}_A(z_i)F$. For the second claim decompose $\mathbf{1}_{[0,1)^d}(z_i) = \sum_{\ell \in \mathbb{Z}^d} \mathbf{1}_{[0,1)^d}(z_i)\mathbf{1}_{\ell+[0,1)^d}(z_j)$ and apply the invariance with $z = -\ell$ to the ℓth term, which becomes $\int \mathbf{1}_{-\ell+[0,1)^d}(z_i)\mathbf{1}_{[0,1)^d}(z_j) F \, d\mathcal{M}$; summing over ℓ gives the claim. □

For an invariant subcoupling with $e(\mathbf{q}) < \infty$, Lemma 3.2 shows that $\int \mathbf{1}_\Lambda(x)|x - y|^p \, d\mathbf{q} = \int \mathbf{1}_\Lambda(y)|x - y|^p \, d\mathbf{q} = e(\mathbf{q})|\Lambda|$ for bounded Borel Λ, and that the defect measures $(m - \mathbf{q}_{\omega,x})(\Omega \times \cdot)$ and $(m' - \mathbf{q}_{\omega,y})(\Omega \times \cdot)$ are multiples of Lebesgue measure, with densities at most $\theta(\mathbf{q})$.

3.2. **Compactness.** For an open bounded $\Lambda \subset \mathbb{R}^d$ we regard $\mathbf{1}_{x \in \Lambda}\mathbf{q}$ as a finite measure on the Polish space $\Omega \times \overline{\Lambda} \times \mathbb{R}^d$ with the narrow topology. We say that $\mathbf{q}_n \to \mathbf{q}$ *locally* if this holds for every open cube Λ. The environment coordinate is never resampled: all measures have environment marginal dominated by $\mathbf{P}$.

Lemma 3.3. *Let $\mathbf{q}_n$ be invariant subcouplings with $\theta(\mathbf{q}_n) \to 0$ and $\sup_n e(\mathbf{q}_n) = E_0 < \infty$. Every subsequence has a further subsequence converging locally to an invariant coupling $\mathbf{q}$, and $e(\mathbf{q}) \leq \liminf e(\mathbf{q}_n)$ along it. Moreover, for every bounded Borel Λ and $R > 0$,*

$$\mathbf{q}_n\{x \in \Lambda, \ |x - y| > R\} + \mathbf{q}_n\{y \in \Lambda, \ |x - y| > R\} \leq 2|\Lambda|E_0 R^{-p}. \tag{3.1}$$

Proof. Step 1: tightness. Fix an open cube Λ. The (ω, x)-marginal of $\mathbf{1}_{x \in \Lambda}\mathbf{q}_n$ is dominated by the finite measure $\mathbf{P} \otimes \mathbf{1}_\Lambda dx$, which is tight. By Lemma 3.2, applied to $\mathbf{1}_{|x-y|>R}|x - y|^p$ with the indicator of Λ placed on either coordinate, each term in (3.1) is at most $|\Lambda|E_0 R^{-p}$. Since x is bounded, this controls the escape of y, and the restrictions are uniformly tight. Prokhorov's theorem and a diagonal argument over an exhaustion by cubes give a locally convergent subsequence. The limit $\mathbf{q}$ is a measure on $\Omega \times \mathbb{R}^d \times \mathbb{R}^d$; source domination ensures that it does not charge $\Omega \times \partial\Lambda \times \mathbb{R}^d$, so the restrictions are consistent. For test functions with unbounded y-support we first cut off $\{|x - y| > R\}$; by (3.1) the error is $O_\Lambda(R^{-p})$ uniformly in n.

Step 2: source marginal. The source marginals satisfy $0 \leq \mathbf{q}_{n,\omega,x} \leq m$ and, for every cube Λ,

$$(m - \mathbf{q}_{n,\omega,x})(\Omega \times \Lambda) \leq \theta(\mathbf{q}_n)|\Lambda| \to 0. \tag{3.2}$$

This total variation convergence identifies the source marginal of $\mathbf{q}$ as m; it is used again in the proof of Lemma 3.7.

Step 3: target marginal. Similarly $0 \leq \mathbf{q}_{n,\omega,y} \leq m'$ and $(m' - \mathbf{q}_{n,\omega,y})(\Omega \times \Lambda) \leq \theta(\mathbf{q}_n)|\Lambda|$. To pass to the limit under $(\omega, x, y) \mapsto (\omega, y)$ we must exclude mass with $y \in \Lambda$ and x far away; this is controlled by the second term in (3.1), which is where the invariance (mass transport) enters. The remaining edges start in the bounded R-neighbourhood of Λ, where local convergence applies. Testing with bounded continuous functions of (ω, y) supported in $\{y \in \Lambda\}$, multiplied by a continuous cutoff in $|x - y|/R$, and letting $n \to \infty$ and then $R \to \infty$, we find that the target marginal of $\mathbf{q}$ is m'.

Step 4: invariance and lower semicontinuity. Invariance under each Θ_z passes to local limits, by testing against bounded continuous functions with bounded x-support, which determine the restrictions $\mathbf{1}_{x \in \Lambda}\mathbf{q}$; this class is preserved by Θ_z, which is continuous because $(z, \omega) \mapsto \tau_z\omega$ is. (Ω is not locally compact, so compactly supported test functions would not suffice.) Hence $\mathbf{q}$ is an invariant coupling. Since $(x, y) \mapsto |x - y|^p \wedge M$ is

continuous and bounded, and $\partial[0,1)^d$ is not charged, $\int \mathbf{1}_{[0,1)^d}(x)(|x-y|^p \wedge M)\,d\mathbf{q} = \lim_n \int \mathbf{1}_{[0,1)^d}(x)(|x-y|^p \wedge M)\,d\mathbf{q}_n \le \liminf_n e(\mathbf{q}_n)$, and monotone convergence in M gives the lower semicontinuity. □

3.3. **Existence, uniqueness and structure of the optimizer.**

Proposition 3.4. *Assume (A1) and $e_\nu < \infty$, and if $p = 1$ assume that ν is almost surely purely atomic. Then:*

(i) an optimal invariant coupling exists;

(ii) every optimal invariant coupling $\mathbf{q} = \mathbf{P}(d\omega)q^\omega$ is induced by a map: for $\mathbf{P}$-a.e. ω, q^ω_x is a Dirac mass for a.e. x;

(iii) the optimal invariant coupling is unique, $\mathbf{q}_\nu = \mathbf{P}(d\omega)dx\,\delta_{T_\nu(\omega,x)}(dy)$, and T_ν can be chosen jointly measurable and exactly equivariant: $T_\nu(\omega,x) = x - \beta_\nu(\tau_x\omega)$ for a measurable $\beta_\nu : \Omega \to \mathbb{R}^d$;

(iv) the displacement $b_\nu(\omega,x) = x - T_\nu(\omega,x) = \beta_\nu(\tau_x\omega)$ satisfies $\mathbb{E}|\beta_\nu|^p = e_\nu$, and $h^{-d}\int_{[0,h)^d} b_\nu(\omega,x)\,dx \to \mathbb{E}[\beta_\nu \mid \mathcal{I}]$ almost surely and in L^p as $h \to \infty$ along the reals; in particular, under (1.8),

$$\mathcal{A}_p(h) := \mathbb{E}\Big|h^{-d}\int_{[0,h)^d} b_\nu(\omega,x)\,dx\Big|^p \to 0 \qquad (h\to\infty); \tag{3.3}$$

(v) $e_\nu > 0$ unless ν is almost surely Lebesgue measure.

Proof. (i) Take invariant couplings with $e(\mathbf{q}_n) \to e_\nu$ and apply Lemma 3.3 (with $\theta(\mathbf{q}_n) = 0$).

(ii) *Step 1: cyclical monotonicity.* Let $\mathbf{q}$ be optimal and $c(x,y) = |x-y|^p$. For $n \ge 2$ let

$$\Delta_n = \sum_{i=1}^n c(x_i,y_i) - \sum_{i=1}^n c(x_i,y_{i+1}) \qquad (y_{n+1} = y_1)$$

on $(\mathbb{R}^d \times \mathbb{R}^d)^n$. We show that for $\mathbf{P}$-a.e. ω, $\Delta_n \le 0$ holds $(q^\omega)^{\otimes n}$-a.e., for every n. If not, there are n and $R \in \mathbb{N}$ such that, with $\mathsf{B} = \{\Delta_n > 0,\ |x_i - x_1| < R\ \forall i\}$, the measure $\gamma^\omega = s\,(q^\omega)^{\otimes n}|_{\mathsf{B}}$, with $s = (n|B_{2R}|^{n-1})^{-1}$, is nonzero with positive probability; it depends measurably on ω. Let γ^ω_i be its image under $(x_j,y_j)_j \mapsto (x_i,y_i)$ and $\gamma^\omega_{i,i+1}$ its image under $(x_j,y_j)_j \mapsto (x_i,y_{i+1})$. On B all source points lie in $B_{2R}(x_i)$, and $q^\omega(\{x' \in A\}) = |A|$, so $\gamma^\omega_i \le s|B_{2R}|^{n-1}q^\omega = q^\omega/n$ setwise. Hence

$$\tilde q^\omega = q^\omega - \sum_{i=1}^n \gamma^\omega_i + \sum_{i=1}^n \gamma^\omega_{i,i+1} \ge 0,$$

and $\tilde q^\omega$ has the same marginals as q^ω, since $\gamma^\omega_{i,i+1}$ has the x-marginal of γ^ω_i and the y-marginal of γ^ω_{i+1}. As B depends only on differences of coordinates and $\mathbf{P}(d\omega)(q^\omega)^{\otimes n}$ is invariant under the diagonal maps, $\tilde{\mathbf{q}} = \mathbf{P}(d\omega)\tilde q^\omega$ is an invariant coupling. All energy terms are finite: on B, $c(x_i,y_{i+1}) \le 2^{p-1}((2R)^p + c(x_{i+1},y_{i+1}))$, and by Lemma 3.2 the integrals of $\mathbf{1}_{[0,1)^d}(x_i)c(x_j,y_j)$ against $(q^\omega)^{\otimes n}|_{\mathsf{B}}$ are at most $|B_{2R}|^{n-1}e(\mathbf{q})$. Moving each indicator $\mathbf{1}_{[0,1)^d}(x_i)$ onto x_1 by Lemma 3.2, term by term, gives

$$e(\tilde{\mathbf{q}}) - e(\mathbf{q}) = -\mathbb{E}\int \mathbf{1}_{[0,1)^d}(x_1)\,\Delta_n\,d\gamma^\omega < 0;$$

the strict inequality holds because $A \mapsto \mathbb{E}\int \mathbf{1}_A(x_1)\Delta_n\,d\gamma^\omega$ is a locally finite invariant measure, hence a multiple of Lebesgue measure, and it is nonzero. This contradicts optimality. Since n and R range over countable sets, Step 1 is proved.

Step 2: the case $p > 1$. Fix ω in the full-probability event of Step 1 for which q^ω couples dx and ω, and fix $r, K \in \mathbb{N}$. Let $\pi = \mathbf{1}_{\{|x|<r,\ |x-y|\le K\}}q^\omega$, a finite measure with bounded support whose first marginal is at most dx. Since $\pi \le q^\omega$, Step 1 gives $\Delta_n \le 0$ $\pi^{\otimes n}$-a.e. As $\{\Delta_n > 0\}$ is open, an n-tuple of points of $\operatorname{supp}\pi$ with $\Delta_n > 0$ would have

a neighbourhood of positive $\pi^{\otimes n}$-measure in $\{\Delta_n > 0\}$. Hence $\operatorname{supp}\pi$ is c-cyclically monotone (cyclic permutations suffice, since every permutation is a product of disjoint cycles), and π is optimal between its marginals [25, Chapter 5]. Since $c(x,y) = h(x-y)$ with $h = |\cdot|^p$ strictly convex and superlinear, and the first marginal of π is absolutely continuous, π is induced by a map, by the theorem of [9] [see also 25, Chapter 10]. By uniqueness of disintegrations, $q^\omega_x|_{\overline{B}_K(x)}$ is a multiple of a Dirac mass for a.e. x with $|x| < r$. If q^ω_x charged two distinct points, both would lie in $\overline{B}_K(x)$ for K large. Letting $K, r \to \infty$ along the integers, q^ω_x is a Dirac mass for a.e. x.

Step 3: the case $p = 1$. Let ω be as in Step 2 and purely atomic. Since $q^\omega(\mathbb{R}^d \times N) = \omega(N) = 0$ for the complement N of the countable set of atoms, q^ω_x is carried by the atoms for a.e. x. For distinct atoms $y_i \neq y_j$ let $h_k(x) = q^\omega_x(\{y_k\})$, $S_k = \{h_k > 0\}$ and $f(x) = c(x,y_i) - c(x,y_j)$. The restriction of $q^\omega \otimes q^\omega$ to $\{y = y_i, y' = y_j\}$ is $h_i(x)dx\, h_j(x')dx'$, and $\Delta_2 \le 0$ there means $f(x) \le f(x')$. Hence $f \le t := \operatorname{ess\,inf}_{S_j} f$ a.e. on S_i and $f \ge s := \operatorname{ess\,sup}_{S_i} f$ a.e. on S_j, while $f \le s$ a.e. on S_i and $f \ge t$ a.e. on S_j. Thus $f = s = t$ a.e. on $S_i \cap S_j$. The function f is real analytic on the connected open set $\mathbb{R}^d \setminus \{y_i, y_j\}$ ($d \ge 2$), and is not constant there, since by continuity it takes values close to $f(y_i) < 0$ and to $f(y_j) > 0$. The zero set of a real-analytic function that does not vanish identically on a connected open set has Lebesgue measure zero [22]; applied to $f - s$, this gives $|S_i \cap S_j| = 0$. Since there are countably many pairs of atoms, q^ω_x is a Dirac mass for a.e. x. (This argument works for every $p \ge 1$ when ν is atomic.)

(iii) If $\mathbf{q}, \mathbf{q}'$ are optimal, so is $(\mathbf{q} + \mathbf{q}')/2$, which by (ii) is induced by a map; since the average of two probability measures is a Dirac mass only if both are equal to it, $q^\omega_x = q'^\omega_x$ for a.e. (ω, x). Let $T_\nu(\omega, x)$ be the barycenter of q^ω_x where it exists (and x otherwise), a measurable map with $q^\omega_x = \delta_{T_\nu(\omega,x)}$ for a.e. (ω, x), and $b(\omega, x) = x - T_\nu(\omega, x)$. Invariance of $\mathbf{q}_\nu$ gives, for each z, $b(\tau_z\omega, x - z) = b(\omega, x)$ for m-a.e. (ω, x). Put $\tilde{b}(\omega, x) = b(\tau_{-x}\omega, x)$. Then, for each z, $\tilde{b}(\omega, x + z) = \tilde{b}(\omega, x)$ for m-a.e. (ω, x) (the map $(\omega, x) \mapsto (\tau_{-x}\omega, x)$ preserves m). By Fubini, for a.e. ω the function $\tilde{b}(\omega, \cdot)$ is a.e. equal to a constant $\beta_\nu(\omega)$, which can be chosen measurably (for instance as $\int_{[0,1)^d} \tilde{b}(\omega, x)\, dx$, since $\tilde{b}(\omega, \cdot) \in L^p_{\mathrm{loc}}$ a.s.: $\mathbb{E}\int_{[0,1)^d + k} |\tilde{b}|^p\, dx = e_\nu$ for every $k \in \mathbb{Z}^d$, by invariance of $\mathbf{q}$). Then $b(\omega, x) = \beta_\nu(\tau_x\omega)$ for m-a.e. (ω, x), and we redefine $T_\nu(\omega, x) = x - \beta_\nu(\tau_x\omega)$, which satisfies $T_\nu(\tau_z\omega, x - z) = T_\nu(\omega, x) - z$ for all ω, x, z.

(iv) $\mathbb{E}|\beta_\nu|^p = \mathbb{E}\int_{[0,1)^d} |b_\nu|^p dx = e_\nu$ by stationarity. The map $(\omega, x) \mapsto \beta_\nu(\tau_x\omega)$ is jointly measurable since the action is continuous, and $\beta_\nu \in L^p(\mathbf{P})$. The multiparameter ergodic theorem for the continuous-parameter action of $\mathbb{R}^d$ and the regular family of cubes $[0,h)^d$, $h \in (0,\infty)$ [see 18], gives $h^{-d}\int_{[0,h)^d} \beta_\nu(\tau_x\omega)\, dx \to \mathbb{E}[\beta_\nu \mid \mathcal{I}]$ almost surely and in L^1. The pth powers of these averages are bounded by the averages of $|\beta_\nu|^p$, which converge in L^1 and are therefore uniformly integrable; hence the convergence holds in L^p, and (3.3) follows under (1.8). (The σ-algebra $\mathcal{I}$ is the one of the action of $\mathbb{R}^d$; for the action of $\mathbb{Z}^d$ it may be larger.)

(v) If $e_\nu = 0$, an optimal invariant coupling is concentrated on $\{y = x\}$, so its second marginal ω equals its first marginal dx for $\mathbf{P}$-a.e. ω. □

Proposition 3.5. *Under the assumptions of Proposition 3.4, the centering condition (1.8) holds in each of the following cases:*

(a) $p = 2$;

(b) the law of ν is ergodic under translations and invariant under the reflection $x \mapsto -x$.

Proof. (a) Let $A \in \mathcal{I}$ and $v \in \mathbb{R}^d$, and put $T_{v,A}(\omega, x) = T_\nu(\omega, x - \mathbf{1}_A(\omega)v)$. Since $\mathbf{1}_A(\tau_z\omega) = \mathbf{1}_A(\omega)$, $T_{v,A}$ is exactly equivariant, and $(T_{v,A})_\# dx = (T_\nu)_\# dx = \omega$ because translations preserve dx. So $(\mathrm{id}, T_{v,A})_\# dx$ defines an invariant coupling. Its displacement is $\mathbf{1}_A v + \beta_\nu(\tau_{x - \mathbf{1}_A v}\omega)$, and since $\mathbb{E}[\mathbf{1}_A f(\tau_y\omega)] = \mathbb{E}[\mathbf{1}_A f]$ for $A \in \mathcal{I}$, its energy is $\mathbb{E}[\mathbf{1}_A|v +$

$\beta_\nu|^2] + \mathbb{E}[\mathbf{1}_{A^c}|\beta_\nu|^2]$. Optimality of T_ν gives $2v \cdot \mathbb{E}[\mathbf{1}_A \beta_\nu] + |v|^2 \mathbf{P}(A) \geq 0$ for every v, hence $\mathbb{E}[\mathbf{1}_A \beta_\nu] = 0$.

(b) Let $\mathcal{R}$ be the reflection $x \mapsto -x$, acting also on Ω. The image of $\mathbf{q}_\nu$ under $(\omega, x, y) \mapsto (\mathcal{R}\omega, -x, -y)$ is an invariant coupling (the law of ν is reflection invariant, and Θ_z is conjugated to Θ_{-z}) with the same energy, since $\int \mathbf{1}_{(-1,0]^d}(x)|x-y|^p \, d\mathbf{q}_\nu = e_\nu$. By uniqueness it equals $\mathbf{q}_\nu$, so $b_\nu(\mathcal{R}\omega, -x) = -b_\nu(\omega, x)$ a.e. Hence, using the reflection invariance of $\mathbf{P}$ and then stationarity, $\mathbb{E}\int_{[0,1)^d} b_\nu \, dx = -\mathbb{E}\int_{(-1,0]^d} b_\nu \, dx = -\mathbb{E}\int_{[0,1)^d} b_\nu \, dx$, so $\mathbb{E}\beta_\nu = 0$. By ergodicity, $\mathbb{E}[\beta_\nu \mid \mathcal{I}] = \mathbb{E}\beta_\nu = 0$. □

For $p \neq 2$, the variation in (a) only gives $\mathbb{E}[\mathbf{1}_A |\beta_\nu|^{p-2}\beta_\nu] = 0$, which does not imply (1.8).

Remark 3.6. Proposition 3.4 is a version, adapted to our setting, of known results. For an equivariant integer-valued point process and a cost $\vartheta(|x-y|)$ with ϑ continuous and strictly increasing, which includes the linear cost, Huesmann and Sturm [15, Theorem 1.1] prove existence and uniqueness of the optimal coupling of Lebesgue measure and the point process, and its representation by a factor map; finiteness of the cost for the Poisson process, for every $p > 0$ when $d \geq 3$, is [15, Corollary 1.4]. For general equivariant random measures on a manifold with a cocompact group of isometries, Huesmann [13, Theorems 1.1–1.3] prove existence, uniqueness, representation by a map and approximation by optimal semicouplings on bounded sets, under the standing assumption that the Monge problem between a compactly supported absolutely continuous measure and an arbitrary compactly supported measure has a unique solution induced by a map [13, Section 2.10]. For $p > 1$ this holds for Euclidean power costs; for finitely many atoms with arbitrary positive masses it holds for every continuous strictly increasing ϑ, in particular for $p = 1$ [13, Lemma 3.2], which is [15, Lemma 6.1]. These works use equivariance with respect to a lattice (or a discrete group), and Huesmann and Sturm [15] require integer masses. We include the short proof above because we need real masses, full translation invariance and the joint-law formulation of Definition 3.1. For $p = 1$ and diffuse ν the Monge problem need not have a unique solution, and we do not know whether the optimal stationary coupling is unique. Finiteness of e_ν is proved in Corollary 4.2.

3.4. **Strong convergence to the optimal graph.** The next lemma is the rigidity statement behind the strong screening of Section 4: invariant subcouplings with vanishing defects and asymptotically optimal energy converge to the graph of T_ν *in the original environment*. Keeping the environment coordinate in the joint law is essential, since convergence in law of the edges alone could not be compared with $T_\nu(\omega, x)$.

The passage from vague convergence of couplings to convergence in measure of the transport maps is [15, Lemma 4.7], and for $p = 2$ the vague convergence of stationarized optimizers of the finite-volume problems to the stationary optimizer is [16, Corollary 1.3]; the version below keeps the environment and upgrades to L^p convergence using the convergence of the energies.

Lemma 3.7. *Assume $e_\nu < \infty$. Let $\mathbf{q}_n$ be invariant subcouplings with $\theta(\mathbf{q}_n) \to 0$, $\sup_n e(\mathbf{q}_n) < \infty$ and $\limsup_n e(\mathbf{q}_n) \leq e_\nu$. Then, with the unique optimal invariant coupling $\mathbf{q}_\nu = \mathbf{P}(d\omega)dx\,\delta_{T_\nu(\omega,x)}(dy)$ of Proposition 3.4,*

$$e(\mathbf{q}_n) \to e_\nu, \qquad \int \mathbf{1}_{[0,1)^d}(x)|y - T_\nu(\omega, x)|^p \, d\mathbf{q}_n \to 0. \tag{3.4}$$

Proof. By Lemma 3.3, every subsequence of $(\mathbf{q}_n)$ has a locally convergent further subsequence, and every local limit point $\mathbf{q}$ is an invariant coupling with $e(\mathbf{q}) \leq e_\nu$, hence optimal, hence $\mathbf{q} = \mathbf{q}_\nu$. So the whole sequence converges locally to $\mathbf{q}_\nu$, and along a subsequence realizing $\liminf e(\mathbf{q}_n)$ we get $e_\nu \leq \liminf e(\mathbf{q}_n) \leq \limsup e(\mathbf{q}_n) \leq e_\nu$, so

$e(\mathbf{q}_n) \to e_\nu$. Fix $\Lambda = (0,1)^d$ and $m_\Lambda = \mathbf{P} \otimes \mathbf{1}_\Lambda dx$. Given $\varepsilon > 0$, Lusin's theorem gives a compact $\mathcal{K}_\varepsilon \subset \Omega \times \Lambda$ with $m_\Lambda(\mathcal{K}_\varepsilon^c) < \varepsilon$ on which T_ν is continuous; by Tietze's theorem there is a bounded continuous f on $\Omega \times \overline{\Lambda}$ equal to T_ν on $\mathcal{K}_\varepsilon$. By source domination the $\mathbf{q}_n$-mass of $\{(\omega,x) \notin \mathcal{K}_\varepsilon\}$ is less than ε. Testing local convergence with $1 \wedge (|y - f(\omega,x)|/s)$, whose integral against $\mathbf{q}_\nu$ is at most ε, gives $\limsup_n \mathbf{q}_n\{x \in \Lambda,\ |y - T_\nu(\omega,x)| > s\} \le 2\varepsilon$ for every $s > 0$: convergence in measure to the graph of T_ν, in the original environment. For the upgrade to L^p, local convergence gives $\int_{x\in\Lambda}(|x-y|^p \wedge M)\, d\mathbf{q}_n \to \int_{x\in\Lambda}(|x-y|^p \wedge M)\, d\mathbf{q}_\nu$ while the full integrals converge to e_ν; hence $\lim_{M\to\infty} \limsup_n \int_{x\in\Lambda}(|x-y|^p - M)_+\, d\mathbf{q}_n = 0$, i.e. $|x-y|^p$ is uniformly integrable with respect to $\mathbf{1}_{x\in\Lambda}\mathbf{q}_n$. By source domination, $\int_{x\in\Lambda} |x - T_\nu|^p \mathbf{1}_{\{|x-T_\nu|^p > M\}}\, d\mathbf{q}_n \le \int |x - T_\nu|^p \mathbf{1}_{\{|x-T_\nu|^p > M\}}\, dm_\Lambda$, which tends to zero as $M \to \infty$. Thus $G = |y - T_\nu|^p \le 2^{p-1}(|y-x|^p + |x - T_\nu|^p)$ is uniformly integrable with respect to $\mathbf{1}_{x\in\Lambda}\mathbf{q}_n$. Since these measures have mass at most one, $\int_{x\in\Lambda} G\, d\mathbf{q}_n \le s^p + M\, \mathbf{q}_n\{x \in \Lambda,\ G > s^p\} + \int_{x\in\Lambda} G\mathbf{1}_{\{G>M\}}\, d\mathbf{q}_n$. For fixed s and M, $\mathbf{q}_n\{x \in \Lambda,\ G > s^p\} = \mathbf{q}_n\{x \in \Lambda,\ |y - T_\nu| > s\} \to 0$ by the convergence in measure just proved, so $\limsup_n \int_{x\in\Lambda} G\, d\mathbf{q}_n \le s^p + \sup_n \int_{x\in\Lambda} G\mathbf{1}_{\{G>M\}}\, d\mathbf{q}_n$; letting $M \to \infty$, then $s \to 0$ gives (3.4). □

4. Stationarization of boundary optimizers and strong screening

We now return to the cube. The direct part of a boundary-optimal plan is turned into a stationary subcoupling by tiling space with copies of the cube in the same environment. This gives the lower bound $e_\nu \le \liminf L^{-d}\mathbb{E}B_L$; the matching upper bound is obtained by cutting the edges of the stationary optimizer at the boundary. Lemma 3.7 then shows that the direct edges of boundary-optimal plans are close to the optimal stationary map.

From now on, ν satisfies (A1)–(A2), and if $p = 1$ it is almost surely purely atomic, so that Proposition 3.4 applies once $e_\nu < \infty$ is known (Corollary 4.2). The centering condition (1.8) is used only in Section 5 and in Proposition 7.2.

4.1. **The direct-edge displacement field.** Select, measurably in ν_L, a $c_{b,p,L}$-optimal coupling π_L of λ_L and ν_L (zero on $\{\Sigma_L = 0\}$); this is possible since $c_{b,p,L}$ is continuous on $\overline{Q}_L^2$ [25, Corollary 5.22], applied to the probability measures λ_L/Σ_L and ν_L/Σ_L on $\overline{Q}_L$ (an optimal plan does not charge ∂Q_L, since λ_L does not). All the objects below are then jointly measurable in the environment and the space variables (for instance $(c,\omega) \mapsto \tau_c\omega$ is continuous), which justifies the uses of Fubini's theorem. On $\{\Sigma_L > 0\}$, disintegrate

$$\pi_L(dx, dy) = a_L dx\, \pi_{L,x}(dy). \tag{4.1}$$

Pairs for which the two branches of $c_{b,p,L}$ coincide are assigned to the direct region:

$$\mathsf{D}_L = \{(x,y) : |x-y|^p \le \rho_L(x)^p + \rho_L(y)^p\}, \qquad \mathsf{S}_L = Q_L^2 \setminus \mathsf{D}_L, \tag{4.2}$$

and we write $\mathsf{D}_{L,x} = \{y : (x,y) \in \mathsf{D}_L\}$, $\mathsf{S}_{L,x} = \{y : (x,y) \in \mathsf{S}_L\}$. The pairs in S_L are the strict shortcuts. The shortcut energy, the shortcut mass, the direct fraction and the barycentric displacement of the direct edges are

$$K_L = \int_{\mathsf{S}_L} (\rho_L(x)^p + \rho_L(y)^p)\, d\pi_L, \qquad M_L^b = \pi_L(\mathsf{S}_L), \tag{4.3}$$

$$w_L(x) = \pi_{L,x}(\mathsf{D}_{L,x}), \qquad g_L(x) = \int_{\mathsf{D}_{L,x}} (x-y)\, \pi_{L,x}(dy), \tag{4.4}$$

with $g_L = w_L = 0$ on $\{\Sigma_L = 0\}$. By Jensen's inequality,

$$|g_L(x)| \le \sqrt{d}L, \qquad \int_{Q_L} |g_L|^p dx \le a_L^{-1} \int_{\mathsf{D}_L} |x-y|^p d\pi_L. \tag{4.5}$$

4.2. **Clipping and stationarization.** Fix $L > 1$ and put

$$\alpha_L = \begin{cases} \min(1, a_L^{-1}) & \Sigma_L > 0, \\ 1 & \Sigma_L = 0. \end{cases}$$

The plan $\alpha_L \mathbf{1}_{\mathsf{D}_L} \pi_L$ has first marginal at most $dx|_{Q_L}$ and second marginal at most ν_L. Since $L^d - \alpha_L \Sigma_L = (L^d - \Sigma_L)_+$ and $\Sigma_L - \alpha_L \Sigma_L = (\Sigma_L - L^d)_+$, the source and target masses missing from it are

$$\Delta_x = (L^d - \Sigma_L)_+ + \alpha_L M_L^b, \qquad \Delta_y = (\Sigma_L - L^d)_+ + \alpha_L M_L^b. \tag{4.6}$$

If $\rho_L(x) \le R$, the source density of $\alpha_L \pi_L$ is at most one, and if $\rho_L(x) > R$, every shortcut leaving x costs more than R^p. Hence

$$\alpha_L M_L^b \le 2dRL^{d-1} + R^{-p} B_L. \tag{4.7}$$

Lemma 4.1. *There is an invariant subcoupling $\mathbf{q}_L$ such that, for every nonnegative measurable F on $\Omega \times \mathbb{R}^d \times \mathbb{R}^d$ invariant under all Θ_z,*

$$\int \mathbf{1}_{[0,1)^d}(x) F \, d\mathbf{q}_L = L^{-d} \mathbb{E}\Big[\alpha_L \int_{\mathsf{D}_L} F(\nu, x, y) \, d\pi_L(x, y) \Big]. \tag{4.8}$$

In particular $e(\mathbf{q}_L) = L^{-d} \mathbb{E}[\alpha_L \int_{\mathsf{D}_L} |x - y|^p d\pi_L]$,

$$L^{-d} \mathbb{E} B_L - e(\mathbf{q}_L) = L^{-d} \mathbb{E}\Big[K_L + (1 - \alpha_L) \int_{\mathsf{D}_L} |x - y|^p \, d\pi_L \Big] \ge 0, \tag{4.9}$$

and, for every $R \ge 1$,

$$\theta(\mathbf{q}_L) \le L^{-d} \mathbb{E} |\Sigma_L - L^d| + 2d \frac{R}{L} + \frac{\mathbb{E} B_L}{L^d R^p}. \tag{4.10}$$

Proof. For a measure $\omega \in \Omega$ let $P_L(\omega)$ denote the plan $\alpha_L \mathbf{1}_{\mathsf{D}_L} \pi_L$ computed from $\omega|_{Q_L}$. For $\varsigma \in [0, L)^d$, let $q^\omega_{L,\varsigma}$ be the sum over $k \in \mathbb{Z}^d$ of the images of $P_L(\tau_{\varsigma + Lk}\omega)$ under $(x, y) \mapsto (x + \varsigma + Lk, y + \varsigma + Lk)$: on each cell $\varsigma + Lk + Q_L$ we apply the clipped optimizer to the restriction of the *same* environment to that cell, in local coordinates. Put $q^\omega_L = L^{-d} \int_{[0,L)^d} q^\omega_{L,\varsigma} \, d\varsigma$ and $\mathbf{q}_L = \mathbf{P}(d\omega) q^\omega_L$. The marginal bounds hold for each ς, hence after averaging. Since $\tau_{\varsigma + Lk} \tau_z = \tau_{\varsigma + z + Lk}$, the kernel for $\tau_z \omega$ with offset ς is the translate by $-z$ of the kernel for ω with offset $\varsigma + z$; as $q^\omega_{L,\varsigma}$ is L-periodic in ς, averaging gives exact equivariance of q^ω_L, so $\mathbf{q}_L$ is invariant.

For invariant $F \ge 0$, $F(\omega, x' + c, y' + c) = F(\tau_c \omega, x', y')$, so

$$\int \mathbf{1}_{[0,1)^d}(x) F \, d\mathbf{q}_L = L^{-d} \int_{\mathbb{R}^d} dc \, \mathbb{E} \int \mathbf{1}_{[0,1)^d}(x' + c) F(\tau_c \omega, x', y') \, dP_L(\tau_c \omega)(x', y'),$$

where we combined $\int_{[0,L)^d} d\varsigma$ and $\sum_k$ into $\int_{\mathbb{R}^d} dc$. By stationarity we may replace $\tau_c \omega$ by ω, and then $\int_{\mathbb{R}^d} \mathbf{1}_{[0,1)^d}(x' + c) \, dc = 1$ gives (4.8). With $F = |x - y|^p$ we get the expression for $e(\mathbf{q}_L)$, and splitting $B_L = \int_{\mathsf{D}_L} |x - y|^p d\pi_L + K_L$ gives (4.9). With $F = 1$, the retained source intensity is $L^{-d} \mathbb{E}[\alpha_L (\Sigma_L - M_L^b)]$, so the source defect is $L^{-d} \mathbb{E} \Delta_x$. By Lemma 3.2 with $F = 1$, the retained target intensity is the same, and since $\mathbb{E}\nu([0,1)^d) = 1$ the target defect equals the source defect (equivalently, $\mathbb{E}\Delta_x = \mathbb{E}\Delta_y$ because $\mathbb{E}\Sigma_L = L^d$). Now (4.6) and (4.7) give (4.10). □

By (2.1), (2.2) and the choice $R = L^{1/(p+1)}$,

$$\theta(\mathbf{q}_L) \le C\big(L^{(\alpha - d)/2} + L^{-p/(p+1)}\big) \to 0. \tag{4.11}$$

Corollary 4.2. $e_\nu \le \liminf_{L \to \infty} L^{-d} \mathbb{E} B_L < \infty$. *Consequently the conclusions of Proposition 3.4 hold.*

Proof. By (4.9) and (2.2), $\sup_L e(\mathbf{q}_L) < \infty$, and $\theta(\mathbf{q}_L) \to 0$ by (4.11). Along a sequence realizing the lower limit of $L^{-d}\mathbb{E}B_L$, Lemma 3.3 gives an invariant coupling $\mathbf{q}$ with $e_\nu \le e(\mathbf{q}) \le \liminf e(\mathbf{q}_L) \le \liminf L^{-d}\mathbb{E}B_L$. □

4.3. **The limit of the boundary cost.** The limit of the specific boundary cost is the optimal stationary cost. For jointly equivariant random measures and the unbalanced cost along integer side lengths this is [16, Theorem 2.2]; the lower bound is Corollary 4.2, which follows their stationarization [16, Proposition 3.2]. For the upper bound we use the following exact inequality, obtained by cutting the edges of a stationary coupling at the boundary; it is a variant of [16, Proposition 3.1 and Lemma 3.3] without error term.

Lemma 4.3. *Let $\mathbf{q} = \mathbf{P}(d\omega)q^\omega$ be an invariant coupling with $e(\mathbf{q}) < \infty$, and let $Q \subset \mathbb{R}^d$ be a bounded open convex set. Then*

$$\mathbb{E}\, W_Q^{b,p}(dx|_Q, \nu|_Q) \le e(\mathbf{q})\, |Q|.$$

Proof. Write $\rho_Q = \operatorname{dist}(\cdot, Q^c)$; the cost (2.6) on Q is defined as for Q_L. Fix ω such that q^ω couples dx and ω. For an edge (x, y) let $z_t = (1-t)x + ty$ and let $\ell(x,y) = |x-y| \int_0^1 \mathbf{1}_Q(z_t)\, dt$ be the length of $[x,y] \cap Q$, an interval since Q is convex. Map x to itself if $x \in Q$ and to the point ∂ otherwise, and likewise y, and let γ be the image of q^ω restricted to $\{x \in Q\} \cup \{y \in Q\}$. On Q the marginals of γ are $dx|_Q$ and $\nu|_Q$, and the two marginals of γ have the same total mass, so γ is admissible in (2.6) (which allows different masses s, t at ∂). An edge with both endpoints in Q costs at most $|x-y|^p = \ell^p$. If $x \in Q$ and $y \notin Q$, the segment leaves Q after length ℓ, so the edge (x, ∂) costs $\rho_Q(x)^p \le \ell^p$; the case $x \notin Q$, $y \in Q$ is symmetric. Since $\ell \le |x-y|$ and $p \ge 1$,

$$W_Q^{b,p}(dx|_Q, \nu|_Q) \le \int \ell(x,y)^p\, dq^\omega \le \int_0^1 \!\!\int |x-y|^p \mathbf{1}_Q(z_t)\, dq^\omega\, dt.$$

For fixed t, the image of $\mathbf{q}$ under $(\omega, x, y) \mapsto (\omega, x, y, z_t)$ is invariant under the diagonal maps of Lemma 3.2, since z_t commutes with translations, and $|x-y|^p$ is invariant. By Lemma 3.2, applied with the coordinates z_t and x, $\mathbb{E}\int |x-y|^p \mathbf{1}_Q(z_t)\, dq^\omega = |Q|\, e(\mathbf{q})$. Integrating in t proves the lemma. □

Proposition 4.4. *Along all real $L \to \infty$,*

$$\lim_{L\to\infty} L^{-d}\mathbb{E}B_L = e_\nu. \tag{4.12}$$

Proof. The lower bound is Corollary 4.2. By Lemma 2.4(i)–(iii), using the symmetry of $W_{Q_L}^{b,p}$ and the intermediate measure $dx|_{Q_L}$,

$$B_L \le (1+\varepsilon)^{p-1} W_{Q_L}^{b,p}(dx|_{Q_L}, \nu_L) + (1+\varepsilon^{-1})^{p-1} C_{d,p} L^{d+p} |a_L - 1|^p$$

for every $\varepsilon \in (0, 1]$. By Lemma 4.3 with $\mathbf{q} = \mathbf{q}_\nu$ and $Q = Q_L$, the expectation of the first term is at most $(1+\varepsilon)^{p-1} e_\nu L^d$, and $\mathbb{E}|a_L - 1|^p \le C L^{-p(d-\alpha)/2}$ by (2.1). Hence $\limsup_L L^{-d}\mathbb{E}B_L \le (1+\varepsilon)^{p-1} e_\nu + C_\varepsilon \limsup_L L^{p(2+\alpha-d)/2}$, and the last term vanishes since $\alpha < d-2$. Let $\varepsilon \to 0$. □

Remark 4.5. For the Poisson process, (4.12) along integers follows from [16, Theorem 1.2 and Proposition 4.1], and the existence of the limit along all real L is [4, Proposition A.3]. One could also deduce (4.12) from [16, Theorem 2.2], the balancing estimate above, and monotonicity of the unbalanced cost under restriction to a subcube for real L, after identifying the class of equivariant couplings of [16] with that of Definition 3.1. Lemma 4.3 avoids this identification and gives all real L directly.

4.4. **Strong screening.**

Theorem 4.6. *Along all real $L \to \infty$,*

$$L^{-d}\mathbb{E}K_L \longrightarrow 0, \tag{4.13}$$

$$L^{-d}\mathbb{E}\int_{Q_L} |g_L - b_\nu|^p dx \longrightarrow 0, \tag{4.14}$$

and the direct edges are uniformly integrable:

$$\lim_{M\to\infty} \limsup_{L\to\infty} L^{-d}\mathbb{E}\int_{\mathsf{D}_L} |x-y|^p \mathbf{1}_{\{|x-y|>M\}} d\pi_L = 0. \tag{4.15}$$

Proof. By (4.9), (4.11) and (4.12), the invariant subcouplings $\mathbf{q}_L$ satisfy the hypotheses of Lemma 3.7. Hence

$$e(\mathbf{q}_L) \to e_\nu, \qquad \gamma_L := \int \mathbf{1}_{[0,1)^d}(x)|y - T_\nu(\omega,x)|^p \, d\mathbf{q}_L \to 0. \tag{4.16}$$

Subtracting $e(\mathbf{q}_L)$ from $L^{-d}\mathbb{E}B_L$ in (4.9) proves (4.13). By (4.7), letting $L \to \infty$ at fixed R and then $R \to \infty$,

$$L^{-d}\mathbb{E}[\alpha_L M_L^b] \to 0. \tag{4.17}$$

Let $G_L = \{1/2 \le a_L \le 2\}$; on G_L, $a_L^{-1} \le 2\alpha_L$ and $\alpha_L \ge 1/2$. Since $x - b_\nu(\omega,x) = T_\nu(\omega,x)$, $g_L - b_\nu = \int_{\mathsf{D}_{L,x}} (T_\nu - y)\,\pi_{L,x}(dy) - (1-w_L)b_\nu$, and Jensen's inequality, $0 \le w_L \le 1$ and $(1-w_L)^p \le 1 - w_L$ give

$$|g_L - b_\nu|^p \le 2^{p-1}\Big[\int_{\mathsf{D}_{L,x}} |y - T_\nu(\omega,x)|^p \pi_{L,x}(dy) + (1-w_L)|b_\nu|^p\Big]. \tag{4.18}$$

The function $|y - T_\nu(\omega,x)|^p$ is invariant, since T_ν is exactly equivariant, and it is evaluated in the global environment. By (4.8),

$$L^{-d}\mathbb{E}\Big[\mathbf{1}_{G_L}\int_{Q_L}\int_{\mathsf{D}_{L,x}} |y - T_\nu|^p \pi_{L,x}(dy)dx\Big] \le 2\gamma_L \to 0. \tag{4.19}$$

Since $\int_{Q_L}(1-w_L)\,dx = a_L^{-1}M_L^b$,

$$L^{-d}\mathbb{E}\Big[\mathbf{1}_{G_L}\int_{Q_L}(1-w_L)\,dx\Big] \le 2L^{-d}\mathbb{E}[\alpha_L M_L^b] \to 0. \tag{4.20}$$

For every $H > 0$ and every random weight $0 \le v(\omega,x) \le 1$, stationarity gives

$$L^{-d}\mathbb{E}\int_{Q_L} v|b_\nu|^p dx \le HL^{-d}\mathbb{E}\int_{Q_L} v\,dx + \mathbb{E}\int_{[0,1)^d} |b_\nu|^p \mathbf{1}_{\{|b_\nu|^p > H\}} dx, \tag{4.21}$$

without any independence between v and b_ν. We apply it with $v = \mathbf{1}_{G_L}(1-w_L)$ and let $L \to \infty$, then $H \to \infty$. On G_L^c we use $|g_L| \le \sqrt{d}L$, $\mathbb{P}(G_L^c) \le C_k L^{-k}$ for every k (Lemma 2.1) and (4.21) with $v = \mathbf{1}_{G_L^c}$. This proves (4.14).

For (4.15), write $z = x - y = b_\nu(\omega,x) + (T_\nu(\omega,x) - y)$ along a direct edge. For $M > 0$,

$$|z|^p \mathbf{1}_{\{|z|>M\}} \le 2^p\big[|b_\nu|^p \mathbf{1}_{\{|b_\nu|>M/2\}} + |y - T_\nu|^p\big]. \tag{4.22}$$

On G_L, integrating against $\mathbf{1}_{\mathsf{D}_L}\pi_L$ and normalizing by L^d, the first term contributes at most $2^{p+1}\mathbb{E}\int_{[0,1)^d}|b_\nu|^p\mathbf{1}_{\{|b_\nu|>M/2\}}dx$, by source domination and $a_L \le 2$, and the second at most $2^{p+1}\gamma_L$, by (4.8) and $\alpha_L \ge 1/2$. On G_L^c, $\int |x-y|^p d\pi_L \le d^{p/2}L^p\Sigma_L$ and $L^{p-d}\mathbb{E}[\Sigma_L \mathbf{1}_{G_L^c}] \le L^{p-d}(\mathbb{E}\Sigma_L^2)^{1/2}\mathbb{P}(G_L^c)^{1/2} \le CL^p\mathbb{P}(G_L^c)^{1/2} \to 0$ by Lemma 2.1. Let $L \to \infty$ and then $M \to \infty$. □

Remark 4.7. The uniform integrability (4.15) of the direct-edge energies is related to the higher integrability estimates for optimal displacements obtained by [10, Theorem 1.2] for periodic bipartite matching with $p \geq 2$ and $d \geq 3$: for every $q \geq 1$, the qth moment of the displacement, per unit volume, is bounded by $C(\log L)^{q(1+1/d)}$. An analogous estimate with $q = 2p$ for boundary-optimal plans, even with this logarithmic loss, would give $L^{-d}\mathbb{E}J_L \lesssim L^{-p}(\log L)^{2p(1+1/d)}$ for the Taylor remainder J_L of Lemma 6.3, but a logarithmically growing bound would not in itself imply (4.15). Here uniform integrability follows instead from the convergence of the specific energies and the uniqueness of the stationary optimizer, for every $p \geq 1$, without higher integrability or large-scale regularity. It is qualitative and gives neither higher moments nor rates.

5. A mesoscopic Poincaré inequality and flux cancellation

For $f \in L^p(Q_L)$ with zero mean set

$$\|f\|_{W^{-1,p}_{\#}(Q_L)} = \sup\Big\{\Big|\int_{Q_L} f\varphi\,dx\Big| : \varphi \in W^{1,p'}(Q_L),\ \int_{Q_L}\varphi\,dx = 0,\ \|\nabla\varphi\|_{L^{p'}(Q_L)} \leq 1\Big\}. \tag{5.1}$$

For $p = 1$ this is the dual Lipschitz seminorm on the convex cube. The supremum may be restricted to a countable family (a dense subset of $W^{1,p'}$ if $p' < \infty$, and, by the Arzelà–Ascoli theorem, of the 1-Lipschitz functions in $C(\overline{Q}_L)$ if $p' = \infty$), so the norms of random densities below are measurable. For general $f \in L^p$ we apply the norm to $f - \fint_{Q_L} f$; since test functions have zero mean, this does not change the pairing. The Poincaré inequality $\|\varphi\|_{L^{p'}(Q_L)} \leq C_{d,p}L\|\nabla\varphi\|_{L^{p'}(Q_L)}$ for mean-zero φ, valid also for $p' = \infty$, gives

$$\Big\|f - \fint_{Q_L} f\Big\|_{W^{-1,p}_{\#}} \leq C_{d,p}L\|f\|_{L^p(Q_L)}. \tag{5.2}$$

Lemma 5.1. *Let $\eta \in C^2([0,1]^d)$ and $\eta_L(x) = \eta(x/L)$. Let $1 \leq h \leq L$ with $L/h \in \mathbb{N}$, tile Q_L by the $(L/h)^d$ cubes $h(k + [0,1)^d)$, $k \in \{0,\dots,L/h-1\}^d$, and let $P_h b$ be the function equal on each of these cubes to the average of b over it. For every $b \in L^p(Q_L;\mathbb{R}^d)$,*

$$\Big\|b\cdot\nabla\eta_L - \fint_{Q_L} b\cdot\nabla\eta_L\Big\|_{W^{-1,p}_{\#}} \leq C_{d,p,\eta}\Big[\frac{h}{L}\|b\|_p + \|P_h b\|_p\Big]. \tag{5.3}$$

Proof. Let φ have zero mean and $\|\nabla\varphi\|_{p'} \leq 1$, so $\|\varphi\|_{p'} \leq CL$, and put $\psi = \varphi\nabla\eta_L$, so that $|\nabla\psi| \leq C(L^{-2}|\varphi| + L^{-1}|\nabla\varphi|)$ and $\|\nabla\psi\|_{p'} \leq C/L$. On each cube Q' of the tiling, $\int_{Q'} b\cdot\psi = \int_{Q'}(b - (b)_{Q'})\cdot(\psi - (\psi)_{Q'}) + \int_{Q'} P_h b\cdot\psi$. By Hölder's inequality and the Poincaré inequality on Q', the first term is at most $Ch\|b - (b)_{Q'}\|_{L^p(Q')}\|\nabla\psi\|_{L^{p'}(Q')}$ (for $p' = \infty$ this is the oscillation bound for the Lipschitz function ψ). Summing over the cubes and using $\|b - P_h b\|_p \leq 2\|b\|_p$, the first terms contribute at most $C(h/L)\|b\|_p$. The second terms contribute at most $\|P_h b\|_p\|\nabla\eta_L\|_\infty\|\varphi\|_{p'} \leq C\|P_h b\|_p$. Since φ has zero mean, subtracting the mean of $b\cdot\nabla\eta_L$ does not change the pairing. □

Proposition 5.2. *Assume (1.8). For every $\eta \in C^2([0,1]^d)$, with $\eta_L(x) = \eta(x/L)$, along all real $L \to \infty$,*

$$L^{-d}\mathbb{E}\Big\|g_L\cdot\nabla\eta_L - \fint_{Q_L} g_L\cdot\nabla\eta_L\Big\|^p_{W^{-1,p}_{\#}} \longrightarrow 0. \tag{5.4}$$

Here $\nabla\eta_L(x) = L^{-1}(\nabla\eta)(x/L)$.

Proof. For $L \geq 1$ let $N_L = \lfloor\sqrt{L}\rfloor$ and $h_L = L/N_L$, so that $L/h_L \in \mathbb{N}$, $h_L \to \infty$ and $h_L/L \to 0$. Since P_{h_L} is a contraction in L^p, Minkowski's inequality in $L^p(\Omega\times Q_L)$ gives

$$L^{-d/p}\big(\mathbb{E}\|P_{h_L}g_L\|_p^p\big)^{1/p} \leq L^{-d/p}\big(\mathbb{E}\|g_L - b_\nu\|_p^p\big)^{1/p} + \mathcal{A}_p(h_L)^{1/p},$$

where we used that, by stationarity of b_ν, each of the N_L^d cubes of the tiling contributes $h_L^d \mathcal{A}_p(h_L)$ to $\mathbb{E}\|P_{h_L} b_\nu\|_p^p$. The first term tends to zero by (4.14) and the second by (3.3), since $h_L \to \infty$. Moreover $\sup_{L \geq L_0} L^{-d}\mathbb{E}\|g_L\|_p^p < \infty$ for some L_0, by (4.14) and $\mathbb{E}|\beta_\nu|^p = e_\nu$, so $(h_L/L)^p L^{-d}\mathbb{E}\|g_L\|_p^p \to 0$. Raising (5.3) with $h = h_L$ to the power p and taking expectations proves (5.4). □

Remark 5.3. For the layer cutoff $\eta = \eta_{\delta,L}$ and a constant vector v, $v \cdot \nabla\eta$ has zero mean. If $v \neq 0$, the affine test function $L^{-d/p'}(v/|v|) \cdot (x - x_0)$, where x_0 is the centre of the cube, has gradient norm one in (5.1). Since $\eta = 1$ on ∂Q_L, integration by parts with $\eta - 1$ gives

$$L^{-d/p}\|v \cdot \nabla\eta\|_{W_{\#}^{-1,p}} \geq |v|(1 - L^{-d}V_\eta).$$

Thus shrinking the layer does not make the contribution of a constant displacement small in this norm.

6. The deterministic comparison

In this section L and the measures are fixed; nothing is random, and the target is an arbitrary finite measure. The only external input is the following transport inequality.

Lemma 6.1. *Let $p \geq 1$, $a \in [1/2, 2]$, $\lambda = a\mathbf{1}_{Q_L} dx$, and let $\nu' = f\,dx$ be a measure on Q_L with bounded density and $\nu'(Q_L) = \lambda(Q_L)$. Then*

$$W_p^p(\nu', \lambda) \leq C_{p,d}\|f - a\|_{W_{\#}^{-1,p}(Q_L)}^p,$$

and one can take $C_{1,d} = 1$.

Proof. For $p = 1$ this is Kantorovich–Rubinstein duality: $W_1(\nu', \lambda)$ is the supremum of $\int \varphi(f - a)\,dx$ over 1-Lipschitz φ; one may subtract the mean of φ, and on the convex cube the 1-Lipschitz functions are those with $\|\nabla\varphi\|_\infty \leq 1$. For $p > 1$ this is [14, Lemma 3.6], applied on the Lipschitz domain Q_1 with weight identically one (for which the weighted Poincaré inequality is the usual one) and lower bound $1/2$ on the density of λ (denoted g_0 there); the negative norm of [14, (2.10)] coincides with (5.1) for zero-mean densities. Under the dilation $x = Ly$, both W_p^p and $\|\cdot\|_{W_{\#}^{-1,p}}^p$ are multiplied by L^{d+p}, while the densities are unchanged, so the constant does not depend on L. □

Fix $0 < \delta \leq 1/16$ and $r = \delta L$. In this section η denotes a cutoff on Q_L (in the notation of Section 5, $\eta_{\delta,L} = (\eta_{\delta,1})_L$). We use the cutoff of [14, (4.3)–(4.4)]: with $f_0(t) = 10t^3 - 15t^4 + 6t^5$, let $\xi_\delta : \mathbb{R} \to [0,1]$ be even, equal to 1 on $[0, 1/2 - 2\delta]$, to $1 - f_0((t - 1/2 + 2\delta)/\delta)$ on $[1/2 - 2\delta, 1/2 - \delta]$, and to 0 on $[1/2 - \delta, \infty)$, and let

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

Then $\eta = \eta_{\delta,L}$ satisfies

$$\begin{gathered} 0 \leq \eta \leq 1, \quad \eta = 1 \text{ on } \{\rho_L \leq r\}, \quad \eta = 0 \text{ on } \{\rho_L \geq 2r\}, \\ \|\nabla\eta\|_\infty \leq C_d/r, \quad \|D^2\eta\|_\infty \leq C_d/r^2, \quad c_d\delta L^d \leq V_\eta := \int_{Q_L} \eta\,dx \leq C_d\delta L^d. \end{gathered} \tag{6.1}$$

For a finite measure μ on Q_L set

$$S_\eta = \int_{Q_L} \eta\,d\mu, \qquad A_\eta = W_p^p\Big(\eta\mu, \frac{S_\eta}{V_\eta}\eta\,dx\Big). \tag{6.2}$$

Proposition 6.2. *Let $p \geq 1$, $a \in [1/2, 2]$, $\lambda = a\mathbf{1}_{Q_L}dx$, let μ be a finite measure on Q_L with $\mu(Q_L) = aL^d$, and let $\pi = a\,dx\,\pi_x$ be a $c_{b,p,L}$-optimal coupling of λ and μ. Define $\mathsf{D}, \mathsf{S}, K, g$ from π as in (4.2)–(4.4), let $\eta = \eta_{\delta,L}$, and put*

$$\kappa = \frac{S_\eta}{aV_\eta}, \qquad J = \int_{\mathsf{D}} |x-y|^p \Big(\frac{|x-y|}{L}\Big)^p d\pi.$$

Let $U = W_p^p(\lambda,\mu)$ and $B = \int c_{b,p,L}\,d\pi$. Then for every $\varepsilon \in (0,1]$,

$$U - B \leq C\varepsilon B + C\varepsilon^{1-p}\Big[A_\eta + L^p|\kappa-1|^p \int_{Q_L} \eta^p + \Big\|g\cdot\nabla\eta - \fint_{Q_L} g\cdot\nabla\eta\Big\|^p_{W^{-1,p}_{\#}} + \delta^{-p}K + \delta^{-2p}J\Big], \tag{6.3}$$

with $C = C(p,d)$. For $p = 1$ the inequality holds without the term $C\varepsilon B$.

Proof. Write $\zeta = 1 - \eta$ and $\pi_x(\zeta) = \int \zeta\,d\pi_x$, and let

$$\mu_1 = \eta\mu + a\pi_x(\zeta)\,dx, \qquad \mu_2 = \frac{S_\eta}{V_\eta}\eta\,dx + a\pi_x(\zeta)\,dx. \tag{6.4}$$

Since $a\int \pi_x(\zeta)\,dx = \int \zeta\,d\mu = aL^d - S_\eta$, both have mass aL^d. By the triangle inequality for W_p and $(s+t)^p \leq (1+\varepsilon)^{p-1}s^p + (1+\varepsilon^{-1})^{p-1}t^p$,

$$U \leq (1 + C_p\varepsilon)W_p^p(\mu,\mu_1) + C_p\varepsilon^{1-p}\{W_p^p(\mu_1,\mu_2) + W_p^p(\mu_2,\lambda)\}, \tag{6.5}$$

and for $p = 1$ the first coefficient is one.

First leg. The plan $\zeta(y)\pi(dx,dy)$ couples $a\pi_x(\zeta)dx$ with $\zeta\mu$; with the diagonal coupling of $\eta\mu$ it transports μ_1 to μ. On D its cost is at most the boundary cost. If $\zeta(y) > 0$ on a strict shortcut, then $\rho_L(y) > r$, and since $|x-y| \leq \sqrt{d}L$ the edge costs at most $d^{p/2}\delta^{-p}(\rho_L(x)^p + \rho_L(y)^p)$. Hence

$$W_p^p(\mu,\mu_1) \leq B + d^{p/2}\delta^{-p}K. \tag{6.6}$$

Second leg. The measures μ_1, μ_2 share $a\pi_x(\zeta)dx$, so

$$W_p^p(\mu_1,\mu_2) \leq A_\eta. \tag{6.7}$$

Third leg. Since π_x is a probability measure,

$$\begin{aligned} a^{-1}\frac{d(\mu_2-\lambda)}{dx} &= \kappa\eta - \pi_x(\eta) = (\kappa-1)\eta + g\cdot\nabla\eta + R_{\mathrm{dir}} + R_{\mathrm{short}}, \\ R_{\mathrm{dir}}(x) &= \int_{\mathsf{D}_x} [\eta(x) - \eta(y) - (x-y)\cdot\nabla\eta(x)]\,\pi_x(dy), \\ R_{\mathrm{short}}(x) &= \int_{\mathsf{S}_x} [\eta(x) - \eta(y)]\,\pi_x(dy). \end{aligned} \tag{6.8}$$

By Lemma 6.1, which applies since μ_2 has the bounded density $a[\kappa\eta + \pi_x(\zeta)] \leq a(\kappa+1)$ and the same mass as λ, $W_p^p(\mu_2,\lambda) \leq C\|\cdot\|^p_{W^{-1,p}_{\#}}$ of the right side of (6.8), since $a \leq 2$. We subtract the spatial mean from each term; their sum has zero mean. By (5.2), the term $(\kappa-1)\eta$ contributes $CL^p|\kappa-1|^p\int\eta^p$, and the flux is kept. If $\eta(x) \neq \eta(y)$ then $\max(\rho_L(x),\rho_L(y)) > r$, so on S this forces $c_{b,p,L}(x,y) > r^p$; as $|R_{\mathrm{short}}| \leq 1$, $\int |R_{\mathrm{short}}|^p \leq a^{-1}\pi\{\mathsf{S}\cap[\eta(x)\neq\eta(y)]\} \leq K/(ar^p)$. Taylor's formula on the convex cube and Jensen's inequality give $\int |R_{\mathrm{dir}}|^p \leq Ca^{-1}r^{-2p}\int_{\mathsf{D}}|x-y|^{2p}d\pi$. After multiplication by L^p (from (5.2)) these become $C\delta^{-p}K$ and $C\delta^{-2p}J$. Inserting (6.6), (6.7) and this bound into (6.5), and using $1 + C_p\varepsilon \leq C\varepsilon^{1-p}$ in front of the shortcut term, proves (6.3). □

Lemma 6.3. *Along all real $L \to \infty$,*

$$L^{-d}\mathbb{E}\int_{\mathsf{D}_L} |x-y|^p\Big(\frac{|x-y|}{L}\Big)^p d\pi_L \longrightarrow 0. \tag{6.9}$$

Proof. For $M > 0$, split according to $|x-y| \le M$ or $|x-y| > M$. The first part is at most $(M/L)^p B_L$, since $|x-y|^p \le c_{b,p,L}$ on D_L; the second at most $d^{p/2}\int_{\mathsf{D}_L} |x-y|^p \mathbf{1}_{\{|x-y|>M\}} d\pi_L$. Use (4.12) and (4.15), letting first $L \to \infty$ and then $M \to \infty$. No moment of order $2p$ is needed. □

7. The boundary layer and the proof of the main theorem

It remains to estimate the transport cost A_η inside the boundary layer and to collect the error terms of Proposition 6.2.

Lemma 7.1. *Let $\delta = 2^{-k}$ with $k \ge 4$ an integer, let $\eta = \eta_{\delta,L}$, and let $S_\eta = \int_{Q_L} \eta \, d\nu_L$. There is a function ω_p on dyadic numbers, depending only on p, d and the constants in (A2), with $\omega_p(\delta) \to 0$ as $\delta \to 0$, such that*

$$\limsup_{L\to\infty} L^{-d}\mathbb{E}W_p^p\Big(\eta\nu_L, \frac{S_\eta}{V_\eta}\eta\,dx\Big) \le \omega_p(\delta). \tag{7.1}$$

Proof. Suppose first that $p > 1$. We use the dyadic construction of [14, Lemma 4.5], treating all scales above $r = \delta L$ in a single comparison. Write $m = \eta\,dx$, and let $\mathcal{Q}_j$ be the dyadic partition of Q_L at side length $\ell_j = L2^{-j}$. For a cell Q with $m(Q) > 0$ put $\kappa_Q = m(Q)^{-1}\int_Q \eta\,d\nu_L$, and define $\lambda^{(j)} = \sum_{Q\in\mathcal{Q}_j} \kappa_Q m|_Q$, with zero contribution from cells with $m(Q) = 0$. Stationarity and intensity one imply that the deterministic subdivision boundaries and the sets with $m(Q) = 0$ carry no $\eta\nu_L$ mass almost surely. All $\lambda^{(j)}$ have mass S_η, and $\lambda^{(0)} = (S_\eta/V_\eta)m$.

We use two properties of the weight. For $j \ge k$ and $Q \in \mathcal{Q}_j$ with $m(Q) > 0$, one has $m(Q) \ge c_d|Q|\sup_Q \eta$ and, for $s = p'$,

$$\inf_{t\in\mathbb{R}} \|\varphi - t\|_{L^s(m|_Q)} \le C_{d,s}\ell_j \|\nabla\varphi\|_{L^s(m|_Q)} \qquad (\varphi \in C^1(\overline{Q})).$$

The same Poincaré inequality holds on Q_L with right side $C_{d,s,\delta}L\|\nabla\varphi\|_{L^s(m)}$. To see this, η is comparable, within factors depending only on d, to the sum of its $2d$ face weights. A face weight at distance t from that face is 1 for $0 \le t \le r$, is $f_0((2r-t)/r)$ for $r \le t \le 2r$, and vanishes beyond $2r$. In particular it is monotone in the normal coordinate and is comparable on its support to $\min\{1, ((2r-t)/r)^3\}$. Each such weight on a rectangle satisfies a Poincaré inequality with constant bounded by a multiple of the largest side. In the normal direction, for a nondecreasing weight w on (a,b) use $|\varphi(x)-\varphi(b)|^s \le (b-a)^{s-1}\int_x^b |\varphi'|^s$ and $\int_a^z w \le (b-a)w(z)$. Subtracting the weighted mean costs at most a factor two. Averaging in the remaining coordinates and applying the ordinary Poincaré inequality on their rectangle gives the product estimate.

At scales $\ell_j \le r$ the planes $t = r, 2r$ lie on the grid. Every active face weight is bounded below by a fixed fraction of its supremum on the half of Q nearer its face. There are no opposite active faces in one cell, so these halves have a common subrectangle of volume at least $2^{-d}|Q|$. This gives the stated bound on $m(Q)$. It also allows us to replace each face mean by the unweighted average on that subrectangle: by Jensen's inequality the L^s error of this replacement for each face weight is bounded by its Poincaré error times a uniform constant. On Q_L, the strips of width r along any two nonopposite faces overlap in a rectangle on which both weights are one. The graph connecting such faces is connected for $d \ge 2$. Jensen's inequality on the overlaps bounds differences of the face means by the sum of their Poincaré errors divided by the sth root of the overlap volume. Using one face mean as the common constant yields the claimed inequality on Q_L, with a constant depending on δ.

Consider a parent Q and a partition of it into cells Q' at the next scale, or take $Q = Q_L$ and the partition $\mathcal{Q}_k$. Let $F = \sum_{Q'} \kappa_{Q'}\mathbf{1}_{Q'}$. On $\{\kappa_Q \ge 1/2\}$ the weighted

transport estimate of [14, Lemma 3.6] and the preceding Poincaré inequality imply

$$W_p^p(Fm|_Q, \kappa_Q m|_Q) \le C\,\mathrm{diam}(Q)^p \int_Q |F-\kappa_Q|^p\,dm,$$

where C may depend on δ for $Q = Q_L$. To apply the estimate with the possibly vanishing weight η, add $\epsilon\kappa_Q\,dx|_Q$ to both measures and use the positive weight $\eta+\epsilon$. The difference pairs with a test function as $\int_Q (F-\kappa_Q)\varphi\,dm$ and annihilates constants. The weighted Poincaré inequality on $m|_Q$ bounds this pairing by the displayed right side to the power $1/p$ times the gradient norm in the larger weight. The constant in the negative-norm bound of HGT is independent of the weight, as seen from its proof by dynamical interpolation. Letting $\epsilon \downarrow 0$ and taking a weak limit of optimal plans on $\overline{Q}^2$ gives the estimate. On $\{\kappa_Q < 1/2\}$, use the diameter bound $W_p^p \le \mathrm{diam}(Q)^p m(Q)\kappa_Q$ and $\mathbb{P}(\kappa_Q < 1/2) \le 2^p\mathbb{E}|\kappa_Q - 1|^p$.

For cells at scales $\ell_j \ge 1$ with $j \ge k$, (A2) and $m(Q) \ge c_d|Q|\sup_Q \eta$ give $\|\kappa_Q - 1\|_{L^p(\mathbb{P})} \le C\ell_j^{-(d-\alpha)/2}$. For $Q = Q_L$ the bound is $C_\delta L^{-(d-\alpha)/2}$ by (6.1). The same bound with C_δ holds on $\mathcal{Q}_k$ since $\ell_k = \delta L$. Put $\gamma = (d-\alpha)/2 - 1 > 0$. Taking expectations in the parent estimates and adding disjoint cells yields

$$\mathbb{E}W_p^p(\lambda^{(0)}, \lambda^{(k)}) \le C_\delta V_\eta L^{-p\gamma},$$
$$\mathbb{E}W_p^p(\lambda^{(j)}, \lambda^{(j+1)}) \le C V_\eta \ell_j^{-p\gamma} \qquad (j \ge k,\ \ell_{j+1} \ge 1).$$

Here the integral of $|F-\kappa_Q|^p$ is estimated using the moment bounds for the parent and its children, and $\sum_{Q\in\mathcal{Q}_j} m(Q) = V_\eta$. For fixed δ and sufficiently large L, choose $J \ge k$ with $1 \le \ell_J < 2$. The diameter bound in the finest cells gives $\mathbb{E}W_p^p(\lambda^{(J)}, \eta\nu_L) \le CV_\eta$. The triangle inequality for W_p, followed by Minkowski in probability, allows us to sum the pth roots of these bounds. Since $\sum_{j=k}^{J-1} \ell_j^{-\gamma} \le \sum_{i\ge 1} 2^{-i\gamma}$, we obtain

$$(\mathbb{E}A_\eta)^{1/p} \le V_\eta^{1/p}(C + C_\delta L^{-\gamma}).$$

Thus (7.1) holds with $\omega_p(\delta) = C\delta$.

For $p = 1$ we reduce to $p = 2$. The two measures in (7.1) have the common mass S_η, and for measures of equal mass m the unnormalized costs satisfy $W_1 \le m^{1/2}(W_2^2)^{1/2}$, by Jensen's inequality applied to an optimal plan for W_2^2. Hence, by the Cauchy–Schwarz inequality and $\mathbb{E}S_\eta = V_\eta \le C\delta L^d$,

$$L^{-d}\mathbb{E}W_1\Big(\eta\nu_L, \frac{S_\eta}{V_\eta}\eta\,dx\Big) \le (L^{-d}V_\eta)^{1/2}\Big(L^{-d}\mathbb{E}W_2^2\Big(\eta\nu_L, \frac{S_\eta}{V_\eta}\eta\,dx\Big)\Big)^{1/2},$$

and the case $p = 2$ gives (7.1) with $\omega_1(\delta) = (C\delta\,\omega_2(\delta))^{1/2}$. □

Proposition 7.2. *Under the assumptions of Theorem 1.2, including* (1.8), $L^{-d}\mathbb{E}[U_L - B_L] \to 0$ *along all real* $L \to \infty$.

Proof. Let $G_L = \{1/2 \le a_L \le 2\}$, fix $\varepsilon \in (0,1]$ and a dyadic $\delta \le 1/16$, and apply Proposition 6.2 on G_L with $\mu = \nu_L$, $\pi = \pi_L$ and $\eta = \eta_{\delta,L}$. We bound the normalized expectation of each term in (6.3).

The layer term has $\limsup_L L^{-d}\mathbb{E}A_{\eta_{\delta,L}} \le \omega_p(\delta)$ by Lemma 7.1. For the mass correction, on G_L, $|\kappa - 1| \le 2V_\eta^{-1}(|S_\eta - V_\eta| + |a_L - 1|V_\eta)$. By Lemma 2.1, $\mathbb{E}|S_\eta - V_\eta|^p \le CL^{p(d+\alpha)/2}$ and $\mathbb{E}|a_L - 1|^p \le CL^{-p(d-\alpha)/2}$. Since $\int \eta^p \le V_\eta$ and $V_\eta \ge c\delta L^d$,

$$L^{-d}\mathbb{E}\Big[\mathbf{1}_{G_L}L^p|\kappa - 1|^p \int \eta^p\Big] \le C(\delta)L^{p(2+\alpha-d)/2} \to 0.$$

The flux term vanishes by Proposition 5.2, applied to the fixed profile $\eta_{\delta,1} \in C^2([0,1]^d)$; the shortcut term vanishes by (4.13); the Taylor term vanishes by Lemma 6.3. On G_L^c we

use $0 \le U_L - B_L \le d^{p/2} L^p \Sigma_L$ and $L^{p-d}\mathbb{E}[\Sigma_L \mathbf{1}_{G_L^c}] \le C L^p \mathbb{P}(G_L^c)^{1/2} \to 0$ by Lemma 2.1. Using also (4.12),

$$0 \le \limsup_{L\to\infty} L^{-d}\mathbb{E}(U_L - B_L) \le C\{\varepsilon e_\nu + \varepsilon^{1-p}\omega_p(\delta)\},$$

where for $p = 1$ the term εe_ν is absent. Let $\delta \downarrow 0$ along dyadic values, and then $\varepsilon \downarrow 0$. □

Proof of Theorem 1.2. By Corollary 4.2 and Proposition 3.4, $e_\nu < \infty$, the optimal stationary coupling exists, is unique and is induced by the equivariant map T_ν, and $e_\nu > 0$ unless $\nu = dx$ almost surely. Under (1.8), Proposition 7.2 gives (1.9). By (1.7), $0 \le L^{-d}\mathbb{E}[U_L - T_L] \le L^{-d}\mathbb{E}[U_L - B_L] \to 0$, and similarly for $T_L - B_L$. By (4.12) all three specific costs converge to e_ν. □

Proposition 7.3. *Let $d \ge 3$, $1 \le p < \infty$, and let ν satisfy (A1)–(A2), almost surely purely atomic if $p = 1$; the centering condition (1.8) is* not *assumed. Then $L^{-d}\mathbb{E}B_L \to e_\nu$ along all real $L \to \infty$, and*

$$\liminf_{L\to\infty} L^{-d}\mathbb{E}U_L = e_\nu \quad \Longrightarrow \quad (1.8) \text{ holds.}$$

Consequently (1.8), $\lim_L L^{-d}\mathbb{E}U_L = e_\nu$ and (1.9) are equivalent.

Proof. By Corollary 4.2, Proposition 4.4 and $B_L \le U_L$, $e_\nu \le \lim_L L^{-d}\mathbb{E}B_L \le \liminf_L L^{-d}\mathbb{E}U_L$, where the first limit exists and equals e_ν; none of this uses (1.8). Let $L_n \to \infty$ with $L_n^{-d}\mathbb{E}U_{L_n} \to e_\nu$.

Stationarization of ordinary optimizers. Choose measurably a $|x-y|^p$-optimal plan π_L^U of λ_L and ν_L. The proof of Lemma 4.1, with D_L replaced by Q_L^2 (so that $M_L^b = 0$), gives invariant subcouplings $\mathbf{q}_L$ satisfying (4.8) with π_L replaced by π_L^U, with $e(\mathbf{q}_L) \le L^{-d}\mathbb{E}U_L$ and, by (4.10) and (4.6), $\theta(\mathbf{q}_L) \le L^{-d}\mathbb{E}|\Sigma_L - L^d| \to 0$. By Lemma 3.7, $\int \mathbf{1}_{[0,1)^d}(x)|y - T_\nu|^p \, d\mathbf{q}_{L_n} \to 0$.

Mean displacement of ordinary plans. Since $\int_{Q_L}(x - x_0)\,dx = 0$ for the centre x_0 of Q_L and λ_L, ν_L have the same mass, every coupling π of λ_L, ν_L satisfies $\int (x-y)\,d\pi = -\int_{Q_L}(y - x_0)\,(d\nu - dy)$. Applying (1.1) to each coordinate of $y - x_0$, shifted by $\sqrt{d}L/2$ to be nonnegative and bounded by $\sqrt{d}L$ (the shift costs a term $L\,|\Sigma_L - L^d|$, of the same order),

$$L^{-d}\,\mathbb{E}\Big|\int (x-y)\,d\pi_L^U\Big| \le C L^{-d}\, L\, L^{(d+\alpha)/2} = C L^{(2+\alpha-d)/2} \to 0,$$

because $\alpha < d - 2$.

Conclusion. Let $A \in \mathcal{I}$ and $i \le d$. The functions $F^\pm = \mathbf{1}_A(\omega)(x-y)_i^\pm$ are invariant, and integrable against $\mathbf{1}_{[0,1)^d}(x)\mathbf{q}_L$ since $p \ge 1$ and $e(\mathbf{q}_L) < \infty$. By (4.8), with $0 \le \alpha_L \le 1$,

$$\Big|\int \mathbf{1}_{[0,1)^d}(x)\mathbf{1}_A(x-y)_i\,d\mathbf{q}_L\Big| \le L^{-d}\,\mathbb{E}\Big|\int (x-y)\,d\pi_L^U\Big| \to 0.$$

Since $\mathbf{q}_{L_n}$ converges to the graph of T_ν in the sense above, and its source marginal on $\Omega \times [0,1)^d$ has a density $\varrho_n \le 1$ with respect to $m_\Lambda = \mathbf{P} \otimes \mathbf{1}_\Lambda dx$, $\Lambda = [0,1)^d$, and $\int \varrho_n\,dm_\Lambda = 1 - \theta(\mathbf{q}_{L_n}) \to 1$, we get $\varrho_n \to 1$ in $L^1(m_\Lambda)$, and

$$\int \mathbf{1}_\Lambda(x)\mathbf{1}_A(x-y)_i\,d\mathbf{q}_{L_n} = \int \mathbf{1}_A b_{\nu,i}\varrho_n\,dm_\Lambda + O\Big(\int \mathbf{1}_\Lambda(x)|y - T_\nu|\,d\mathbf{q}_{L_n}\Big)$$
$$\longrightarrow \mathbb{E}[\mathbf{1}_A\beta_{\nu,i}].$$

Indeed the error term tends to 0 by Jensen (the measures have mass at most one), and $\int \mathbf{1}_A b_{\nu,i}\varrho_n\,dm_\Lambda \to \int \mathbf{1}_A b_{\nu,i}\,dm_\Lambda$ because $|b_\nu| \in L^1(m_\Lambda)$ and $0 \le 1 - \varrho_n \le 1$ with $1 - \varrho_n \to 0$ in measure; finally $\int \mathbf{1}_A b_{\nu,i}\,dm_\Lambda = \mathbb{E}[\mathbf{1}_A\beta_{\nu,i}]$ by stationarity and $A \in \mathcal{I}$. Hence $\mathbb{E}[\mathbf{1}_A\beta_\nu] = 0$ for every $A \in \mathcal{I}$, which is (1.8). The last assertion follows from Proposition 7.2, $\lim L^{-d}\mathbb{E}B_L = e_\nu$ and (1.7). □

Remark 7.4. For $p = 3/2$, there is a stationary measure satisfying (A1)–(A2) with $\alpha = 0$ for which $L^{-d}\mathbb{E}U_L$ has strictly larger lower limit than e_ν. To construct it, fix a smooth function ψ supported in B_1 such that $\int |\nabla\psi|\nabla\psi\,dx \neq 0$. Indeed, take $f(x_1)g(x'/R)$ with $g \geq 0$ smooth, compactly supported and nonzero, and f' the difference of two disjoint smooth nonnegative bumps of equal integral and different squared integrals. After division by R^{d-1}, the first component of the displacement integral converges as $R \to \infty$ to $\int |f'|f' \int g^2 \neq 0$. A spatial dilation puts the support in B_1.

Let $\mathcal{Z}$ consist of points of a unit Poisson process with no other point within distance 4. For a sufficiently small constant $a > 0$, set

$$u(x) = a \sum_{z\in\mathcal{Z}} \psi(x - z), \qquad T(x) = x - |\nabla u(x)|\nabla u(x), \qquad \nu = T_{\#}dx.$$

The supports are disjoint, so we may choose a such that $\operatorname{osc} u < 2/3$, $\|D^2u\|_\infty < 1/2$, and $|T(x) - x| < 1$. The map and its graph are equivariant on the Poisson space. The bounded displacement ensures local finiteness, and mass transport gives intensity one. The restriction of ν to any unit cube is determined by the Poisson process within distance 6 of that cube. Its mass is bounded by the volume of the 1-neighbourhood of the cube. Partition the unit cubes of the integer grid into finitely many classes, with disjoint 6-neighbourhoods within each class. For a test function in (1.1), the centred integrals on the grid cubes of each class are independent and bounded in absolute value by $C_d\|f\|_\infty$. A sum of N such variables has L^q norm at most $C_{d,q}\sqrt{N}\|f\|_\infty$, by expanding an even moment of order at least q and using centering. There are $O_d(\operatorname{diam}(Q)^d)$ cubes to consider, proving (A2).

For $h(z) = \frac{2}{3}|z|^{3/2}$ put $v(y) = \inf_x(h(x - y) - u(x))$. All minimizers lie in $B_1(y)$ by comparison with $x = y$. On that ball $h(\cdot - y)$ is $1/2$-strongly convex, and hence $h(\cdot - y) - u$ is strictly convex there. For $y = T(x)$, the point x is in the ball and solves its first-order condition. Therefore $u(x) + v(y) \leq h(x - y)$, with equality on the graph. The potentials are bounded and equivariant on the Poisson space. Their expected integrals per unit volume depend only on the marginals, using Lemma 3.2 to place the indicator on the target for v. Any stationary competitor on the space of ν can be lifted by evaluating its kernel at the ν constructed on the Poisson space. This proves optimality of the projected graph coupling, and Proposition 3.4 identifies it with $\mathbf{q}_\nu$. By stationarity, $\mathcal{Z}$ has some positive intensity $\lambda_\mathcal{Z}$ (positivity follows by requiring one point in a small ball and none in its 4-neighbourhood outside that ball). Thus

$$\mathbb{E}\beta_\nu = \mathbb{E}\int_{[0,1)^d} (x - T(x))\,dx = \lambda_\mathcal{Z} a^2 \int |\nabla\psi|\nabla\psi\,dx \neq 0.$$

The strict gap follows from Proposition 7.3.

Remark 7.5. For the randomly shifted lattice $\nu = \sum_{k\in\mathbb{Z}^d} \delta_{k+U}$, with U uniform on $(0, 1)^d$, the costs can differ even for $p = 2$ (see [16, Remark 1.4]). The lattice satisfies (A1), but (A2) fails: for integer L the test $f(x) = \mathbf{1}_{\{\{x_1\}<1/2\}}$ on Q_L, with $\{x_1\}$ the fractional part, gives an absolute discrepancy $L^d/2$. The nearest-point allocation is stationary, is optimal by pointwise minimization, and has specific quadratic cost $d/12$. Its displacement is centered by Proposition 3.5(a). On integer cubes $a_L = 1$, so Lemmas 2.4 and 4.3 give $L^{-d}\mathbb{E}B_L \leq d/12$. For the ordinary cost the optimal plan is the product of the one-dimensional monotone transports, each of which sends $[k, k + 1)$ to $k + U_i$. Indeed, quadratic cost is additive in the coordinates and the coordinate projections of any coupling are admissible in the one-dimensional problems after normalizing the mass. It follows that

$$L^{-d}U_L = \sum_{i=1}^{d} (\tfrac{1}{12} + (U_i - \tfrac{1}{2})^2), \qquad L^{-d}\mathbb{E}(U_L - B_L) \geq \tfrac{d}{12}.$$

Proof of Theorem 1.3. By Lemma 2.2, ν^W satisfies (A1)–(A2) with $\alpha = 0 < d-2$, it is purely atomic, and its law is ergodic and invariant under $x \mapsto -x$. By Proposition 3.5(b) the centering condition holds, and Theorem 1.2 applies. Finally ν^W is not Lebesgue measure, so $e_W > 0$. □

Remark 7.6. The parameters are removed in the order $L \to \infty$ (with the mesoscopic block size h_L of Proposition 5.2 tied to L), then $\delta \downarrow 0$ along dyadic values, then $\varepsilon \downarrow 0$; no rate is claimed. The condition $\alpha < d-2$ enters through the decay $L^{p(2+\alpha-d)/2} \to 0$ of the mass corrections, the boundary-layer estimate and the ordinary cost bound. The centering condition (1.8) is used only in Proposition 5.2, and atomicity of ν only for the uniqueness of the stationary optimizer when $p = 1$.

8. Applications to matching

8.1. **Unit marks: semi-discrete matching.** For $W \equiv 1$, ν is the Poisson process, $e_1(p,d) = e_*(p,d)$ is the cost of the optimal stationary allocation of [15], and Theorem 1.3 states that the periodic, ordinary and boundary semi-discrete Poisson costs per unit volume converge to $e_*(p,d)$ for every $p \geq 1$, $d \geq 3$. Let $Q = (0,1)^d$, let $X_1, \dots, X_n$ be independent uniform points in Q and $\mu_n = \sum_{i=1}^n \delta_{X_i}$. Caglioti et al. [4] consider

$$\bar{c}_{p,d} = \limsup_{n\to\infty} n^{p/d-1}\mathbb{E}W_p^p(\mu_n, n\mathbf{1}_Q), \qquad \underline{c}_{p,d} = \liminf_{n\to\infty} n^{p/d-1}\mathbb{E}W_Q^{b,p}(\mu_n, n\mathbf{1}_Q),$$

and conjecture that $\underline{c}_{p,d} = \bar{c}_{p,d}$ for every p and d [4, Remark 1.3]. For measures of equal mass, $W_Q^{b,p}$ is the transport cost for $c_{b,p,1}$ (Lemma 2.4 and Remark 2.5).

Corollary 8.1. *For every $d \geq 3$ and $p \geq 1$, with $c_{\mathrm{per}} = d_1^{\mathrm{per}}(x,y)^p$,*

$$\begin{aligned}\lim_{n\to\infty} n^{p/d-1}\mathbb{E}W_p^p(\mu_n, n\mathbf{1}_Q) &= \lim_{n\to\infty} n^{p/d-1}\mathbb{E}\,\mathrm{OT}_{c_{\mathrm{per}}}(\mu_n, n\mathbf{1}_Q)\\ &= \lim_{n\to\infty} n^{p/d-1}\mathbb{E}W_Q^{b,p}(\mu_n, n\mathbf{1}_Q) = e_*(p,d).\end{aligned}$$

In particular $\underline{c}_{p,d} = \bar{c}_{p,d} = e_*(p,d)$.

Proof. Let $f(m) = \mathbb{E}W_p^p(\mu_m, m\mathbf{1}_Q)$, $f^{\mathrm{per}}(m)$ and $f^b(m)$ the analogous expectations, let N_n be Poisson with mean n, independent of the points, and put $L = n^{1/d}$. After the rescaling $x \mapsto x/L$, the pair (λ_L, ν_L) for unit marks has the law of $(N_n\mathbf{1}_Q, \mu_{N_n})$, and the three costs are multiplied by L^{-p}. Since $n^{p/d-1} = L^{p-d}$, Theorem 1.3 gives $n^{p/d-1}\mathbb{E}f(N_n) = L^{-d}\mathbb{E}U_L^1 \to e_*$ and likewise for f^b. By the de-Poissonization of [4, Proposition A.1], which applies to the ordinary and the boundary costs, $\limsup_n n^{p/d-1}f(n) \leq e_*$ and $\liminf_n n^{p/d-1}f^b(n) \geq e_*$. Since $f^b \leq f^{\mathrm{per}} \leq f$ by (1.6), all three sequences converge to e_*. □

The equality of the constants also gives the limit for samples of a general law. Huesmann et al. [14, Theorem 1.1] collect, from [6], [3] and [12], asymptotic upper and lower bounds in which the constants are the limits of $L^{-d}\mathbb{E}U_L^1$ and $L^{-d}\mathbb{E}B_L^1$; they note that the existence of the limit would follow from the equality of these constants [14, Section 1.1].

Corollary 8.2. *Let $d \geq 3$, $p \geq 1$, and let λ be a probability measure on $\mathbb{R}^d$ with absolutely continuous part $\lambda_a\,dx$. Assume that either*

(i) $p < d/2$ and $\int |x|^q\,d\lambda < \infty$ for some $q > dp/(d-p)$; or
(ii) λ is supported on the closure of a bounded connected open set D which is C^2 or convex, and $\lambda = \lambda_a\mathbf{1}_D dx$ with λ_a Hölder continuous and bounded above and below by positive constants on D.

In case (i) use $D = \mathbb{R}^d$ if (ii) is not assumed. If $X_1, X_2, \dots$ are independent with law λ, then

$$\lim_{n\to\infty} n^{p/d}\,\mathbb{E}W_p^p\Big(\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i},\lambda\Big) = e_*(p,d)\int_D \lambda_a^{1-p/d}\,dx.$$

Proof. In case (i), Huesmann et al. [14, Theorem 1.1] bound the lower and upper limits by the respective limits of $L^{-d}\mathbb{E}B_L^1$ and $L^{-d}\mathbb{E}U_L^1$ multiplied by $\int_{\mathbb{R}^d}\lambda_a^{1-p/d}\,dx$. These constants agree by Theorem 1.3. In case (ii), Caglioti et al. [4, Theorems A.2 and A.5 in the arXiv version] give the corresponding upper and lower bounds with the integral over D. In the lower bound, the sum of $|Q|(\fint_Q \lambda_a)^{1-p/d}$ over interior grid cubes converges to this integral by uniform continuity and the positive lower bound on λ_a, also when $p > d$. The uniform constants in these bounds agree by Corollary 8.1. □

The corollary is a statement about expectations, under exactly the hypotheses of [14, Theorem 1.1]; in the range $p \ge d/2$ some structural assumption such as (ii) is needed for the upper bound [14, Section 1]. It does not concern bipartite matching between two samples of a nonuniform law.

Dereich, Scheutzow and Schottstedt. For $1 \le p < d/2$, Dereich et al. [6, Theorem 2(i)] prove that $N^{1/d}\big(\mathbb{E}W_p^p(\widehat{\mu}_N, U)\big)^{1/p}$ converges to a constant κ_p^{unif} (we take the Euclidean norm), where $\widehat{\mu}_N$ is the empirical probability measure of N uniform points and U the uniform probability on $[0,1)^d$. They also prove convergence for a lower functional in which U may be replaced by any probability measure U' with $U' \le U$ on $(0,1)^d$, the missing mass being placed on the boundary, and they conjecture that the two constants coincide [6, Remark 2]. The upper constant satisfies $(\kappa_p^{\mathrm{unif}})^p = e_*(p,d)$ by Corollary 8.1. For the lower constant we use the clipping argument with the boundary-source edges removed.

For unit marks let C_L be the infimum of the ordinary cost over couplings with target ν_L and source on $\overline{Q}_L$ whose restriction to Q_L is dominated by λ_L. This infimum is attained and has a measurable optimizer by compactness and measurable selection on $\overline{Q}_L^2$. Indeed, the marginal constraint can be tested against continuous nonnegative functions supported in Q_L. We have $C_L \le U_L^1$. Delete the edges starting on ∂Q_L in an optimizer and multiply by α_L as in Lemma 4.1. The retained marginals are dominated by $dx|_{Q_L}$ and ν_L. If M_L is the deleted mass then

$$M_L \le \nu_L\{\rho_L \le R\} + R^{-p}C_L.$$

Indeed, a boundary-source edge to a target with $\rho_L > R$ costs at least R^p. The first term has expectation at most $2dRL^{d-1}$ by intensity one. The stationarization of Lemma 4.1 applies to the retained plan. Its energy is at most $L^{-d}\mathbb{E}C_L$, and its defect is at most $L^{-d}\mathbb{E}[|\Sigma_L - L^d| + M_L]$. This tends to zero by first sending $L \to \infty$ at fixed R and then $R \to \infty$. Compactness gives $e_*(p,d) \le \liminf L^{-d}\mathbb{E}C_L$. The upper bound follows from $C_L \le U_L^1$.

Let $f_N = \mathbb{E}\inf_{U'} W_p^p(U', \widehat{\mu}_N)$, with the constraint from [6]. The cost as a function of the empirical probability measure is convex by mixing plans together with their admissible sources. Averaging over samples with one point omitted shows that f_N is nonincreasing. Conditional on the Poisson count N_L, scaling the cube and dividing by the common mass gives $\mathbb{E}C_L = L^p\mathbb{E}[N_L f_{N_L}]$, with the zero-count contribution set to zero. The sequence $F_N = Nf_N$ satisfies the hypotheses of Lemma 8.6 below, since $f_N \le d^{p/2}$ and $L^{p-d}\mathbb{E}F_{N_L} \to e_*(p,d)$. Thus $N^{p/d}f_N \to e_*(p,d)$, proving $\underline{\kappa}_p^{\mathrm{unif}} = \kappa_p^{\mathrm{unif}}$.

8.2. **Exponential marks: bipartite matching.** Fix $Q = (0,1)^d$. For each $n \ge 1$ let $X_1, \dots, X_n, Y_1, \dots, Y_n$ be independent uniform points in Q. For $\sharp \in \{\mathrm{per}, \mathrm{cube}, \mathrm{bd}\}$ let $c_{\mathrm{per}} = d_1^{\mathrm{per}}(x,y)^p$, $c_{\mathrm{cube}} = |x-y|^p$, $c_{\mathrm{bd}} = c_{b,p,1}$ (and $c_{\sharp,L}$ the corresponding costs

$d_L^{\mathrm{per}}(x,y)^p$, $|x-y|^p$, $c_{b,p,L}$ on Q_L), and

$$F_n^\sharp = \mathbb{E}\Big[\min_{\sigma\in\mathcal{S}_n}\sum_{i=1}^n c_\sharp(X_i, Y_{\sigma(i)})\Big], \qquad F_0^\sharp = 0. \tag{8.1}$$

The boundary assignment equals the optimal *partial* matching with penalty $\rho_1(z)^p$ for every unmatched point: discard each pair using the boundary alternative or, conversely, complete a partial matching by pairing the remaining points through the boundary. For $p=2$ this is the boundary convention of [26, Appendix A.2].

Theorem 8.3. *For $d\ge 3$ and $1\le p<\infty$, with $e_{\mathrm{Exp}}(p,d)$ the constant of Theorem 1.3 for $W\sim\mathrm{Exp}(1)$,*

$$\lim_{n\to\infty} n^{p/d-1}F_n^{\mathrm{per}} = \lim_{n\to\infty} n^{p/d-1}F_n^{\mathrm{cube}} = \lim_{n\to\infty} n^{p/d-1}F_n^{\mathrm{bd}} = e_{\mathrm{Exp}}(p,d)\in(0,\infty). \tag{8.2}$$

For the cube cost, the existence of the limit is due to [11]. For $p=2$, the equality of the ordinary and boundary limits, for every $d\ge 3$, for two samples whose common size is Poisson, is obtained in [26, Appendix A.2] by combining his identity with the quadratic comparison of [14, Theorem 4.3]; there the identification with the constants of [3] is carried out for $d\ge 5$.

Consider now the marked measure ν^W of Section 1.2 with $W\sim\mathrm{Exp}(1)$, let $N_L = \mathcal{P}(Q_L)$ be the number of its points in Q_L, and write $H_{L,\mathrm{per}} = T_L^{\mathrm{Exp}}$, $H_{L,\mathrm{cube}} = U_L^{\mathrm{Exp}}$, $H_{L,\mathrm{bd}} = B_L^{\mathrm{Exp}}$ for its three costs (1.3)–(1.5).

Lemma 8.4. *For each $n\ge 0$, $L>0$ and $\sharp$,*

$$\mathbb{E}[H_{L,\sharp}\mid N_L=n] = L^pF_n^\sharp, \qquad \mathbb{E}H_{L,\sharp} = L^p\mathbb{E}F_{N_L}^\sharp. \tag{8.3}$$

Proof. Condition on $N_L=n>0$ and on the positions $Y_1,\dots,Y_n$ of the points in Q_L, which are independent and uniform. Let $\Sigma=\sum_j E_j$ and $A_j=E_j/\Sigma$. Then $\Sigma\sim\Gamma(n,1)$ and $A\sim\mathrm{Dirichlet}(1,\dots,1)$ are independent (Remark A.3). By homogeneity of the transport cost in the common mass, $H_{L,\sharp}=\Sigma\,\mathrm{OT}_{c_{\sharp,L}}(L^{-d}dx|_{Q_L},\sum_j A_j\delta_{Y_j})$, so $\mathbb{E}[H_{L,\sharp}\mid N_L=n,Y] = n\mathbb{E}_A\,\mathrm{OT}_{c_{\sharp,L}}(L^{-d}dx|_{Q_L},\sum_j A_j\delta_{Y_j})$. By Proposition A.1, applied to the bounded functions $c_j=c_{\sharp,L}(\cdot,Y_j)$ and the uniform law on Q_L, the right side is the expected optimal assignment cost of n independent uniform points of Q_L to $Y_1,\dots,Y_n$. Averaging over Y and scaling Q_L to Q, which multiplies each of the three pair costs by L^{-p}, gives the first identity; the case $n=0$ is trivial, and the second identity follows. □

Lemma 8.5. *For each $\sharp$, $f_n^\sharp := F_n^\sharp/n$ is nonincreasing in $n\ge 1$.*

Proof. Let $\widehat{\mu}_{n+1},\widehat{\nu}_{n+1}$ be the empirical probability measures of $X_1,\dots,X_{n+1}$ and $Y_1,\dots,Y_{n+1}$, and let $\widehat{\mu}_n^{(-j)},\widehat{\nu}_n^{(-j)}$ be those obtained by deleting the *same* index j from both samples. Then $\widehat{\mu}_{n+1}=\frac{1}{n+1}\sum_j\widehat{\mu}_n^{(-j)}$ and likewise for $\widehat{\nu}_{n+1}$. The optimal transport cost between probability measures is jointly convex for any fixed pair cost, so $\mathrm{OT}_{c_\sharp}(\widehat{\mu}_{n+1},\widehat{\nu}_{n+1})\le\frac{1}{n+1}\sum_j\mathrm{OT}_{c_\sharp}(\widehat{\mu}_n^{(-j)},\widehat{\nu}_n^{(-j)})$. By the Birkhoff–von Neumann theorem, the transport cost between two empirical measures with k atoms each is k^{-1} times the assignment cost; taking expectations gives $f_{n+1}^\sharp\le f_n^\sharp$. □

Lemma 8.6. *Let $F_n\ge 0$ satisfy $F_n\le Cn$ and F_n/n nonincreasing, and let $d,p>0$. If, for integers $M\to\infty$ and $N_M\sim\mathrm{Pois}(M^d)$,*

$$M^{p-d}\mathbb{E}F_{N_M}\longrightarrow e\in[0,\infty), \tag{8.4}$$

then $n^{p/d-1}F_n\to e$ as $n\to\infty$ through all integers.

Proof. Put $f_n=F_n/n$. Fix $\varepsilon\in(0,1/2)$ and let $m=L^d$ with L an integer. Choose integers $M_\pm(L)$ with $M_\pm(L)/L\to(1\pm\varepsilon)^{1/d}$ and let $N_\pm\sim\mathrm{Pois}(M_\pm^d)$. Poisson concentration gives $\mathbb{E}[N_-\mathbf{1}_{\{N_-\le m\}}]=M_-^d(1+o(1))$ and $\mathbb{E}[N_+\mathbf{1}_{\{N_+\ge m\}}]=M_+^d(1+o(1))$,

with exponentially small complementary expectations. On $\{N_- \le m\}$, $F_{N_-} \ge N_- f_m$, so $L^p f_m \le L^p \mathbb{E}F_{N_-}/\mathbb{E}[N_- \mathbf{1}_{\{N_- \le m\}}]$, whose upper limit is at most $e(1-\varepsilon)^{-p/d}$. On $\{N_+ \ge m\}$, $F_{N_+} \le N_+ f_m$, while $\mathbb{E}[F_{N_+}\mathbf{1}_{\{N_+<m\}}] \le C\mathbb{E}[N_+\mathbf{1}_{\{N_+<m\}}] = o(M_+^{d-p})$; hence $L^p f_m \ge L^p(\mathbb{E}F_{N_+} - o(M_+^{d-p}))/\mathbb{E}[N_+\mathbf{1}_{\{N_+\ge m\}}]$, whose lower limit is at least $e(1+\varepsilon)^{-p/d}$. Letting $\varepsilon \downarrow 0$, $(L^d)^{p/d} f_{L^d} \to e$. For $L^d \le n \le (L+1)^d$, monotonicity gives $f_{(L+1)^d} \le f_n \le f_{L^d}$, and $(L+1)/L \to 1$ extends the limit to all n. □

Proof of Theorem 8.3. The law Exp(1) satisfies (H). By Theorem 1.3, each $L^{-d}\mathbb{E}H_{L,\sharp}$ converges to $e_{\mathrm{Exp}}(p,d) \in (0,\infty)$, and by (8.3) this reads $L^{p-d}\mathbb{E}F^{\sharp}_{N_L} \to e_{\mathrm{Exp}}(p,d)$. For each cost $0 \le F^{\sharp}_n \le d^{p/2}n$, and Lemma 8.5 gives the monotonicity required by Lemma 8.6. □

Corollary 8.7. *Let $d \ge 3$ and $1 \le p < d/2$. Let $\beta_p(d)$ be the constant of [3] for uniform samples, that is, the limit of $n^{p/d-1}$ times the optimal matching cost of n independent uniform points with n independent uniform points in Q, and let $\mathcal{X}, \mathcal{Y}$ be independent Poisson processes of intensity $n\mathbf{1}_Q$ and $\mathcal{L}_{\partial Q}$ the penalized boundary matching cost of [3, (18)] (denoted $L_{\partial Q}$ there). Then $\beta'_p(d) = \liminf_n n^{p/d-1}\mathbb{E}\mathcal{L}_{\partial Q}(\mathcal{X},\mathcal{Y})$ exists and*

$$\beta_p(d) = \beta'_p(d) = e_{\mathrm{Exp}}(p,d).$$

Consequently, if μ is a probability measure on $\mathbb{R}^d$ with absolutely continuous part $f\,dx$ and $\int |x|^\alpha d\mu < \infty$ for some $\alpha > 4dp/(d-2p)$, and $(X_i)_{i\ge1}$, $(Y_j)_{j\ge1}$ are independent with law μ, then almost surely

$$\lim_{n\to\infty} n^{p/d-1} \min_{\sigma\in\mathcal{S}_n} \sum_{i=1}^n |X_i - Y_{\sigma(i)}|^p = e_{\mathrm{Exp}}(p,d) \int_{\mathbb{R}^d} f^{1-p/d}\,dx. \tag{8.5}$$

Proof. Let M_n be the optimal matching cost of the two n-samples, so that $\mathbb{E}M_n = F^{\mathrm{cube}}_n$. Replacing one point changes M_n by at most $d^{p/2}$, so the Efron–Stein inequality gives $\operatorname{Var} M_n \le Cn$. Since $p < d/2$, $n^{p/d-1}n^{1/2} \to 0$, hence $n^{p/d-1}(M_n - F^{\mathrm{cube}}_n) \to 0$ in L^2, and by Theorem 8.3 $n^{p/d-1}M_n \to e_{\mathrm{Exp}}(p,d)$ in probability. The almost sure limit $\beta_p(d)$ of [3, Theorem 2] is therefore $e_{\mathrm{Exp}}(p,d)$.

For $p \ge 1$ the weight $2^{p-1} \wedge 1$ in [3, (18)] equals one, so $\mathcal{L}_{\partial Q}(X,Y)$ is the minimum, over matchings of a subset of X with a subset of Y of the same cardinality, of the sum of $|x-y|^p$ over matched pairs plus $\rho_1(z)^p$ over unmatched points. For samples of the same size this is the boundary assignment cost. Let $\mathcal{X}, \mathcal{Y}$ have N and N' points, $m = \min\{N, N'\}$, and let X^m, Y^m be the first m points of each sample for a labelling independent of the positions. Since $\rho_1 \le 1/2$, each unmatched point costs at most one, and adding or removing the $|N - N'|$ surplus points changes the optimal cost by at most $|N - N'|$ in each direction. Conditionally on (N, N'), X^m, Y^m are independent uniform samples, so $|\mathbb{E}\mathcal{L}_{\partial Q}(\mathcal{X},\mathcal{Y}) - \mathbb{E}F^{\mathrm{bd}}_m| \le \mathbb{E}|N - N'| \le Cn^{1/2}$. By Theorem 8.3, $F^{\mathrm{bd}}_k \le Ck^{1-p/d}$ and $k^{p/d-1}F^{\mathrm{bd}}_k \to e_{\mathrm{Exp}}(p,d)$; since $m/n \to 1$ in probability and N/n has bounded moments, $n^{p/d-1}\mathbb{E}F^{\mathrm{bd}}_m \to e_{\mathrm{Exp}}(p,d)$ by uniform integrability. Finally $n^{p/d-1}n^{1/2} \to 0$ because $p < d/2$.

Under the stated assumptions on μ, Barthe and Bordenave [3, Theorem 2] bound the upper limit in (8.5) by $\beta_p(d)\int f^{1-p/d}$ and the lower limit from below by $\beta'_p(d)\int f^{1-p/d}$, almost surely. Since $\beta_p(d) = \beta'_p(d)$, (8.5) follows. □

Remark 8.8. The common-count Poissonization in (8.3) is an exact representation of the expected fixed-size assignment cost, valid for every n and every $p \ge 1$. It differs from the comparison of two independent Poisson processes, whose counts disagree by order $n^{1/2}$: repairing this discrepancy costs $O(n^{1/2})$ in the unit cube, which is negligible against $n^{1-p/d}$ only when $p < d/2$. Besides the range of [3, Theorem 2], this and the variance

bound in the first paragraph are the only places where the proof of Corollary 8.7 uses $p < d/2$; it is not a limitation of Theorem 1.3 or Theorem 8.3. For $p = 2$ and $d \geq 5$ and compactly supported densities, (8.5) is [26, Corollary 4].

8.3. **Brownian interlacements.** We follow [23, Section 2]. Let W^* be the space of doubly infinite continuous trajectories in $\mathbb{R}^d$, $d \geq 3$, tending to infinity at $\pm\infty$, modulo time shift, and for compact $K \subset \mathbb{R}^d$ let $W^*_K \subset W^*$ be the set of trajectories that visit K. Brownian interlacements are given by a Poisson point process $\tilde{\omega} = \sum_i \delta_{(w_i^*, \alpha_i)}$ on $W^* \times [0, \infty)$ with intensity $\mathsf{n} \otimes d\alpha$, where n is the σ-finite measure of [23, Theorem 2.2]; by [23, (2.7) and (2.22)], $\mathsf{n}(W^*_K) = \mathrm{cap}(K)$. The occupation measure at level one is

$$\mathcal{L}_1 = \sum_{i:\alpha_i \leq 1} \int_{\mathbb{R}} \delta_{w_i(s)}\, ds$$

[23, (2.37)], a random locally finite measure with intensity dy [23, (2.39)].

Lemma 8.9. *Let $d \geq 3$. The random measure $\mathcal{L}_1$ satisfies (A1) and the concentration estimate (1.1) with $\alpha = 2$, which is (A2) when $d \geq 5$. Its law is invariant under $x \mapsto -x$ and mixing, hence ergodic, under translations. With positive probability $\mathcal{L}_1(B_1) = 0$.*

Proof. Invariance. The law of $\tilde{\omega}$ is invariant under translations and under linear isometries of $\mathbb{R}^d$ acting on trajectories [23, Proposition 2.4], and $\mathcal{L}_1$ is an equivariant function of $\tilde{\omega}$: translating or reflecting all trajectories translates or reflects $\mathcal{L}_1$. With the intensity dy this gives (A1) and reflection invariance.

Concentration. Huesmann et al. [14, Lemma 4.7] give (1.1) with $\alpha = 2$ for $q \geq 2$ (see also [21, Lemma 4.3]); the case $q \in [1, 2)$ follows by Hölder's inequality.

Mixing. For compact K, the restriction of $\mathcal{L}_1$ to K depends only on the restriction of $\tilde{\omega}$ to $W^*_K \times [0, 1]$, since a trajectory that does not visit K does not charge K. Let E_1, E_2 be events determined by $\mathcal{L}_1|_K$, and E_2^x the event that the translate of $\mathcal{L}_1$ by x lies in E_2; it is determined by the restriction of $\tilde{\omega}$ to $W^*_{K+x} \times [0, 1]$. The restrictions of $\tilde{\omega}$ to $(W^*_K \setminus W^*_{K+x}) \times [0, 1]$, $(W^*_{K+x} \setminus W^*_K) \times [0, 1]$ and $(W^*_K \cap W^*_{K+x}) \times [0, 1]$ are independent. Replacing $\tilde{\omega}$ by 0 on the third set changes $\mathbf{1}_{E_1}$ and $\mathbf{1}_{E_2^x}$ only on the event that $\tilde{\omega}$ charges it, and after this replacement they are independent. Hence

$$\left|\mathbb{P}(E_1 \cap E_2^x) - \mathbb{P}(E_1)\mathbb{P}(E_2)\right| \leq 4\, \mathsf{n}(W^*_K \cap W^*_{K+x}).$$

A trajectory visiting both K and $K' = K + x$ visits K' after its first visit to K, or visits K after its first visit to K'. By [23, (2.22)], the trajectories of W^*_K after their first visit to K have law P_{e_K}, Brownian motion started from the equilibrium measure e_K of K, of mass $\mathrm{cap}(K)$. Therefore

$$\mathsf{n}(W^*_K \cap W^*_{K'}) \leq \mathrm{cap}(K) \sup_{y \in K} P_y[H_{K'} < \infty] + \mathrm{cap}(K') \sup_{y \in K'} P_y[H_K < \infty],$$

where H denotes hitting times. Since Brownian motion is transient for $d \geq 3$ and $\mathrm{cap}(K') = \mathrm{cap}(K)$, both terms tend to zero as $|x| \to \infty$. Thus $\mathbb{P}(E_1 \cap E_2^x) \to \mathbb{P}(E_1)\mathbb{P}(E_2)$ for events determined by the restriction of $\mathcal{L}_1$ to a compact set, and by approximation for all events, which is mixing. The same argument is used by [20, Proposition 1.5] for the interlacement set.

Vacant balls. The support of $\mathcal{L}_1$ is the set of points visited by the trajectories with $\alpha_i \leq 1$ [23, (2.40)], and the probability that no such trajectory visits $\overline{B}_1$ is $e^{-\mathrm{cap}(\overline{B}_1)} > 0$ [23, (2.32)]. □

Corollary 8.10. *Let $d \geq 5$ and $1 < p < \infty$, and let $\nu = \mathcal{L}_1$ be the occupation measure of Brownian interlacements at level one. Then the optimal stationary coupling of dx and $\mathcal{L}_1$ is unique and induced by an equivariant map, and the periodic, ordinary and boundary specific costs converge, along all real L, to the same limit $e_{\mathcal{L}_1}(p, d) \in (0, \infty)$.*

Proof. By Lemma 8.9, (A1)–(A2) hold with $\alpha = 2 < d-2$, and Proposition 3.5(b) gives (1.8). Theorem 1.2 applies; $\mathcal{L}_1$ is not Lebesgue measure by the last assertion of Lemma 8.9. □

For $p = 2$, Corollary 8.10 is [14, Theorem 4.3], where the interlacement occupation measure is one of the main examples. Since $\mathcal{L}_1$ is diffuse, the case $p = 1$ is not covered: we do not know whether the optimal stationary coupling is unique.

Appendix A. Xu's averaging identity

For completeness we prove the identity used in Lemma 8.4, following [26, Theorem 2, Section 2.1 and Appendix A.1], for arbitrary integrable costs.

Let $(\mathsf{X}, \mathcal{A}, \mu)$ be a probability space, $n \geq 1$, and let $c_j : \mathsf{X} \to \mathbb{R}$ be measurable with $\int |c_j|\, d\mu < \infty$ for $j = 1, \dots, n$. For w in the simplex $\Delta_{n-1} = \{w_j \geq 0 : \sum_j w_j = 1\}$ let

$$\mathcal{T}_c(\mu, w) = \inf\Big\{ \int_{\mathsf{X}} \sum_{j=1}^{n} c_j \alpha_j \, d\mu : \ \alpha_j \geq 0,\ \sum_j \alpha_j = 1 \ \mu\text{-a.e.},\ \int \alpha_j \, d\mu = w_j \Big\}. \tag{A.1}$$

For $c_j = c(\cdot, y_j)$ this is the semi-discrete transport cost from μ to $\sum_j w_j \delta_{y_j}$. The function $w \mapsto \mathcal{T}_c(\mu, w)$ is finite and convex on the simplex (mixtures of admissible allocations are admissible). A finite convex function is continuous on the relative interior of each face of the simplex, so $\mathcal{T}_c(\mu, \cdot)$ is Borel.

Proposition A.1. *If $X_1, \dots, X_n$ are independent with law μ and $A \sim \mathrm{Dirichlet}(1, \dots, 1)$, then*

$$\mathbb{E} \min_{\sigma \in \mathcal{S}_n} \sum_{i=1}^{n} c_{\sigma(i)}(X_i) = n\, \mathbb{E}\, \mathcal{T}_c(\mu, A). \tag{A.2}$$

A cost matrix $(C_{ij})_{1 \leq i \leq N, 1 \leq j \leq n}$ is *generic* if every simple alternating cycle in the complete bipartite graph has nonzero alternating cost sum.

Lemma A.2. *Let $n \geq 2$, $k \geq 0$, $N = n - 1 + k$, and let (C_{ij}) be generic. For $I \subset \{1, \dots, N\}$ with $|I| = n - 1$, let $D_j(I)$ be the minimal cost of matching the sources in I bijectively with the targets other than j. For $i \notin I$ let $J_I(i) = \arg\min_j \{C_{ij} + D_j(I)\}$ and $\chi_j(I) = |\{i \notin I : J_I(i) = j\}|$. Then these minimizers are unique, and $I \mapsto (\chi_1(I), \dots, \chi_n(I))$ is a bijection onto $\{r \in \mathbb{Z}^n_{\geq 0} : \sum_j r_j = k\}$.*

Proof. Fix such an r. Consider the transportation problem with N sources of unit mass and target j of mass $r_j + \tau$, $\tau = (n-1)/n$; both sides have mass N. Let t_{ij} be optimal and choose dual variables u_i, q_j, $q_n = 0$, with

$$u_i \leq C_{ij} + q_j, \qquad t_{ij} > 0 \Rightarrow u_i = C_{ij} + q_j. \tag{A.3}$$

The graph of edges with $t_{ij} > 0$ contains every vertex and has no cycle, since a cycle would consist of dual-equality edges, contradicting genericity. It is connected: if a component contains b targets, its target mass $\sum r_j + b\tau$ equals its integer source mass, so $n \mid b$ since $\gcd(n-1, n) = 1$, whence $b = n$. Thus it is a spanning tree with $N + n - 1$ edges. Call a source split if it has at least two positive edges, and let s be their number. Counting edges, $N + s \leq N + n - 1$, so $s \leq n - 1$. If ℓ_j unsplit sources are sent to j, the remaining demand $r_j + \tau - \ell_j$ is nonnegative and of the form integer plus τ, hence at least τ; the split sources carry exactly this demand, so $s \geq n\tau = n - 1$. Hence $s = n - 1$, each split source has exactly two positive edges, and $\ell_j = r_j$.

Let I be the set of split sources. Contracting each of them to an edge yields a tree on the n targets. Rooting it at j and matching each split source to the child endpoint of its

edge assigns I bijectively to the targets other than j along dual-equality edges, so the dual bound is attained:

$$D_j(I) = \sum_{i\in I} u_i - \sum_{\ell\neq j} q_\ell, \qquad D_j(I) - D_n(I) = q_j. \tag{A.4}$$

For an unsplit source i, its positive edge is its unique dual-equality edge, since another one would close a cycle with the spanning tree. By (A.4), minimizing $C_{ij} + D_j(I)$ is minimizing $C_{ij} + q_j$, so $J_I(i)$ is the target of its transport edge and $\chi(I) = \ell = r$. This proves surjectivity. For every base I, two minimizers $J_I(i)$ would give two optimal full matchings of $I \cup \{i\}$; their symmetric difference is a union of alternating cycles, and swapping either matching along one of them cannot lower its cost, so that cycle has alternating sum both ≥ 0 and ≤ 0, i.e. zero, contradicting genericity. Finally there are $\binom{N}{n-1} = \binom{n+k-1}{n-1}$ bases and as many weak compositions of k into n parts, so the surjection is a bijection. □

Proof of Proposition A.1. For $n = 1$ both sides equal $\int c_1\, d\mu$. Let $n \geq 2$ and assume first that $c_j(X) - c_\ell(X)$ has a nonatomic law for $j \neq \ell$; then all cost matrices obtained from independent sources are generic almost surely (condition on all sources of a given cycle but one).

Condition on $X_1, \dots, X_{n-1}$, let D_j be the minimal cost of matching them to all targets but j, and let $j_*(x)$ be the μ-a.e. unique minimizer of $c_j(x) + D_j$, which is the label given to an additional source at x by an optimal full matching. Let $A_j = \mu\{j_* = j\}$. For any admissible α with masses A, $\int c_{j_*}\, d\mu = \int \min_j(c_j + D_j)\, d\mu - \sum_j D_j A_j \leq \int \sum_j c_j \alpha_j\, d\mu$, so $\mathcal{T}_c(\mu, A) = \int c_{j_*}\, d\mu$. By genericity the optimal assignment σ^* of $X_1, \dots, X_n$ is a.s. unique (two optimal assignments would produce a zero-cost alternating cycle, as in the proof of Lemma A.2), and it is equivariant under permutations of the sources, so the pairs $(X_i, \sigma^*(i))$ have the same law and the expected total is $n\mathbb{E}c_{\sigma^*(n)}(X_n)$. Since the optimum equals $\min_j(c_j(X_n) + D_j)$ with a unique minimizer, $\sigma^*(n) = j_*(X_n)$, and $\mathbb{E}[c_{\sigma^*(n)}(X_n) \mid X_1, \dots, X_{n-1}] = \mathcal{T}_c(\mu, A)$. Hence

$$\mathbb{E}\min_\sigma \sum_i c_{\sigma(i)}(X_i) = n\,\mathbb{E}\,\mathcal{T}_c(\mu, A). \tag{A.5}$$

To identify the law of A, fix $k \geq 1$ and add k independent sources, each choosing a label by the rule j_*. Conditionally on the base sample the count vector is multinomial with parameters $(k; A)$. On the other hand, regarding the $N = n - 1 + k$ sources as exchangeable and choosing a base uniformly among the $\binom{N}{n-1}$ subsets, Lemma A.2 shows that the count vector is uniform on the weak compositions of k. Hence, for $\sum_j r_j = k$,

$$\mathbb{E}\prod_j A_j^{r_j} = \frac{(n-1)!\prod_j r_j!}{(n+k-1)!}, \tag{A.6}$$

which are the moments of the uniform law on Δ_{n-1}. Moments determine laws on the compact simplex, so $A \sim \mathrm{Dirichlet}(1, \dots, 1)$.

For general costs, enlarge the space by independent uniform $Z \in [0,1]^n$ and replace c_j by $c_j + \epsilon z_j$. The assignment minimum increases by at most $n\epsilon$. If $\mathcal{T}_\epsilon(w)$ denotes the cost (A.1) on $\mathsf{X} \times [0,1]^n$ for the costs $c_j + \epsilon z_j$, then $\mathcal{T}_c(\mu, w) \leq \mathcal{T}_\epsilon(w) \leq \mathcal{T}_c(\mu, w) + \epsilon$ (average over z for the lower bound, lift an allocation independently of z for the upper bound). Let $\epsilon \downarrow 0$. □

Remark A.3. If $E_1, \dots, E_n$ are independent $\mathrm{Exp}(1)$, $\Sigma = \sum_j E_j$ and $A_j = E_j/\Sigma$, the change of variables $(E_j) = (\Sigma A_j)$ turns the joint density into $e^{-\Sigma}\Sigma^{n-1}$ on $(0,\infty) \times \Delta_{n-1}$, which factors into a $\Gamma(n,1)$ density for Σ and the uniform density $(n-1)!$ on the simplex for A. Thus Σ and A are independent, $\mathbb{E}\Sigma = n$ and $A \sim \mathrm{Dirichlet}(1, \dots, 1)$.

Declaration on the use of AI

The starting point of this work was the preprint of [16], combined with the recent bridge between bipartite and semi-discrete matching due to [26] we then prompted the LLMs to use these as a starting point to extend the result from [14] to every $p \geq 1$ and also to the bi-partite problem, which they did relatively quickly. We checked the proofs at a relatively detailed level and prompted further the models for simplifications and extensions. The authors take full responsibility for the content of the paper and will provide a digested version in the near future.

References

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

[2] L. Ambrosio, F. Stra, and D. Trevisan. A PDE approach to a 2-dimensional matching problem. *Probability Theory and Related Fields* **173** (2019), 433–477.

[3] F. Barthe and C. Bordenave. Combinatorial optimization over two random point sets. In *Séminaire de Probabilités XLV*, Lecture Notes in Mathematics 2078, Springer, 2013, 483–535. arXiv:1103.2734v2; equation and lemma numbers refer to this version.

[4] E. Caglioti, M. Goldman, F. Pieroni, and D. Trevisan. Subadditivity and optimal matching of unbounded samples. *Annales de la Faculté des Sciences de Toulouse: Mathématiques* **35** (2026), no. 3, 697–742. doi:10.5802/afst.1861; arXiv:2407.06352.

[5] S. Caracciolo, C. Lucibello, G. Parisi, and G. Sicuro. Scaling hypothesis for the Euclidean bipartite matching problem. *Physical Review E* **90** (2014), no. 1, 012118.

[6] S. Dereich, M. Scheutzow, and R. Schottstedt. Constructive quantization: approximation by empirical measures. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **49** (2013), no. 4, 1183–1203. arXiv:1108.5346.

[7] V. Dobrić and J. E. Yukich. Asymptotics for transportation cost in high dimensions. *Journal of Theoretical Probability* **8** (1995), 97–118.

[8] A. Figalli and N. Gigli. A new transportation distance between non-negative measures, with applications to gradient flows with Dirichlet boundary conditions. *Journal de Mathématiques Pures et Appliquées (9)* **94** (2010), no. 2, 107–130.

[9] W. Gangbo and R. J. McCann. The geometry of optimal transportation. *Acta Mathematica* **177** (1996), no. 2, 113–161.

[10] M. Goldman, L. Koch, and D. Trevisan. Large-scale regularity theory for optimal p-transport maps and application to concentration in the matching problem. arXiv:2611.25660v1, 19 November 2026. https://arxiv.org/abs/2611.25660.

[11] M. Goldman and D. Trevisan. Convergence of asymptotic costs for random Euclidean matching problems. *Probability and Mathematical Physics* **2** (2021), no. 2, 341–362. doi:10.2140/pmp.2021.2.341; arXiv:2009.04128.

[12] M. Goldman and D. Trevisan. Optimal transport methods for combinatorial optimization over two random point sets. *Probability Theory and Related Fields* (2023), 1–70. arXiv:2209.14615.

[13] M. Huesmann. Optimal transport between random measures. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **52** (2016), no. 1, 196–232. doi:10.1214/14-AIHP634; arXiv:1206.3672, whose numbering we use.

[14] M. Huesmann, M. Goldman, and D. Trevisan. Asymptotics for random quadratic transportation costs. arXiv:2409.08612v2, 17 February 2025. https://arxiv.org/abs/2409.08612.

[15] M. Huesmann and K.-T. Sturm. Optimal transport from Lebesgue to Poisson. *Annals of Probability* **41** (2013), no. 4, 2426–2478. arXiv:1012.3845v2; theorem and lemma numbers refer to this version. doi:10.1214/12-AOP814.

[16] M. Huesmann and A. Wachtel. Asymptotic transportation cost of the optimal matching subject to different boundary conditions. arXiv:2611.17066v1, 12 November 2026. https://arxiv.org/abs/2611.17066.
[17] O. Kallenberg. *Random Measures, Theory and Applications*. Probability Theory and Stochastic Modelling 77, Springer, Cham, 2017.
[18] U. Krengel. *Ergodic Theorems*. de Gruyter Studies in Mathematics 6, Walter de Gruyter, Berlin, 1985.
[19] G. Last and M. Penrose. *Lectures on the Poisson Process*. Institute of Mathematical Statistics Textbooks 7, Cambridge University Press, Cambridge, 2018.
[20] X. Li. Percolative properties of Brownian interlacements and its vacant set. arXiv:1610.08204v2, 18 September 2019.
[21] M. Mariani and D. Trevisan. Wasserstein asymptotics for Brownian motion on the flat torus and Brownian interlacements. arXiv:2307.10325v2, 15 February 2024.
[22] B. S. Mityagin. The zero set of a real analytic function. *Mathematical Notes* **107** (2020), no. 3–4, 529–530. arXiv:1512.07276.
[23] A.-S. Sznitman. On scaling limits and Brownian interlacements. *Bulletin of the Brazilian Mathematical Society, New Series* **44** (2013), no. 4, 555–592. arXiv:1209.4531; equation numbers refer to this version.
[24] M. Talagrand. Matching random samples in many dimensions. *The Annals of Applied Probability* **2** (1992), no. 4, 846–856.
[25] C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften 338, Springer, Berlin, 2009.
[26] Y. Xu. Random matching between distinct distributions in two dimensions. arXiv:2610.40509v1, 29 October 2026. https://arxiv.org/abs/2610.40509.

Univ Paris Est Creteil, CNRS, LAMA, F-94010 Creteil, France
Email address: michael.goldman@cnrs.fr
URL: https://orcid.org/0000-0002-9523-9721

Institut für Mathematische Stochastik, Universität Münster, Orléans-Ring 10, 48149 Münster, Germany
Email address: martin.huesmann@uni-muenster.de

Dipartimento di Matematica, Università di Pisa, Largo Bruno Pontecorvo 5, 56127 Pisa, Italy
Email address: dario.trevisan@unipi.it
URL: https://orcid.org/0000-0002-4563-5638

Institut für Mathematische Stochastik, Universität Münster, Orléans-Ring 10, 48149 Münster, Germany
Email address: arthur.wachtel@uni-muenster.de