The Annals of Probability
2013, Vol. 41, No. 4, 2426–2478
DOI: 10.1214/12-AOP814

OPTIMAL TRANSPORT FROM LEBESGUE TO POISSON

BY MARTIN HUESMANN AND KARL-THEODOR STURM

University of Bonn

This paper is devoted to the study of couplings of the Lebesgue measure and the Poisson point process. We prove existence and uniqueness of an optimal coupling whenever the asymptotic mean transportation cost is finite. Moreover, we give precise conditions for the latter which demonstrate a sharp threshold at $d = 2$. The cost will be defined in terms of an arbitrary increasing function of the distance.

The coupling will be realized by means of a transport map ("allocation map") which assigns to each Poisson point a set ("cell") of Lebesgue measure 1. In the case of quadratic costs, all these cells will be convex polytopes.

1. Introduction and statement of main results. (a) The theory of *optimal transportation* studies couplings between two probability measures λ and ν on $\mathbb{R}^d$ which minimize the total transportation cost. A coupling is interpreted as a plan how to transport λ into ν. Transporting a unit of mass from a to b produces cost of amount $c(a, b)$, where $c(\cdot, \cdot)$ is a given cost function. Of particular interest are couplings q which are induced by transport maps, that is, $q = (\mathrm{id}, \psi)_*\lambda$ for some map $\psi : \mathbb{R}^d \to \mathbb{R}^d$ with $\psi_*\lambda = \nu$.

A *fair allocation* for a simple point process in $\mathbb{R}^d$ is a coupling of the Lebesgue measure $\mathfrak{L}$ and the point process $\mu^\bullet$ induced by a transport map, that is, there is a map $\Psi : \Omega \times \mathbb{R}^d \to \mathbb{R}^d$ such that for $\mathbb{P}$-almost every $\omega \in \Omega$ the map $\Psi^\omega : \mathbb{R}^d \to \mathbb{R}^d$ transports the Lebesgue measure into the point process: $\Psi^\omega_* \mathfrak{L} = \mu^\omega$. Such an allocation is called *factor allocation* if it is a measurable function of the point process (i.e., it measurably depends only on the given point process).

In this article we connect these two theories by constructing fair allocations between the Lebesgue measure and point processes using tools from optimal transportation. Instead of considering the total transportation cost we ask for minimizers of the *cost per unit mass*. Good estimates on the transportation cost will directly imply good tail estimates for the distribution of the transport distance.

Moreover, the techniques developed in this article allow us to construct a *fair factor allocation* with the best possible tail estimate and also to derive new estimates on the transportation cost between the Lebesgue measure and a Poisson point process.

We now describe our results in more detail.

Received April 2011; revised August 2012.
MSC2010 subject classifications. Primary 60D05; secondary 52A22, 49Q20.
Key words and phrases. Optimal transportation, fair allocation, Laguerre tessellation, Poisson point process.

(b) A point process $\mu^\bullet : \Omega \to \mathcal{N}(\mathbb{R}^d)$ is a random variable with values in the space of integer valued Radon measure. Put $\Xi(\omega) = \mathrm{supp}(\mu^\omega)$. Then, $\mu^\bullet$ has the representation $\mu^\bullet : \omega \mapsto \mu^\omega = \sum_{\xi \in \Xi(\omega)} k(\xi) \cdot \delta_\xi$ with $k(\xi) \in \mathbb{N}$. $\mu^\bullet$ is called equivariant if for all Borel sets $A \in \mathcal{B}(\mathbb{R}^d)$ we have $\mu^{\omega+z}(A+z) = \mu^\omega(A)$. Here, we interprete $\omega + z$ as the support of μ^ω translated by z; see Section 2.2.

Given an equivariant point process $\mu^\bullet : \omega \mapsto \mu^\omega = \sum_{\xi \in \Xi(\omega)} k(\xi) \cdot \delta_\xi$ on $\mathbb{R}^d$ with unit intensity, we consider the set Π of all *couplings* $q^\bullet$ of the Lebesgue measure $\mathfrak{L}$ and the point process—that is, the set of measure-valued random variables $\omega \mapsto q^\omega$ s.t. for a.e. ω the measure q^ω on $\mathbb{R}^d \times \mathbb{R}^d$ is a coupling of $\mathfrak{L}$ and μ^ω—and we ask for a minimizer of the *asymptotic mean cost functional*

$$\mathfrak{C}_\infty(q^\bullet) := \liminf_{n\to\infty} \frac{1}{\mathfrak{L}(B_n)} \mathbb{E}\left[\int_{\mathbb{R}^d \times B_n} \vartheta\big(|x-y|\big)\, dq^\bullet(x,y)\right].$$

Here $B_n := [0, 2^n)^d \subset \mathbb{R}^d$. The *scale* $\vartheta : \mathbb{R}_+ \to \mathbb{R}_+$ will always be some strictly increasing, continuous function with $\vartheta(0) = 0$ and $\lim_{r\to\infty} \vartheta(r) = \infty$.

A coupling $\omega \mapsto q^\omega$ of the Lebesgue measure and the point process is called *optimal* if it minimizes the asymptotic mean cost functional and if it is *equivariant* in the sense that $q^{\omega+z}(A+z, B+z) = q^\omega(A,B)$ for all $z \in \mathbb{R}^d$ and Borel sets $A, B \in \mathcal{B}(\mathbb{R}^d)$. Our main result states the following:

THEOREM 1.1. *If the asymptotic mean transportation cost*

$$\mathfrak{c}_\infty := \liminf_{n\to\infty} \inf_{q^\bullet \in \Pi} \frac{1}{\mathfrak{L}(B_n)} \mathbb{E}\left[\int_{\mathbb{R}^d \times B_n} \vartheta\big(|x-y|\big)\, dq^\bullet(x,y)\right] \tag{1}$$

is finite, then there exists a unique optimal coupling of the Lebesgue measure and the point process $\mu^\bullet$.

(c) The unique optimal coupling q^ω can be represented as $(\mathrm{id}, T^\omega)_* \mathfrak{L}$ for some map $T^\omega : \mathbb{R}^d \to \mathrm{supp}(\mu^\omega) \subset \mathbb{R}^d$ measurably only dependent on the sigma algebra generated by the point process. In other words, T^ω defines a *fair factor allocation*. Its inverse map assigns to each point ξ of the point process ("center") a set ("cell") of Lebesgue measure $\mu^\omega(\xi) \in \mathbb{N}$. If the point process is simple, then all these cells have volume 1. In the case of quadratic cost, that is, $\vartheta(r) = r^2$, the cells will be convex polytopes. The transport map will be given as $T^\omega = \nabla\varphi^\omega$ for some convex function $\varphi^\omega : \mathbb{R}^d \to \mathbb{R}$ and induces a Laguerre tessellation; see [18].

In the case $\vartheta(r) = r$ the transportation map induces a Johnson–Mehl diagram; see [5]. For the many results on and applications of these tessellations see the references in [18] and [5]. In the light of these results one might interpret the optimal coupling as a generalized tessellation.

(d) As a particular corollary to Theorem 1.1 we conclude that $\mathfrak{c}_\infty = \inf_{q^\bullet \in \Pi} \mathfrak{C}_\infty(q^\bullet)$ and that the infimum is always attained; more precisely, it is attained by an equivariant coupling $q^\bullet$. For equivariant couplings $q^\bullet$ the mean cost functional $\frac{1}{\mathfrak{L}(A)}\mathbb{E}[\int_{\mathbb{R}^d\times A}\vartheta(|x-y|)\,dq^\bullet(x,y)]$, however, is independent of $A\subset\mathbb{R}^d$. Hence,

$$\mathfrak{c}_\infty=\inf_{q^\bullet\in\Pi_{\text{eqv}}}\mathbb{E}\left[\int_{\mathbb{R}^d\times[0,1)^d}\vartheta(|x-y|)\,dq^\bullet(x,y)\right],$$

where Π_{eqv} now denotes the set of all equivariant couplings of the Lebesgue measure and the point process.

Moreover, for equivariant couplings, $\mathbb{E}[\vartheta(|x-T^\bullet(x)|)]$ the mean cost of transportation of a Lebesgue point x to the center of its cell is independent of $x\in\mathbb{R}^d$. Hence,

$$\mathfrak{c}_\infty=\inf_{T^\bullet}\mathbb{E}[\vartheta(|0-T^\bullet(0)|)],\tag{2}$$

where the infimum is taken over all equivariant maps $T:\mathbb{R}^d\times\Omega\to\mathbb{R}^d$ with $T^\omega{}_*\mathfrak{L}=\mu^\omega$ for a.e. ω. And again: the infimum is attained by a unique such T. Let us point out that identity (2) allows us to resolve the asymmetry in the integration domain in equation (1): we equally well may replace the domain of integration $\mathbb{R}^d\times B_n$ by $B_n\times\mathbb{R}^d$.

(e) Analogous results will be obtained in the more general case of optimal "semicouplings" between the Lebesgue measure and point processes of "subunit" intensity.

We develop the theory of optimal semicouplings as a concept of independent interest. Optimal semicouplings are solutions of a twofold optimization problem: the optimal choice of a density $\rho\le 1$ of the first marginal μ_1 and subsequently the optimal choice of a coupling between $\rho\mu_1$ and μ_2. This twofold optimization problem can also be interpreted as a transport problem with free boundary values; see Figure 1.

Given a point process of subunit intensity and finite mean transportation cost, we prove that there exists a unique optimal semicoupling between the Lebesgue measure and the point process. It can be represented on $\mathbb{R}^d\times\mathbb{R}^d$ as before as

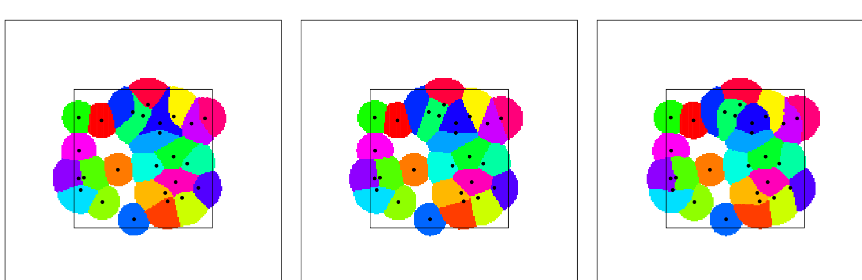

FIG. 1. *Optimal semicoupling of Lebesgue and* 25 *points in the cube with cost function* $c(x,y)=|x-y|^p$ *and (from left to right)* $p=1,2,4$, *respectively.*

$q^\omega = (\mathrm{id}, T^\omega)_* \mathfrak{L}$ in terms of a transport map $T^\omega : \mathbb{R}^d \to \mathrm{supp}[\mu^\omega] \cup \{\eth\}$ where $\eth$ now denotes an isolated point ("cemetery") added to $\mathbb{R}^d$.

(f) In any case, we prove that the unique transport map T^ω can be obtained as the limit of a suitable sequence of transport maps which solve the optimal transportation problem between the Lebesgue measure and the point process restricted to bounded sets.

More precisely, for $z \in \mathbb{Z}^d$ and $\gamma \in \Gamma := (\{0,1\}^d)^{\mathbb{N}}$ consider the "doubling sequence" of cubes

$$B_n(z,\gamma) = z - \sum_{k=1}^{n} 2^{k-1}\gamma_k + [0, 2^n)^d.$$

Note that the cube $B_n(z,\gamma)$ is one of the subcubes obtained by subdividing $B_{n+1}(z,\gamma)$ into 2^d cubes of half edge length. Let $T_{z,n}(\cdot,\omega,\gamma) : \mathbb{R}^d \to \mathrm{supp}[\mu^\omega] \cup \{\eth\}$ be the transport map for the unique optimal semicoupling between $\mathfrak{L}$ and $1_{B_n(z,\gamma)} \cdot \mu^\omega$, that is, for the optimal transport of an optimal "submeasure" $\rho^\omega \cdot \mathfrak{L}$ to the point process restricted to the cube $B_n(z,\gamma)$.

THEOREM 1.2. *For every $z \in Z^d$ and every bounded Borel set $M \subset \mathbb{R}^d$,*

$$\lim_{n\to\infty} (\mathfrak{L} \otimes \mathbb{P} \otimes \nu)\big(\{(x,\omega,\gamma) \in M \times \Omega \times \Gamma : T_{z,n}(x,\omega,\gamma) \neq T(x,\omega)\}\big) = 0,$$

where ν denotes the Bernoulli measure on Γ.

(g) If $\mu^\bullet$ is a Poisson point process with intensity $\beta \le 1$ we have rather sharp estimates for the asymptotic mean transportation cost to be finite.

THEOREM 1.3. (i) *Assume $d \ge 3$ (and $\beta \le 1$) or $\beta < 1$ (and $d \ge 1$). Then there exists a constant $0 < \kappa < \infty$ s.t.*

$$\limsup_{r\to\infty} \frac{\log \vartheta(r)}{r^d} < \kappa \quad \Longrightarrow \quad \mathfrak{c}_\infty < \infty \quad \Longrightarrow \quad \liminf_{r\to\infty} \frac{\log \vartheta(r)}{r^d} \le \kappa.$$

(ii) *Assume $d \le 2$ and $\beta = 1$. Then for any concave $\hat{\vartheta} : [1,\infty) \to \mathbb{R}$ dominating ϑ*

$$\int_1^\infty \frac{\hat{\vartheta}(r)}{r^{1+d/2}}\, dr < \infty \quad \Longrightarrow \quad \mathfrak{c}_\infty < \infty \quad \Longrightarrow \quad \liminf_{r\to\infty} \frac{\vartheta(r)}{r^{d/2}} = 0.$$

The first implication in assertion (ii) is new. Assertion (i) in the case $\beta = 1$ is due to Holroyd and Peres [16], based on a fundamental result of Talagrand [28]. The first implication in assertion (i) in the case $\beta < 1$ was proven by Hoffman, Holroyd and Peres [13]. The second implication in assertion (ii) is due to [14].

Now let us consider the particular case of L^p transportation cost, that is, $\vartheta(r) = r^p$.

COROLLARY 1.4. (i) *For all $d \in \mathbb{N}$, all $\beta \le 1$ and $p \in (0,\infty)$ the asymptotic mean L^p-transportation cost $\mathfrak{c}_\infty$ is finite if and only if*

$$p < \overline{p} := \begin{cases} \infty, & \text{for } d \ge 3 \text{ or } \beta < 1; \\ \dfrac{d}{2}, & \text{for } d \le 2 \text{ and } \beta = 1. \end{cases}$$

(ii) *If $\beta = 1$, then for all $p \in (0,\infty)$ there exist constants $0 < k \le k' < \infty$ s.t. for all $d > 2(p \wedge 1)$*

$$k \cdot d^{p/2} \le \mathfrak{c}_\infty \le k' \cdot d^{p/2}.$$

(h) The study of fair allocations for point processes is an important and hot topic of current research; see, for example, [15, 16, 29] and references therein. A landmark contribution was the construction of the *stable marriage* between Lebesgue measure and an ergodic translation invariant simple point process [13]. One of the challenges is to produce allocations with fast decay of the distance of a typical point in a cell to its center or of the diameter of the cell. The *gravitational allocation* [8, 9] in $d \ge 3$ was the first allocation with exponential decay. Moreover, all the cells are connected and contain their center. However, the decay was not yet as good as the decay of a *random allocation* constructed in [16].

On the other hand, during the last decade the theory of optimal transportation (see, e.g., [25, 31]) has attracted lot of interest and has produced an enormous amount of deep results, striking applications and stimulating new developments, among others in PDEs (e.g., [3, 7, 23]), evolution semigroups (e.g., [4, 22, 24]) and geometry (e.g., [19, 21, 26, 27, 32]). Ajtai, Komlós and Tusnády as well as Talagrand and others studied the problem of matchings and allocation of independently distributed points in the unit cube in terms of transportation cost ([1, 28] and references therein). For further studies of invariant transports between random measures in more general spaces we refer to [17][1].

(i) In all the optimal transportation problems considered in the aforementioned contributions, however, the marginals have finite total mass. Our paper seems to be the first to prove existence and uniqueness of a solution to an optimal transportation problems for which the total transportation cost is infinite.

More precisely, the main contributions of the current paper are:

- We present a concept of "optimality" for (semi-) couplings between the Lebesgue measure and a point process.
- We prove existence and uniqueness of an optimal semicoupling whenever there exists a semicoupling with finite asymptotic mean transportation cost.

[1] In the course of the refereeing process of this paper a construction of a fair allocation for the Poisson point process with optimal tail behavior of the diameter of a typical cell was presented by Markó and Timar [20] using the algorithm of Ajtai, Komlós and Tusnády.

- We prove that for a.e. doubling sequence of boxes $(B_n(z,\gamma))_{n\in\mathbb{N}}$ the sequence of optimal semicouplings $q^\bullet_{n,z,\gamma}$ between the Lebesgue measure and the point process restricted to the box $B_n(z,\gamma)$ will converge. More precisely, the sequence $q^\bullet_{n,z,\gamma}$ will converge as $n\to\infty$ toward a unique optimal semicoupling $q^\bullet$ between the Lebesgue measure and the point process.
- We prove that the asymptotic mean transportation cost for the Poisson point process in $d\leq 2$ is finite for L^p-costs with $p<d/2$ and also for more general scale functions like $\vartheta(r)=r^{d/2}\cdot\frac{1}{(\log r)^\alpha}$ with $\alpha>1$.

1.1. *Outline*. The article is divided into five parts. The core material with the proofs of the main theorems is contained in Sections 3 to 5. These three sections are rather independent of each other.

In Section 2 we start by recalling the relevant definitions and objects we work with. We also state an importation technical result, Theorem 2.1, the existence and uniqueness result of optimal semicouplings on bounded sets. The proof of this theorem is deferred to Section 6 because it is a purely deterministic result on transportation problems between *finite* measures whereas the rest of the article deals with transportation problems between random measures with *infinite* mass. The key idea for the proof is to show that every minimizer has to be concentrated on a certain graph. Then, existence can be shown via lower semicontinuity plus compactness. Uniqueness follows from the observation that a convex combination of optimal semicouplings can only be concentrated on a graph if all optimal semicouplings are concentrated on the same graph.

In Section 3 we proof the uniqueness part of Theorem 1.1. The idea for the proof is again to show that every optimal semicoupling has to be concentrated on the graph of some function. To this end, we introduce the concept of local optimality. A semicoupling $q^\bullet$ is called locally optimal if and only if for $\mathbb{P}$-almost all ω the restriction of q^ω to any bounded Borel set A, $1_{\mathbb{R}^d\times A}q^\omega$ is optimal between its marginals in the classical sense. Using equivariance, we show that every optimal semicoupling is locally optimal. Hence, by applying Theorem 2.1 we get the existence of a transportation map and therefore uniqueness.

The proof of the existence part of Theorem 1.1 is presented in the first part of Section 4. The idea is to approximate the optimal semicoupling by solutions to classical optimal transportation problems on bounded regions. The main problem to overcome is to control the contribution of a small fixed observation window to the total asymptotic mean transportation cost. The solution is not to consider a deterministic exhausting sequence of cubes, but a random sequence of cubes. This second randomization causes a symmetrization and induces tightness of this sequence. It could also be seen as a way to enforce the equivariance of the limiting measure. The uniqueness of optimal semicouplings then allows us to remove the second randomization again and also to deduce "quenched" results in the second part of Section 4 which finally proves Theorem 1.2.

In Section 5, we prove Theorem 1.3. The estimates are based on an explicit construction of a semicoupling between $\mathfrak{L}$ and $1_{[0,2^n)^d}\mu^\bullet$. The transportation cost estimate can thereby be reduced to the estimates of moments, central moments and inverse moments of Poisson random variables. The advantage of this approach is that it allows us to get fairly reasonable estimates of constants and, more importantly, it is also potentially applicable to other cases of interest.

2. Set-up and basic concepts. $\mathfrak{L}$ will always denote the Lebesgue measure on $\mathbb{R}^d$. The complement of a set $A \subset \mathbb{R}^d$ will be denoted by $\complement A$. The push forward of a measure ρ by a map S will be denoted by $S_*\rho$.

2.1. *Couplings and semicouplings.* For each Polish space X (i.e., separable, complete metrizable space) the set of measures on X—equipped with its Borel σ-field—will be denoted by $\mathcal{M}(X)$. Given any ordered pair of Polish spaces X, Y and measures $\lambda \in \mathcal{M}(X)$, $\mu \in \mathcal{M}(Y)$, we say that a measure $q \in \mathcal{M}(X \times Y)$ is a *semicoupling* of λ and μ, briefly $q \in \Pi_s(\lambda, \mu)$, if and only if the (first and second, resp.) marginals satisfy

$$(\pi_1)_* q \le \lambda, \qquad (\pi_2)_* q = \mu,$$

that is, if and only if $q(A \times Y) \le \lambda(A)$ and $q(X \times B) = \mu(B)$ for all Borel sets $A \subset X$, $B \subset Y$. The semicoupling q is called *coupling*, briefly $q \in \Pi(\lambda, \mu)$, if and only if, in addition,

$$(\pi_1)_* q = \lambda.$$

Existence of a coupling requires that the measures λ and μ have the same total mass. If the total masses of λ and μ are finite and equal, then the "renormalized" product measure $q = \frac{1}{\lambda(X)} \lambda \otimes \mu$ is always a coupling of λ and μ.

If λ and μ are Σ-finite, that is, $\lambda = \sum_{n=1}^\infty \lambda_n$, $\mu = \sum_{n=1}^\infty \mu_n$ with finite measures $\lambda_n \in \mathcal{M}(X)$, $\mu_n \in \mathcal{M}(Y)$—which is the case for all Radon measures—and if both of them have infinite total mass, then there always exists a Σ-finite coupling of them. [Indeed, then the λ_n and μ_n can be chosen to have unit mass and $q = \sum_n (\lambda_n \otimes \mu_n)$ does the job.]

See also [11] for the related concept of *partial coupling*.

2.2. *Point processes.* Throughout this paper, $\mu^\bullet$ will denote an equivariant point process of subunit intensity, modeled on some probability space $(\Omega, \mathfrak{A}, \mathbb{P})$. For convenience, we will assume that Ω is a compact separable metric space and $\mathfrak{A}$ its completed Borel field. These technical assumptions are only made to simplify the presentation.

Recall that a *point process* is a measurable map $\mu^\bullet : \Omega \to \mathcal{M}(\mathbb{R}^d)$, $\omega \mapsto \mu^\omega$ with values in the subset $\mathcal{N}(\mathbb{R}^d)$ of locally finite *counting measures* on $\mathbb{R}^d$. It is a particular example of a random measure, characterized by the fact that $\mu^\omega(A) \in \mathbb{N}_0$ for $\mathbb{P}$-a.e. ω and every bounded Borel set $A \subset \mathbb{R}^d$. It can always be written as

$$\mu^\omega = \sum_{\xi \in \Xi(\omega)} k(\xi) \delta_\xi$$

with some countable set $\Xi(\omega) \subset \mathbb{R}^d$ without accumulation points and with numbers $k(\xi) \in \mathbb{N}$. The point process is called *simple* if and only if $k(\xi) = 1$ for all $\xi \in \Xi(\omega)$ and a.e. ω or, in other words, if and only if $\mu(\{x\}) \in \{0, 1\}$ for every $x \in \mathbb{R}^d$ and a.e. ω.

We assume that the probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ admits a measurable flow $\theta : \mathbb{R}^d \times \Omega \to \Omega$ such that the point process $\mu^\bullet$ is $\mathbb{R}^d$*-equivariant* or just equivariant, that is,

$$\mu^{\theta_z(\omega)}(A + z) = \mu^\omega(A)$$

for all Borel sets $A \in \mathcal{B}(\mathbb{R}^d)$. Moreover, we assume that $\mathbb{P}$ is stationary, that is, invariant under the flow

$$\mathbb{P} \circ \theta = \mathbb{P}.$$

In particular, this implies that $\mu^\bullet$ is *translation invariant* in the usual sense, that is,

$$(\tau_z)_* \mu^\bullet \overset{(\mathrm{d})}{=} \mu^\bullet$$

for each $z \in \mathbb{R}^d$. We interpret the flow as a shift of the support of $\mu^\bullet$ and therefore write $\theta_z(\omega) = \omega + z$; see also Example 2.1 of [17].

To split the translation invariance into equivariance and stationarity has the huge advantage that equivariance is stable under addition whereas translation invariance is not. It is not really a restriction as we can always take the canonical realization as a probability space; again see Example 2.1 of [17].

We say that $\mu^\bullet$ has *subunit intensity* if and only if $\mathbb{E}[\mu^\bullet(A)] \le \mathfrak{L}(A)$ for all Borel sets $A \subset \mathbb{R}^d$. If "$=$" holds instead of "$\le$" we say that $\mu^\bullet$ has *unit intensity*. A translation invariant point process has subunit (or unit) intensity if and only if its intensity

$$\beta = \mathbb{E}[\mu^\bullet([0, 1)^d)]$$

is ≤ 1 (or $= 1$, resp.).

Given a point process $\mu^\bullet$, the measure $d(\mu^\bullet \mathbb{P})(y, \omega) := d\mu^\omega(y)\, d\mathbb{P}(\omega)$ on $\mathbb{R}^d \times \Omega$ is called *Campbell measure* of the random measure $\mu^\bullet$.

The most important example of an equivariant simple point process is the *Poisson point process* or *Poisson random measure* with intensity $\beta \le 1$. It is characterized by:

- for each Borel set $A \subset \mathbb{R}^d$ of finite volume the random variable $\omega \mapsto \mu^\omega(A)$ is Poisson distributed with parameter $\beta \cdot \mathfrak{L}(A)$, and
- for disjoint Borel sets $A_1, \dots, A_k \subset \mathbb{R}^d$ the family of random variables $\mu^\omega(A_1), \dots, \mu^\omega(A_k)$ is independent.

There are some instances in which we need additional assumptions on $\mu^\bullet$ (e.g., ergodicity, unit intensity). In each of these cases we will clearly point out the specific assumptions we make.

2.3. *Couplings of Lebesgue measure and the point process.* A (semi-) coupling of the Lebesgue measure $\mathfrak{L} \in \mathcal{M}(\mathbb{R}^d)$ and the point process $\mu^\bullet : \Omega \to \mathcal{M}(\mathbb{R}^d)$ is a measurable map $q^\bullet : \Omega \to \mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d)$ s.t. for $\mathbb{P}$-a.e. $\omega \in \Omega$

$$q^\omega \text{ is a (semi-) coupling of } \mathfrak{L} \text{ and } \mu^\omega.$$

We say that a measure $Q \in \mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d \times \Omega)$ is an *universal* (*semi-*) *coupling* of the Lebesgue measure and the point process if and only if $dQ(x, y, \omega)$ is a (semi-) coupling of the Lebesgue measure $d\mathfrak{L}(x)$ and of the Campbell measure $d(\mu^\bullet\mathbb{P})(y, \omega)$.

Disintegration of a universal (semi-) coupling w.r.t. the third marginal yields a measurable map $q^\bullet : \Omega \to \mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d)$ which is a (semi-) coupling of the Lebesgue measure $\mathfrak{L}$ and the point process $\mu^\bullet$. Conversely, given any (semi-) coupling $q^\bullet$ of the Lebesgue measure $\mathfrak{L}$ and the point process $\mu^\bullet$, then its Campbell measure

$$dQ(x, y, \omega) := dq^\omega(x, y)\, d\mathbb{P}(\omega)$$

defines a universal (semi-) coupling.

According to this one-to-one correspondence between $q^\bullet$ [(semi-) coupling of $\mathfrak{L}$ and $\mu^\bullet$] and $Q = q^\bullet\mathbb{P}$ [(semi-) coupling of $\mathfrak{L}$ and $\mu^\bullet\mathbb{P}$], we will freely switch between them. In many cases, the specification "universal" for (semi-) couplings of $\mathfrak{L}$ and $\mu^\bullet\mathbb{P}$ will be suppressed. And quite often, we will simply speak of (*semi-*) *couplings of* $\mathfrak{L}$ *and* $\mu^\bullet$.

2.4. *Fair allocations.* Let $\mu^\bullet \in \mathcal{N}(\mathbb{R}^d)$ be given. A *fair allocation* of Lebesgue measure $\mathcal{L}$ to $\mu^\bullet$ is a measurable map $\Psi^\bullet : \Omega \times \mathbb{R}^d \to \mathbb{R}^d$, $(\omega, x) \mapsto \Psi^\omega(x)$ such that for $\mathbb{P}$-almost every ω:

(i) $\mathcal{L}(\mathbb{R}^d \setminus \bigcup_{\xi \in \Xi_\omega} \Psi_\omega^{-1}(\xi)) = 0$;
(ii) $\mathcal{L}(\Psi_\omega^{-1}(\xi)) = 1$ for all $\xi \in \Xi(\omega)$.

We call each *configuration point* $\xi \in \Xi(\omega)$ a *center*, and the set $(\Psi^\omega)^{-1}(\xi)$ the *cell* associated to the center ξ. The allocation $\Psi^\bullet$ is called equivariant if and only if $\Psi_\omega(x) = y \Rightarrow \forall z \in \mathbb{R}^d : \Psi_{\theta_z\omega}(x + z) = y + z$. An allocation is called *factor allocation* if the random map $\omega \mapsto \Psi^\omega$ is measurable with respect to the σ-algebra generated by $\mu^\bullet$. For some examples on allocations and their connection to Palm measures we refer to [9, 13, 16] and references therein.

In particular, any allocation $\Psi^\bullet$ for $\mu^\bullet$ induces a coupling $q^\bullet$ between $\mathfrak{L}$ and $\mu^\bullet$ via $q^\bullet = (\mathrm{id}, \Psi^\bullet)_*\mathfrak{L}$.

2.5. *The optimal transportation problem.* Given two probability measures λ, μ on $\mathbb{R}^d$ and a measurable cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, the optimal transportation problem between λ and μ is to find a minimizer of

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)\, dq(x, y)$$

among all couplings q of λ and μ. A minimizer is called *optimal coupling*. Optimal couplings have many nice properties. The most basic and also very intuitive one is that they are concentrated on *c-cyclical monotone* sets. A set $N \subset \mathbb{R}^d \times \mathbb{R}^d$ is called c-cyclical monotone if and only if for all $n \in \mathbb{N}$ and $(x_i, y_i) \in N$ for $i = 1, \dots, n$, we have

$$\sum_{i=1}^{n} c(x_i, y_i) \le \sum_{i=1}^{n} c(x_i, y_{i+1}), \tag{3}$$

where $y_{n+1} = y_1$. The interpretation of cyclical monotonicity is clear. If a coupling is optimal we cannot improve it, produce a coupling with less cost, by breaking up and recoupling finitely many coupled pairs of points. In fact, if the cost function is sufficiently nice (continuous is much more than needed, see [6]) also the reverse direction holds. Any measure that is concentrated on a c-cyclical monotone set is optimal. In many situations, the optimal coupling is induced by a transportation map T, that is, $q = (\mathrm{id}, T)_* \lambda$. Then T is c-cyclically monotone if and only if its graph is c-cyclical monotone set. For more details on optimal transportation and its many applications we refer to [25, 31, 32].

2.6. *Cost functionals*. Throughout this paper, ϑ will be a strictly increasing, continuous function from $\mathbb{R}_+$ to $\mathbb{R}_+$ with $\vartheta(0) = 0$ and $\lim_{r \to \infty} \vartheta(r) = \infty$. Given a *scale function* ϑ as above we define the *cost function*

$$c(x, y) = \vartheta\big(|x - y|\big)$$

on $\mathbb{R}^d \times \mathbb{R}^d$, the *cost functional*

$$\mathsf{Cost}(q) = \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)\, dq(x, y)$$

on $\mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d)$ and the *mean cost functional*

$$\mathfrak{Cost}(Q) = \int_{\mathbb{R}^d \times \mathbb{R}^d \times \Omega} c(x, y)\, dQ(x, y, \omega)$$

on $\mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d \times \Omega)$. We have the following basic result on existence and uniqueness of optimal semicouplings, the proof of which is deferred to the Section 6. The first part of the theorem, the existence and uniqueness of an optimal semicoupling, is very much in the spirit of an analogous result by Figalli [11] on existence and (if enough mass is transported) uniqueness of an optimal partial coupling. However, in our case the second marginal is discrete whereas in [11] it is absolutely continuous.

THEOREM 2.1. (i) *For each bounded Borel set $A \subset \mathbb{R}^d$ there exists a unique semicoupling Q_A of $\mathfrak{L}$ and $(1_A \mu^\bullet)\mathbb{P}$ which minimizes the mean cost functional $\mathfrak{Cost}(\cdot)$.*

(ii) Q_A *can be disintegrated as* $dQ_A(x,y,\omega) := dq_A^\omega(x,y)\,d\mathbb{P}(\omega)$ *where for* $\mathbb{P}$-*a.e.* ω *the measure* q_A^ω *is the unique minimizer of the cost functional* $\mathsf{Cost}(\cdot)$ *among the semicouplings of* $\mathfrak{L}$ *and* $1_A\mu^\omega$.

(iii) $\mathfrak{Cost}(Q_A) = \int_\Omega \mathsf{Cost}(q_A^\omega)\,d\mathbb{P}(\omega)$.

For a bounded Borel set $A \subset \mathbb{R}^d$, the *transportation cost on* A is given by the random variable $\mathsf{C}_A : \Omega \to [0,\infty]$ as

$$\mathsf{C}_A(\omega) := \mathsf{Cost}(q_A^\omega) = \inf\{\mathsf{Cost}(q^\omega) : q^\omega \text{ semicoupling of } \mathfrak{L} \text{ and } 1_A\mu^\omega\}.$$

LEMMA 2.2. (1) *If* $A_1,\dots,A_n$ *are disjoint, then* $\forall\omega\in\Omega$

$$\mathsf{C}_{\bigcup_{i=1}^n A_i}(\omega) \ge \sum_{i=1}^n \mathsf{C}_{A_i}(\omega).$$

(2) *If* A_1 *and* A_2 *are translates of each other, then* C_{A_1} *and* C_{A_2} *are identically distributed.*

(3) *If* $A_1,\dots,A_n$ *are disjoint and* $\mu^\bullet(A_1),\dots,\mu^\bullet(A_n)$ *are independent, then the random variables* C_{A_i}, $i=1,\dots,n$, *are independent.*

PROOF. Properties (ii) and (iii) follow directly from the respective properties of the point process and the invariance of the Lebesgue measure under translations. The intuitive argument for (i) is that minimizing the costs on $\bigcup_i A_i$ is more restrictive than doing it separately on each of the A_i. The more detailed argument is the following. Given any semicoupling q^ω of $\mathfrak{L}$ and $1_{\bigcup_i A_i}\mu^\omega$, then for each i the measure $q_i^\omega := 1_{\mathbb{R}^d\times A_i} q^\omega$ is a semicoupling of $\mathfrak{L}$ and $1_{A_i}\mu^\omega$. Choosing q^ω as the minimizer of $\mathsf{C}_{\bigcup_{i=1}^n A_i}(\omega)$ yields

$$\mathsf{C}_{\bigcup_i A_i}(\omega) = \mathsf{Cost}(q^\omega) = \sum_i \mathsf{Cost}(q_i^\omega) \ge \sum_i \mathsf{C}_{A_i}(\omega). \qquad \square$$

2.7. *Convergence along standard exhaustions.* For $n \in \mathbb{N}_0 := \mathbb{N}\cup\{0\}$ and $z \in \mathbb{Z}^d$ define the *cube* or *box* $B_n(z)$ of generation n with basepoint z by

$$B_n(z) = z + [0,2^n)^d.$$

For $z=0$ simply put $B_n = B_n(0)$. More generally, for $\gamma = (\gamma_k) \in \Gamma := (\{0,1\}^d)^{\mathbb{N}}$ put

$$B_n(z,\gamma) = z - \sum_{k=1}^n 2^{k-1}\gamma_k + [0,2^n)^d.$$

Starting with the unit box $B_0(z,\gamma) = z + [0,1)^d$, for any random vector $\gamma \in \Gamma$ the sequence $(B_n(z,\gamma))_{n\in\mathbb{N}_0}$ can be constructed iteratively as follows: Given the box $B_n(z,\gamma)$ attach $2^d - 1$ copies of it—depending on the random variable $\gamma_{n+1} =$

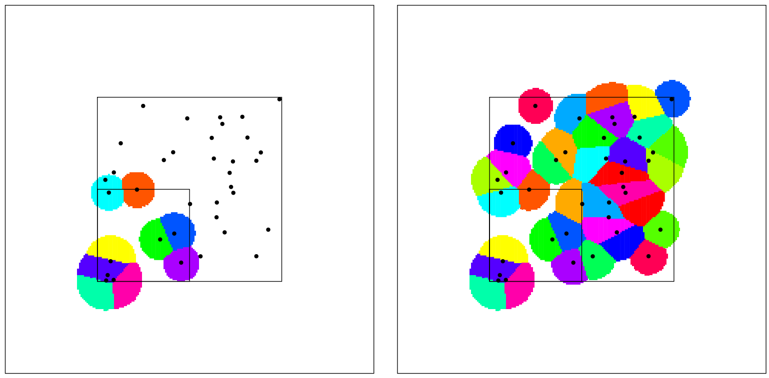

FIG. 2. *Concept of exhausting sequences: start with a small cube and repeatedly double its edge lengths to exhaust space* [*cost function* $c(x, y) = |x - y|^2$].

$(\gamma_{n+1}^1, \dots, \gamma_{n+1}^d)$ with values in $\{0, 1\}^d$—either on the right (if $\gamma_{n+1}^1 = 0$) or on the left (if $\gamma_{n+1}^1 = 1$), either on the backside (if $\gamma_{n+1}^2 = 0$) of on the front (if $\gamma_{n+1}^2 = 1$), either on the top (if $\gamma_{n+1}^3 = 0$) or on the bottom (if $\gamma_{n+1}^3 = 1$), etc; see Figure 2.

The sequence $(B_n(z,\gamma))_{n\in\mathbb{N}_0}$ for fixed z and γ is increasing and for ν-almost every $\gamma \in \Gamma$ it increases to $\mathbb{R}^d$. Each of the boxes $B_n(z,\gamma)$ contains the point z.

Put

$$\mathfrak{c}_n := 2^{-dn} \cdot \mathbb{E}[\mathsf{C}_{B_n(z,\gamma)}].$$

Note that translation invariance (equivariance plus stationarity) implies that the right-hand side does not depend on $z \in \mathbb{Z}^d$ and $\gamma \in \Gamma$.

COROLLARY 2.3. (i) *The sequence* $(\mathfrak{c}_n)_{n\in\mathbb{N}_0}$ *is nondecreasing. The limit*

$$\mathfrak{c}_\infty = \lim_{n\to\infty} \mathfrak{c}_n = \sup_n \mathfrak{c}_n$$

exists in $(0, \infty]$.

(ii) *Assume that* $\mu^\bullet$ *is ergodic. Then, we have for all* $z \in \mathbb{Z}^d$, *for all* $\gamma \in \Gamma$ *and for* $\mathbb{P}$-*almost every* $\omega \in \Omega$,

$$\liminf_{n\to\infty} 2^{-nd} \mathsf{C}_{B_n(z,\gamma)}(\omega) = \mathfrak{c}_\infty.$$

(iii) $\mathfrak{c}_\infty \le \inf_{q\in\Pi_s} \mathfrak{C}_\infty(q)$ *where* Π_s *denotes the set of semicouplings of* $\mathfrak{L}$ *and* $\mu^\bullet$.

PROOF. (i) is an immediate consequence of the previous lemma. For (ii) fix an arbitrary nested sequence of boxes $(B_n)_n$ generated by a standard exhaustion. Then we have by superadditivity $\forall \omega \in \Omega$ for all $n, k \in \mathbb{N}$

$$2^{-d(n+k)} \mathsf{C}_{B_{n+k}}(\omega) \ge 2^{-dk} \sum_{j=1}^{2^{dk}} 2^{-nd} \mathsf{C}_{B_n^j}(\omega),$$

where B_n^j are disjoint copies of B_n such that $\bigcup_{j=1}^{2^{dk}} B_n^j = B_{n+k}$. In the limit of $k \to \infty$ we get by ergodicity for $\mathbb{P}$-a.e. ω

$$\liminf_{k\to\infty} 2^{-kd} \mathsf{C}_{B_k}(\omega) \geq \mathbb{E}[2^{-nd}\mathsf{C}_{B_n}] = \mathfrak{c}_n$$

for each $n \in \mathbb{N}$ and thus

$$\liminf_{k\to\infty} 2^{-kd} \mathsf{C}_{B_k}(\omega) \geq \mathfrak{c}_\infty.$$

On the other hand, Fatou's lemma implies

$$\mathbb{E}\Big[\liminf_{n\to\infty} 2^{-nd}\mathsf{C}_{B_n}\Big] \leq \liminf_{n\to\infty} \mathbb{E}[2^{-nd}\mathsf{C}_{B_n}] = \mathfrak{c}_\infty.$$

Both inequalities together imply the assertion.

For (iii) take any semicoupling $q^\bullet$ of $\mathfrak{L}$ and $\mu^\bullet\mathbb{P}$. Then we have for any n

$$2^{-dn}\mathfrak{Cost}(1_{R^d\times B_n\times\Omega}q^\bullet) \geq \mathfrak{c}_n.$$

Taking the limit yields

$$\mathfrak{C}_\infty(q^\bullet) = \liminf_{n\to\infty} 2^{-dn}\mathfrak{Cost}(1_{R^d\times B_n\times\Omega}q^\bullet) \geq \lim_n \mathfrak{c}_n = \mathfrak{c}_\infty. \qquad \square$$

COROLLARY 2.4. *$\mathfrak{c}_\infty$ only depends on the scale ϑ and on the* distribution *of $\mu^\bullet$, not on the choice of the realization of μ^ω on a particular probability space* $(\Omega, \mathfrak{A}, \mathbb{P})$.

PROOF. It is sufficient to show that $\mathfrak{c}_n$ just depends on the distribution of $\mu^\bullet$. For a given set of points $\Xi(\omega)$ in B_n there is a unique semicoupling $q_{B_n}^\omega$ of $\mathfrak{L}$ and $1_{B_n}\mu^\omega$ minimizing Cost; see Proposition 6.3. Hence, $q_{B_n}^\omega$ just depends on $\Xi(\omega)$. However, the distribution of the points in B_n, $\Xi(\omega)$, just depends on the distribution of $\mu^\bullet$. $\square$

REMARK 2.5. None of the previous definitions and results required that $\mu^\bullet$ have subunit intensity. However, one easily verifies that

$$\beta > 1 \quad \Longrightarrow \quad \mathfrak{c}_\infty = \infty,$$

where $\beta := \mathbb{E}[\mu^\bullet([0, 1)^d)]$ denotes the intensity of the equivariant point process.

REMARK 2.6. The problem of finding an optimal semicoupling between $\mathfrak{L}$ and a Poisson point process $\mu^\bullet$ of intensity $\beta < 1$ is equivalent to the problem of finding an optimal semicoupling between $\mathfrak{L}$ and $\beta \cdot \hat{\mu}^\bullet$ where $\hat{\mu}^\bullet$ is a Poisson point process of unit intensity; see Figure 3.

Indeed, given $\beta \in (0, 1)$ and a semicoupling $q^\bullet$ of $\mathfrak{L}$ and a Poisson point process $\mu^\bullet$ of intensity β. Put $\tau : x \mapsto \beta^{1/d}x$ on $\mathbb{R}^d$ as well as on $\mathbb{R}^d \times \mathbb{R}^d$. Then $\hat{\mu}^\omega := \tau_*\mu^\omega$ is a Poisson point process with intensity 1, and

$$\tilde{q}^\omega := \beta \cdot \tau_* q^\omega$$

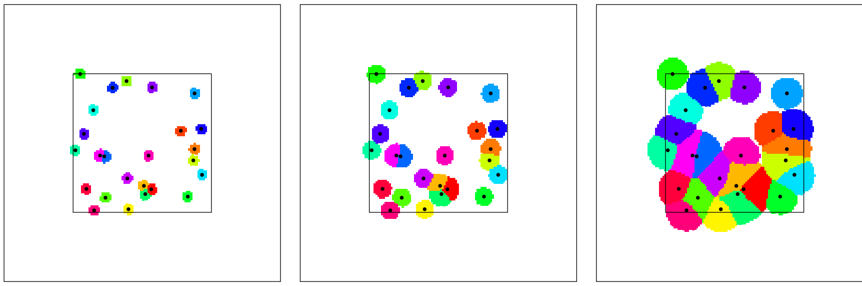

FIG. 3. *Semicoupling of Lebesgue and* 25 *points in the cube with* $c(x, y) = |x - y|$ *where each point gets mass* $1/9, 1/3, 1$, *respectively.*

is a semicoupling of $\mathfrak{L}$ and $\beta \cdot \hat{\mu}^\omega$. Conversely, given any Poisson point process $\hat{\mu}^\omega$ of unit intensity and any semicoupling $\tilde{q}^\omega$ of $\mathfrak{L}$ and $\beta \cdot \hat{\mu}^\omega$, then $q^\omega := \frac{1}{\beta} \cdot (\tau^{-1})_* \tilde{q}^\omega$ is a semicoupling of $\mathfrak{L}$ and $\mu^\omega := (\tau^{-1})_* \hat{\mu}^\omega$, the latter being a Poisson point process of intensity β. In both cases, q is equivariant if and only if $\tilde{q}$ is equivariant.

The asymptotic mean transportation cost for $\tilde{q}^\bullet$ measured with scale ϑ will coincide with the asymptotic mean transportation cost for $q^\bullet$ measured with scale $\vartheta_\beta(r) := \beta \cdot \vartheta(\beta^{-1/d} r)$,

$$\mathbb{E} \int_{\mathbb{R}^d \times [0,1)^d} \vartheta(|x - y|) \, d\tilde{q}^\bullet = \mathbb{E} \int_{\mathbb{R}^d \times [0,1)^d} \vartheta_\beta(|x - y|) \, dq^\bullet.$$

3. Uniqueness. Throughout this section we fix an equivariant point process $\mu^\bullet : \Omega \to \mathcal{M}(\mathbb{R}^d)$ of subunit intensity and with finite asymptotic mean transportation cost $\mathfrak{c}_\infty$.

PROPOSITION 3.1. *Given a counting measure* $\mu \in \mathcal{N}(\mathbb{R}^d)$ *and a semicoupling* q *of* $\mathfrak{L}$ *and* μ, *then the following properties are equivalent*:

(i) *For each bounded Borel set* $A \subset \mathbb{R}^d$, *the measure* $1_{\mathbb{R}^d \times A} q$ *is the unique optimal semicoupling of the measures* $\lambda_A(\cdot) := q(\cdot, A)$ *and* $1_A \mu$; *see Figure* 4.

(ii) *The support of* q *is* c*-cyclically monotone*, *more precisely*,

$$\sum_{i=1}^n c(x_i, y_i) \le \sum_{i=1}^n c(x_i, y_{i+1})$$

for any $n \in \mathbb{N}$ *and any choice of points* $(x_1, y_1), \dots, (x_n, y_n)$ *in* $\mathrm{supp}(q)$ *with the convention* $y_{n+1} = y_1$; *cf.* (3).

(iii) *There exists a density* $\rho : \mathbb{R}^d \to [0, 1]$ *and a* c*-cyclically monotone map* $T^\omega : \{\rho > 0\} \to \mathbb{R}^d$ *such that*

$$q = (\mathrm{id}, T)_*(\rho \mathfrak{L}). \tag{4}$$

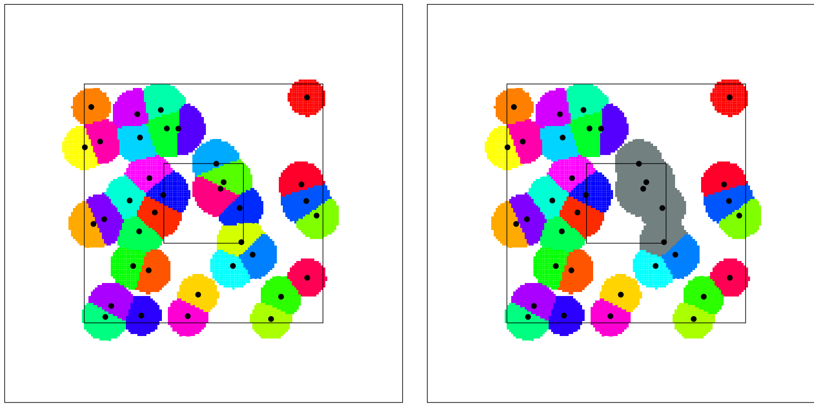

FIG. 4. *The left picture is a semicoupling of Lebesgue and* 36 *points with cost function* $c(x,y)=|x-y|^4$. *In the right picture, the five points within the small cube can choose new partners from the mass that was transported to them in the left picture* (*corresponding to the measure* λ_A). *If the semicoupling on the left-hand side is locally optimal, then the points in the small cube on the right-hand side will choose from the gray region exactly the partners they have in the left picture.*

Recall that, by definition, a map T *is* c*-cyclically monotone if and only if the closure of its graph* $\{(x,T(x)):x\in A^\omega\}$ *is a* c*-cyclically monotone set.*

PROOF. The implications (iii) $\Longrightarrow$ (ii) $\Longrightarrow$ (i) follow from Lemma 6.1.

(i) $\Longrightarrow$ (iii): Fix an exhaustion $(B'_n)_n$ of $\mathbb{R}^d$ by boxes, say $B'_n=[-2^{n-1},2^{n-1})^d$. For each $n\in\mathbb{N}$, let ρ_n be the density of the measure $\lambda_n:=\lambda_{B'_n}$ on $\mathbb{R}^d$. This is the part of Lebesgue measure from which the points inside of B'_n might choose their "partners." Obviously, $0\le\rho_n\le\rho_{n+1}\le 1$. Hence, $\lim_{n\to\infty}\rho_n(x)=\rho(x)\le 1$ exists $\mathfrak{L}$-a.e.

Assuming (i), according to Proposition 6.3 (or, more precisely, a canonical extension of it for semicouplings of $\rho\mathfrak{L}$ and σ), there exists a c-cyclically monotone map $T_n:\{\rho_n>0\}\to\mathbb{R}^d$ such that

$$dq(x,y)=d\delta_{T_n(x)}(y)\rho_n(x)\,d\mathfrak{L}(x)\qquad\text{on } \mathbb{R}^d\times B'_n.$$

Since the left-hand side is independent of n, we have

$$T_{n+1}=T_n\qquad\text{on } \{\rho_n>0\}.$$

This trivially yields the existence of

$$T:=\lim_{n\to\infty}T_n\qquad\text{on } \{\rho>0\}:=\lim_{n\to\infty}\{\rho_n\},$$

defining a c-cyclically monotone map $T:\{\rho>0\}\to\mathbb{R}^d$ with the property that

$$dq(x,y)=d\delta_{T(x)}(y)\rho(x)\,d\mathfrak{L}(x).\qquad\square$$

REMARK 3.2. Set $A = \{\rho > 0\}$. In the sequel, any *transport map* $T : A \to \mathbb{R}^d$ as above will be extended to a map $T : \mathbb{R}^d \to \mathbb{R}^d \cup \{\eth\}$ by putting $T(x) := \eth$ for all $x \in \mathbb{R}^d \setminus A$ where $\eth$ denotes an isolated point added to $\mathbb{R}^d$ ("point at infinity," "cemetery"). Then (4) simplifies to

$$q = (\mathrm{id}, T)_*(\rho\mathfrak{L}) \qquad \text{on } \mathbb{R}^d \times \mathbb{R}^d. \tag{5}$$

Moreover, we put $c(x, T(x)) = c(x, \eth) := 0$ for $x \in \mathbb{R}^d \setminus A$.

DEFINITION 3.3.

- *A semicoupling $Q = q^\bullet\mathbb{P}$ of $\mathfrak{L}$ and $\mu^\bullet$ is called* locally optimal *if and only if some* (*hence every*) *of the properties of the previous proposition are satisfied for $\mathbb{P}$-a.e. $\omega \in \Omega$.*
- *A semicoupling $Q = q^\bullet\mathbb{P}$ of $\mathfrak{L}$ and $\mu^\bullet$ is called* asymptotically optimal *if and only if*
$$\liminf_{n\to\infty} 2^{-nd}\mathfrak{Cost}(1_{\mathbb{R}^d\times B'_n} Q) = \mathfrak{c}_\infty$$
for some exhaustion $(B'_n)_n$ of $\mathbb{R}^d$ by boxes $B'_n = B_n(z, \gamma)$.
- *A semicoupling $Q = q^\bullet\mathbb{P}$ of $\mathfrak{L}$ and $\mu^\bullet$ is called* equivariant *if and only if for each $z \in \mathbb{Z}^d$ the measure Q is equivariant under the diagonal action of $\mathbb{Z}^d$, that is*,
$$q^\omega(A, B) = q^{\omega+z}(A + z, B + z)$$
for all $z \in \mathbb{Z}^d$ and $A, B \in \mathcal{B}(\mathbb{R}^d)$.
- *A semicoupling $Q = q^\bullet\mathbb{P}$ of $\mathfrak{L}$ and $\mu^\bullet$ is called* optimal *if and only if it is equivariant and asymptotically optimal.*

The very same definitions apply to *couplings* instead of semicouplings.

REMARK 3.4. (i) Asymptotic optimality is not sufficient for uniqueness and it does not imply local optimality: Given any asymptotically optimal semicoupling $q^\bullet$ and a bounded Borel set $A \subset \mathbb{R}^d$ of positive volume, choose an arbitrary coupling $\tilde{q}^\omega_A$ of the measures $q^\omega(\cdot, A)$ and $1_A\mu^\omega$, which are the marginals of $q^\omega_A := 1_{\mathbb{R}^d\times A} q^\omega$. If $\mu^\omega(A) \geq 2$ (which happens with positive probability), then one can always achieve that $\tilde{q}^\omega_A$ is a nonoptimal coupling and that it is different from q^ω_A. Put

$$\tilde{q}^\omega := q^\omega + \tilde{q}^\omega_A - q^\omega_A.$$

Then $\tilde{q}^\bullet$ is an asymptotically optimal semicoupling of $\mathfrak{L}$ and $\mu^\bullet$. It is not locally optimal and it does not coincide with $q^\bullet$.

(ii) Local optimality does not imply asymptotic optimality and it is not sufficient for uniqueness: For instance in the case $p = 2$, given any coupling $q^\bullet$ of $\mathfrak{L}$ and $\mu^\bullet$ and $z \in \mathbb{R}^d \setminus \{0\}$, then

$$d\tilde{q}^\omega(x, y) := dq^\omega(x + z, y)$$

defines another locally optimal coupling of $\mathfrak{L}$ and $\mu^\bullet$. At most one of them can be asymptotically optimal.

(iii) Note that local optimality—in contrast to asymptotic optimality and equivariance—is not preserved under convex combinations. We do not claim that local optimality and asymptotic optimality imply uniqueness.

(iv) Local optimality links classical optimal transportation problems, problems between finite measures, with optimal transportation problems between $\mathfrak{L}$ and a point process by locally optimizing the semicouplings.

Given $\gamma, \eta \in \mathcal{M}(\mathbb{R}^d)$ with $\gamma(\mathbb{R}^d) \geq \eta(\mathbb{R}^d)$, we define the *transportation cost* by

$$\mathsf{Cost}(\gamma, \eta) := \inf\{\mathsf{Cost}(q) : q \in \Pi_s(\gamma, \eta)\}.$$

Similarly, given measure valued random variables $\gamma^\bullet, \eta^\bullet : \Omega \to \mathcal{M}(\mathbb{R}^d)$ and a bounded Borel set $A \subset \mathbb{R}^d$ we define the *mean transportation cost* by

$$\mathfrak{Cost}(\gamma^\bullet, \eta^\bullet) := \inf\{\mathfrak{Cost}(q^\bullet\mathbb{P}) : q^\omega \in \Pi_s(\gamma^\omega, \eta^\omega) \text{ for a.e. } \omega\}.$$

Given a (semi-) coupling $Q = q^\bullet\mathbb{P}$ of $\mathfrak{L}$ and $\mu^\bullet\mathbb{P}$, recall the definition of $\lambda_A^\bullet$ from Proposition 3.1. We define the *efficiency of the* (*semi*-) *coupling* Q *on the set* A by

$$\mathfrak{eff}_A(Q) := \frac{\mathfrak{Cost}(\lambda_A^\bullet, 1_A\mu^\bullet)}{\mathfrak{Cost}(1_{\mathbb{R}^d \times A} Q)}.$$

It is a number in (0, 1]. The (semi-) coupling Q is said to be efficient on A if and only if $\mathfrak{eff}_A(Q) = 1$. Otherwise, it is inefficient on A.

LEMMA 3.5. (i) *Q is locally optimal if and only if $\mathfrak{eff}_A(Q) = 1$ for all bounded Borel sets $A \subset \mathbb{R}^d$.*

(ii) *$\mathfrak{eff}_A(Q) = 1$ for some $A \subset \mathbb{R}^d$ implies $\mathfrak{eff}_{A'}(Q) = 1$ for all $A' \subset A$.*

PROOF. (i) Let A be given and $\omega \in \Omega$ be fixed. Then $1_{\mathbb{R}^d \times A} q^\omega$ is the optimal semicoupling of the measures λ_A^ω and $1_A\mu^\omega$ if and only if

$$\mathsf{Cost}(1_{\mathbb{R}^d \times A} q^\omega) = \mathsf{Cost}(\lambda_A^\omega, 1_A\mu^\omega). \tag{6}$$

On the other hand, $\mathfrak{eff}_A(Q) = 1$ is equivalent to

$$\mathbb{E}[\mathsf{Cost}(1_{\mathbb{R}^d \times A} q^\bullet)] = \mathbb{E}[\mathsf{Cost}(\lambda_A^\bullet, 1_A\mu^\omega)].$$

The latter, in turn, is equivalent to (6) for $\mathbb{P}$-a.e. $\omega \in \Omega$.

(ii) If the transport q restricted to $\mathbb{R}^d \times A$ is optimal, then also each of its subtransports; see Theorem 4.6 in [32]. $\square$

THEOREM 3.6. *Every optimal semicoupling of $\mathfrak{L}$ and $\mu^\bullet\mathbb{P}$ is locally optimal.*

PROOF. Assume we are given a semicoupling Q of $\mathfrak{L}$ and $\mu^\bullet\mathbb{P}$ which is equivariant and not locally optimal. According to the previous lemma, the latter implies that there exist $n \in \mathbb{N}$ and $z_0 \in \mathbb{Z}^d$ such that the semicoupling Q is not efficient on the box $B_n(z_0) = z_0 + [0, 2^n)^d$, that is,

$$\eta := \mathfrak{eff}_{B_n(z_0)}(Q) < 1.$$

By equivariance this implies $\mathfrak{eff}_{B_n(z)}(Q) = \eta < 1$ for all $z \in \mathbb{Z}^d$. Hence, for each $z \in \mathbb{Z}^d$ there exists a measure-valued random variable $\tilde{q}^\bullet_{B_n(z)}$ such that $\tilde{q}^\omega_{B_n(z)}$ for a.e. ω is a semicoupling of $\lambda^\omega_{B_n(z)}$ and $1_{B_n(z)}\mu^\omega$ and more efficient than $q^\omega_{B_n(z)} := 1_{\mathbb{R}^d \times B_n(z)} \cdot q^\omega$, that is, such that

$$\mathbb{E}\big[\mathsf{Cost}\big(\tilde{q}^\bullet_{B_n(z)}\big)\big] \le \eta \cdot \mathbb{E}\big[\mathsf{Cost}\big(q^\bullet_{B_n(z)}\big)\big].$$

Put

$$\tilde{q}^\bullet := \sum_{z \in (2^n\mathbb{Z})^d} \tilde{q}^\bullet_{B_n(z)}.$$

Then $\tilde{q}^\bullet$ is a semicoupling of $\mathfrak{L}$ and $\mu^\bullet$ and for all $z \in (2^n\mathbb{Z})^d$

$$\mathbb{E}\big[\mathsf{Cost}\big(1_{\mathbb{R}^d \times B_n(z)}\tilde{q}^\bullet\big)\big] \le \eta \cdot \mathbb{E}\big[\mathsf{Cost}\big(1_{\mathbb{R}^d \times B_n(z)}q^\bullet\big)\big].$$

Equivariance of $q^\bullet$—together with uniqueness of cost minimizers on bounded sets—implies equivariance of $\tilde{q}^\bullet$ under the group $(2^n\mathbb{Z}^d)$. In other words, $\tilde{Q} = \tilde{q}^\bullet\mathbb{P}$ is an $(2^n\mathbb{Z}^d)$-equivariant semicoupling of $\mathfrak{L}$ and $\mu^\bullet\mathbb{P}$ which satisfies

$$\mathfrak{Cost}(1_{\mathbb{R}^d \times B_n(z)}\tilde{Q}) \le \eta \cdot \mathfrak{Cost}(1_{\mathbb{R}^d \times B_n(z)}Q)$$

for all $z \in (2^n\mathbb{Z})^d$. Additivity of the mean cost functional $\mathfrak{Cost}(\cdot)$ implies

$$\mathfrak{Cost}(1_{\mathbb{R}^d \times B_{n+k}}\tilde{Q}) \le \eta \cdot \mathfrak{Cost}(1_{\mathbb{R}^d \times B_{n+k}}Q)$$

for all $k \in \mathbb{N}_0$ and therefore, due to Corollary 2.3(iii), finally

$$\mathfrak{c}_\infty \le \liminf_{k\to\infty} \mathfrak{Cost}(1_{\mathbb{R}^d \times B_k}\tilde{Q}) \le \eta \cdot \liminf_{k\to\infty} \mathfrak{Cost}(1_{\mathbb{R}^d \times B_k}Q)$$

with $\eta < 1$. This proves that Q is not asymptotically optimal. $\square$

LEMMA 3.7. *Let $q^\omega = (\mathrm{id}, T^\omega)_*(\rho^\omega\mathfrak{L})$ be an optimal semicoupling between $\mathfrak{L}$ and $\mu^\bullet$. Then, $\mathbb{P}$-a.s. we have $\rho^\omega(x) \in \{0, 1\}$ $\mathfrak{L}$-a.e.*

PROOF. Assume there is a $n \in \mathbb{N}$ and $B_n(z_0) = z_0 + [0, 2^n)^d$ such that on a set of positive $\mathbb{P}$-measure

$$q_n^\omega := 1_{\mathbb{R}^d \times B_n(z_0)}\, dq^\omega(x, y) = (\mathrm{id}, T^\omega)_*(\rho_n^\omega \mathfrak{L})$$

with $0 < \rho_n^\omega < 1$ on a set of positive $\mathfrak{L}$-measure. However, due to Proposition 6.3 this implies that $Q = q^\bullet \mathbb{P}$ is not efficient on $B_n(z_0)$ because it is possible to construct a semicoupling between $1_{\rho_n^\omega > 0}\mathfrak{L}$ and $1_{B_n(z_0)}\mu^\omega$ with less cost. By the same reasoning as in the last proof, this implies that Q is not optimal. □

Hence, any optimal semicoupling can be written as $q^\omega = (\mathrm{id}, T^\omega)_* \mathfrak{L}$ for some measurable map $T : A^\omega \to \mathbb{R}^d \cup \{\eth\}$; cf Remark 3.2.

THEOREM 3.8. *There exists at most one optimal semicoupling of $\mathfrak{L}$ and $\mu^\bullet \mathbb{P}$.*

PROOF. Assume we are given two optimal semicouplings $q_1^\bullet$ and $q_2^\bullet$. Then also $q^\bullet := \frac{1}{2} q_1^\bullet + \frac{1}{2} q_2^\bullet$ is an optimal semicoupling. Hence, by the previous theorem all three couplings—$q_1^\bullet$, $q_2^\bullet$ and $q^\bullet$—are locally optimal. Thus, for a.e. ω by the results of Proposition 3.1 and the last lemma there exist maps T_1^ω, T_2^ω, T^ω and sets A_1^ω, A_2^ω, A^ω such that

$$\begin{aligned} dq^\omega(x, y) &= d\delta_{T^\omega(x)}(y) 1_{A^\omega}(x)\, d\mathfrak{L}(x) \\ &= \bigl(\tfrac{1}{2} d\delta_{T_1^\omega(x)}(y) 1_{A_1^\omega}(x) + \tfrac{1}{2} d\delta_{T_2^\omega(x)}(y) 1_{A_2^\omega}(x)\bigr)\, d\mathfrak{L}(x). \end{aligned}$$

This, however, implies $T_1^\omega(x) = T_2^\omega(x)$ for a.e. $x \in A_1^\omega \cap A_2^\omega$ and, moreover, $A_1^\omega = A_2^\omega$. Thus $q_1^\omega = q_2^\omega$. □

REMARK 3.9. Note that we only used equivariance under the action of $\mathbb{Z}^d$. However, the minimizer is equivariant under the action of $\mathbb{R}^d$. For the uniqueness it would also have been sufficient to require equivariance under the action of $k\mathbb{Z}^d$ for some $k \in \mathbb{N}$.

THEOREM 3.10. (i) *If $\mu^\bullet$ has unit intensity, then every optimal semicoupling of $\mathfrak{L}$ and $\mu^\bullet$ is indeed a coupling of them.*

(ii) *Conversely, if an optimal coupling exists, then $\mu^\bullet$ must have unit intensity.*

This theorem is in a similar spirit as Theorem 4 in [13].

PROOF. (i) Let Q be an optimal semicoupling. For $n \in \mathbb{N}$ put $B_n(z) = z + [0, 2^n)^d$ and consider the *saturation* $\alpha_k := 2^{-kd} Q(B_k(z) \times B_k(z) \times \Omega) \le 1$. Note that α_k is independent of $z \in \mathbb{Z}^d$. Hence, we have $\alpha_k \le \alpha_{k+1}$. Indeed, $B_{k+1}(z)$ is the disjoint union of 2^d cubes $B_k(y_j)$ for suitable y_j. Therefore,

$$\alpha_{k+1} \ge 2^{-d} \sum_{j=1}^{2^d} 2^{-kd} Q\bigl(B_k(y_j) \times B_k(y_j) \times \Omega\bigr) = \alpha_k.$$

Thus, the limit $\alpha_\infty := \lim_{k\to\infty} \alpha_k$ exists, and we have $\alpha_\infty \in (0, 1]$.

Since $\mu^\bullet$ has unit intensity and since Q is a semicoupling, we have $Q(\mathbb{R}^d \times B_k \times \Omega) = 2^{kd}$. Let us first assume that $\alpha_\infty < 1$ and choose $r = [(1 + \frac{1}{2}(1 - \alpha_\infty))^{1/d} - 1]/2$. Then for all $k \in \mathbb{N}$ mass of a total amount of at least $(1 - \alpha_\infty)2^{kd}$ has to be transported from $\complement B_k$ into B_k. The volume of the $(r2^k)$-neighborhood of the box B_k is less than $\frac{1}{2}(1 - \alpha_\infty)2^{kd}$. Hence, mass of total amount of at least $\frac{1}{2}(1 - \alpha_\infty)2^{kd}$ has to be transported at least the distance $r2^k$. Thus, we can estimate the costs per unit from below by

$$2^{-kd} \int_{\mathbb{R}^d \times B_k \times \Omega} c(x, y)\, dQ(x, y, \omega) \geq \frac{1}{2}(1 - \alpha_\infty)\vartheta(r2^k).$$

The right-hand side diverges as k tends to infinity which contradicts the finiteness of the costs per unit. Thus, we have $\alpha_\infty = 1$. Furthermore, for all k there is a $u \in B_k(0)$ such that

$$\begin{aligned}
\alpha_k &= 2^{-kd} Q\big(B_k(0) \times B_k(0) \times \Omega\big) \\
&= 2^{-kd} \sum_{v \in B_k(0) \cap \mathbb{Z}^d} Q\big(B_0(v) \times B_k(0) \times \Omega\big) \\
&\leq Q\big(B_0(u) \times B_k(0) \times \Omega\big) \leq Q\big(B_0(u) \times \mathbb{R}^d \times \Omega\big).
\end{aligned}$$

However, by translation invariance (equivariance plus stationarity) the quantity $Q(B_0(u) \times \mathbb{R}^d \times \Omega)$ is independent of u. Moreover, it is bounded above by 1 as Q is a semicoupling. Hence, we have for all $v \in \mathbb{R}^d$:

$$1 = \limsup_{k\to\infty} \alpha_k \leq Q\big(B_0(v) \times \mathbb{R}^d \times \Omega\big) \leq 1.$$

Therefore, Q is actually a coupling of the Lebesgue measure and the point process.

(ii) Assume that Q is an optimal coupling and that $\beta < 1$. Then a similar argumentation as above yields that for each box B_k, Lebesgue measure of total mass $\geq (1 - \beta) \cdot 2^{kd}$ has to be transported from the interior of B_k to the exterior. As k tends to ∞, the costs of these transports explode. $\square$

COROLLARY 3.11. *In the case $\vartheta(r) = r^2$, given an optimal coupling $q^\bullet$ of $\mathfrak{L}$ and a point process $\mu^\bullet$ of unit intensity then for a.e. $\omega \in \Omega$ there exists a convex function $\varphi^\omega : \mathbb{R}^d \to \mathbb{R}$ (unique up to additive constants) such that*

$$q^\omega = \big(\mathrm{id}, \nabla\varphi^\omega\big)_* \mathfrak{L}.$$

In particular, a "fair allocation rule" is given by the monotone map $T^\omega = \nabla\varphi^\omega$.

Moreover, for a.e. ω and any center $\xi \in \Xi(\omega) := \mathrm{supp}(\mu^\omega)$, the associated cell

$$S^\omega(\xi) = \big(T^\omega\big)^{-1}\big(\{\xi\}\big)$$

is a convex polyhedron of volume $\mu^\omega(\xi) \in \mathbb{N}$. If the point process is simple, then all these cells have volume 1.

PROOF. By Proposition 3.1 we know that $T^\omega = \lim_{n\to\infty} T_n^\omega$, where T_n^ω is an optimal transportation map from some set A_n^ω to B_n'. From the classical theory (see [7, 12]), we know that $T_n^\omega = \nabla\varphi_n^\omega$ for some convex function φ_n^ω. More precisely,

$$\varphi_n^\omega(x) = \max_{\xi\in\Xi(\omega)\cap B_n'} \big(x^2 - |x-\xi|^2/2 + b_\xi\big)$$

for some constants b_ξ. Moreover, we know that $T_{n+k}^\omega = T_n^\omega$ on A_n^ω for any $k \in \mathbb{N}$. Fix any $\xi_0 \in \Xi(\omega)$. Then there is $n \in \mathbb{N}$ such that $\xi_0 \in B_n'$. Then $(T_{n+k}^\omega)^{-1}(\xi_0) = (T_n^\omega)^{-1}(\xi_0)$ for any $k \in \mathbb{N}$. Furthermore,

$$T_n^\omega(x) = \xi_0 \quad \Leftrightarrow \quad -|x-\xi_0|^2/2 + b_{\xi_0} > -|x-\xi|^2/2 + b_\xi \qquad \forall \xi \in \Xi(\omega)\cap B_n', \xi \neq \xi_0.$$

For fixed $\xi \neq \xi_0$ this equation describes two half-spaces separated by a hyperplane (defined by equality in the equation above). The set $S^\omega(\xi_0)$ is then given as the intersection of all these halfspaces defined by ξ_0 and $\xi \in \Xi(\omega)\cap B_n'$. Hence, it is a convex polytope. Moreover, the last inequality is exactly the defining equation for a Laguerre tessellation wrt $\mathrm{supp}(\mu^\omega)$ and weights b_ξ; see [18]. □

4. Construction of optimal semicouplings. Again we fix an equivariant point process $\mu^\bullet : \Omega \to \mathcal{M}(\mathbb{R}^d)$ of subunit intensity and with finite asymptotic mean transportation cost $\mathfrak{c}_\infty$.

4.1. *Second randomization and annealed limits.* The crucial step in our construction of an optimal coupling of Lebesgue measure and the point process will be the introduction of a *second randomization*, in addition to the first randomness modeled on the probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ which describes the random choice $\omega \mapsto \mu^\omega$ of a realization of the point process. The second randomization describes the random choice $\gamma \mapsto (B_n(z,\gamma))_{n\in\mathbb{N}}$ of an increasing sequence of boxes containing a given starting point $z \in \mathbb{Z}^d$; see also Section 2.7. It is modeled on the *Bernoulli scheme* $(\Gamma, \mathfrak{B}(\Gamma), \nu)$ with $\Gamma = (\{0,1\}^d)^{\mathbb{N}}$, $\mathfrak{B}(\Gamma)$ its Borel σ-field and ν the uniform distribution on $\Gamma = (\{0,1\}^d)^{\mathbb{N}}$ (or, more precisely, the infinite product of the uniform distribution on $\{0,1\}^d$).

For each $z \in \mathbb{Z}^d$, $\gamma \in \Gamma$ and $k \in \mathbb{N}$, recall that $Q_{B_k(z,\gamma)}$ denotes the minimizer of $\mathfrak{Cost}$ among the semicouplings of $\mathfrak{L}$ and $(1_{B_k(z,\gamma)}\mu^\bullet)\mathbb{P}$ as constructed in Theorem 2.1. Equivariance of this minimizer implies that

$$Q_{B_k(z',\gamma)}(A, B, \omega) = Q_{B_k(z,\gamma)}\big(A + z - z', B + z - z', \omega + z - z'\big)$$

for all $z, z' \in \mathbb{Z}^d$ and $A, B \in \mathcal{B}(M)$. Put

$$dQ_z^k(x,y,\omega) := \int_\Gamma dQ_{B_k(z,\gamma)}(x,y,\omega)\, d\nu(\gamma)$$

and $d\dot{Q}_z^k(x,y,\omega) := 1_{B_0(z)}(y)\,dQ_z^k(x,y,\omega)$.

The measure $\dot{Q}_z^k$ defines a semicoupling between the Lebesgue measure and the point process restricted to the box $B_0(z)$. It is a deterministic, fractional allocation in the following sense:

- it is a deterministic function of μ^ω and does not depend on any additional randomness [coming, e.g., from $d\nu(\gamma)$];
- the measure transported into a given point of the point process has density ≤ 1.

The last fact of course implies that the semicoupling $\dot{Q}_z^k$ is *not* optimal. The first fact implies that all the objects derived from $\dot{Q}_z^k$ in the sequel—like $\dot{Q}_z^\infty$ and Q^∞—are also deterministic.

LEMMA 4.1. (i) *For each $k \in \mathbb{N}$ and $z \in \mathbb{Z}^d$*

$$\int_{\mathbb{R}^d \times B_0(z)\times\Omega} c(x,y)\,dQ_z^k(x,y,\omega) \leq \mathfrak{c}_\infty.$$

(ii) *The family $(\dot{Q}_z^k)_{k\in\mathbb{N}}$ of probability measures on $\mathbb{R}^d \times \mathbb{R}^d \times \Omega$ is relatively compact in the weak topology.*

(iii) *There exist probability measures $\dot{Q}_z^\infty$ and a subsequence $(k_l)_{l\in\mathbb{N}}$ such that for all $z \in \mathbb{Z}^d$*

$$\dot{Q}_z^{k_l} \longrightarrow \dot{Q}_z^\infty \qquad \textit{weakly as } l \to \infty.$$

PROOF. (i) Let us fix $z \in \mathbb{Z}^d$ and start with the following important observation: *For given $n \in \mathbb{N}$ the initial box $B_0(z)$ has each possible "relative position within $B_n(z,\gamma)$" with equal probability.*

Hence, together with translation invariance of $Q_{B_k(z,\gamma)}$ (which in turn follows from equivariance and stationarity of $\mathbb{P}$) we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^d \times B_0(z)\times\Omega} c(x,y)\,dQ_z^k(x,y,\omega)\\
&\quad= \int_\Gamma \int_{\mathbb{R}^d \times B_0(z)\times\Omega} c(x,y)\,dQ_{B_k(z,\gamma)}(x,y,\omega)\,d\nu(\gamma)\\
&\quad= 2^{-kd} \sum_{v\in B_k(z)\cap\mathbb{Z}^d} \left[\int_{\mathbb{R}^d \times B_0(v)\times\Omega} c(x,y)\,dQ_{B_k(z)}(x,y,\omega)\right]\\
&\quad= 2^{-kd} \int_{\mathbb{R}^d \times B_k(z)\times\Omega} c(x,y)\,dQ_{B_k(z)}(x,y,\omega)\\
&\quad= \mathfrak{c}_k \leq \mathfrak{c}_\infty.
\end{aligned}$$

(ii) In order to prove tightness of $(\dot{Q}_z^k)_{k\in\mathbb{N}}$, let

$$K_m := \left\{ y \in \mathbb{R}^d : \inf_{x\in B_0(z)} |x-y| \leq m \right\}$$

denote the closed m-neighborhood of the unit box based at z. Then

$$Q_z^k(\complement K_m \times B_0(z) \times \Omega) \le \frac{1}{\vartheta(m)} \int_{\mathbb{R}^d \times B_0(z) \times \Omega} c(x, y)\, dQ_z^k(x, y, \omega) \le \frac{1}{\vartheta(m)} \cdot \mathfrak{c}_\infty.$$

Since $\vartheta(m) \to \infty$ as $m \to \infty$ this proves tightness of the family $(\dot{Q}_z^k)_{k \in \mathbb{N}}$ on $\mathbb{R}^d \times \mathbb{R}^d \times \Omega$. (Recall that Ω was assumed to be compact from the very beginning.)

(iii) Tightness yields the existence of $\dot{Q}_z^\infty$ and of a converging subsequence for each z. A standard argument ("diagonal sequence") then gives convergence for all $z \in \mathbb{Z}^d$ along a common subsequence. □

LEMMA 4.2. (i) *For each $r > 0$ there exist numbers $\varepsilon_k(r)$ with $\varepsilon_k(r) \to 0$ as $k \to \infty$ such that for all $z, z' \in \mathbb{Z}^d$ and all $k \in \mathbb{N}$*

$$\int_\Gamma Q_{B_k(z',\gamma)}(A)\, d\nu(\gamma) \le \int_\Gamma Q_{B_k(z,\gamma)}(A)\, d\nu(\gamma) + \varepsilon_k(|z - z'|) \cdot \sup_\gamma Q_{B_k(z',\gamma)}(A)$$

for any Borel set $A \subset \mathbb{R}^d \times \mathbb{R}^d \times \Omega$.

(ii) *For all $z_1, \dots, z_m \in \mathbb{Z}^d$, all $k \in \mathbb{N}$ and all Borel sets $A \subset \mathbb{R}^d$,*

$$\sum_{i=1}^m \dot{Q}_{z_i}^k(A \times \mathbb{R}^d \times \Omega) \le \left(1 + \sum_{i=1}^m \varepsilon_k(|z_1 - z_i|)\right) \cdot \mathfrak{L}(A).$$

PROOF. (i) First, note that for each $z, z' \in \mathbb{Z}^d$, $k \in \mathbb{N}$, $\gamma \in \Gamma$,

$$z' \in B_k(z, \gamma) \quad \Longleftrightarrow \quad \exists \gamma' : B_k(z, \gamma) = B_k(z', \gamma')$$

and in this case

$$\nu(\{\gamma' : B_k(z', \gamma') = B_k(z, \gamma)\}) = 2^{-kd}.$$

Moreover,

$$\nu(\{\gamma : z' \notin B_k(z, \gamma)\}) \le \varepsilon_k(|z - z'|)$$

for some $\varepsilon_k(r)$ with $\varepsilon_k(r) \to 0$ as $k \to \infty$ for each $r > 0$. It implies that for each pair $z, z' \in \mathbb{Z}^d$ and each $k \in \mathbb{N}$,

$$\nu(\{\gamma \in \Gamma : \exists \gamma' : B_k(z, \gamma) = B_k(z', \gamma')\}) \ge 1 - \varepsilon_k(|z - z'|).$$

Therefore, for each Borel set $A \subset \mathbb{R}^d \times \mathbb{R}^d \times \Omega$,

$$\int_\Gamma Q_{B_k(z',\gamma)}(A)\, d\nu(\gamma) \le \int_\Gamma Q_{B_k(z,\gamma)}(A)\, d\nu(\gamma) + \varepsilon_k(|z - z'|) \cdot \sup_\gamma Q_{B_k(z',\gamma)}(A).$$

(ii) According to the previous part (i), for each Borel set $A \subset \mathbb{R}^d$,

$$
\begin{aligned}
&\sum_{i=1}^{m} \dot{Q}^k_{z_i}(A \times \mathbb{R}^d \times \Omega) \\
&\quad = \sum_{i=1}^{m} \int_\Gamma Q_{B_k(z_i,\gamma)}\big(A \times B_0(z_i) \times \Omega\big)\, d\nu(\gamma) \\
&\quad \le \sum_{i=1}^{m} \bigg[\int_\Gamma Q_{B_k(z_1,\gamma)}\big(A \times B_0(z_i) \times \Omega\big)\, d\nu(\gamma) \\
&\qquad\qquad + \varepsilon_k\big(|z_1 - z_i|\big) \cdot \sup_{\gamma\in\Gamma} Q_{B_k(z_i,\gamma)}\big(A \times B_0(z_i) \times \Omega\big)\bigg] \\
&\quad \le Q_{B_k(z_1,\gamma)}\big(A \times \mathbb{R}^d \times \Omega\big) + \sum_{i=1}^{m} \varepsilon_k\big(|z_1 - z_i|\big) \cdot \mathfrak{L}(A) \\
&\quad \le \left(1 + \sum_{i=1}^{m} \varepsilon_k\big(|z_1 - z_i|\big)\right) \cdot \mathfrak{L}(A).
\end{aligned}
$$

□

THEOREM 4.3. *The measure* $Q^\infty := \sum_{z\in\mathbb{Z}^d} \dot{Q}^\infty_z$ *is an optimal semicoupling of* $\mathfrak{L}$ *and* $\mu^\bullet$.

PROOF. (i) *Second/third marginal*: For any $f \in \mathcal{C}^+_b(\mathbb{R}^d \times \Omega)$ we have due to Lemma 4.1,

$$
\begin{aligned}
&\int_{\mathbb{R}^d\times\Omega} f(y,\omega)\, dQ^\infty(x,y,\omega) \\
&\quad = \sum_{z\in\mathbb{Z}^d} \int_{\mathbb{R}^d\times\Omega} f(y,\omega)\, d\dot{Q}^\infty_z(x,y,\omega) \\
&\quad = \sum_{z\in\mathbb{Z}^d} \lim_{l\to\infty} \int_{\mathbb{R}^d\times\Omega} f(y,\omega)\, d\dot{Q}^{k_l}_z(x,y,\omega) \\
&\quad = \sum_{z\in\mathbb{Z}^d} \int_{\mathbb{R}^d\times\Omega} f(y,\omega) 1_{B_0(z)}(y)\, d\big(\mu^\bullet\mathbb{P}\big)(y,\omega) \\
&\quad = \int_{\mathbb{R}^d\times\Omega} f(y,\omega)\, d\big(\mu^\bullet\mathbb{P}\big)(y,\omega).
\end{aligned}
$$

(ii) *First marginal*: Let an arbitrary bounded open set $A \subset \mathbb{R}^d$ be given, and let $(z_i)_{i\in\mathbb{N}}$ be an enumeration of $\mathbb{Z}^d$. According to the previous Lemma 4.2, for any $m \in \mathbb{N}$ and any $k \in \mathbb{N}$,

$$
\sum_{i=1}^{m} \dot{Q}^k_{z_i}(A \times \mathbb{R}^d \times \Omega) \le \left(1 + \sum_{i=1}^{m} \varepsilon_k\big(|z_1 - z_i|\big)\right) \cdot \mathfrak{L}(A).
$$

Letting first k tend to ∞ yields

$$\sum_{i=1}^{m} \dot{Q}_{z_i}^{\infty}\big(A \times \mathbb{R}^d \times \Omega\big) \le \mathfrak{L}(A).$$

Then with $m \to \infty$ we obtain

$$Q^{\infty}\big(A \times \mathbb{R}^d \times \Omega\big) \le \mathfrak{L}(A),$$

which proves that $(\pi_1)_* Q^{\infty} \le \mathfrak{L}$.

(iii) *Optimality*: By construction, Q^{∞} is $\mathbb{Z}^d$-equivariant. Due to the stationarity of $\mathbb{P}$, the asymptotic cost is given by

$$\begin{aligned}
&\int_{\mathbb{R}^d \times B_0(0) \times \Omega} c(x, y)\, dQ^{\infty}(x, y, \omega) \\
&\quad = \sum_{z \in \mathbb{Z}^d} \int_{\mathbb{R}^d \times B_0(0) \times \Omega} c(x, y)\, d\dot{Q}_z^{\infty}(x, y, \omega) \\
&\quad = \int_{\mathbb{R}^d \times B_0(0) \times \Omega} c(x, y)\, d\dot{Q}_0^{\infty}(x, y, \omega) \le \mathfrak{c}_{\infty}.
\end{aligned}$$

Here the final *inequality* is due to Lemma 4.1, property (i) (which remains true in the limit $k = \infty$), and the last *equality* comes from the fact that

$$\int_{\mathbb{R}^d \times B_0(u) \times \Omega} c(x, y)\, d\dot{Q}_z^k(x, y, \omega) = 0$$

for all $z \ne u$ and for all $k \in \mathbb{N}$ (which also remains true in the limit $k = \infty$). □

COROLLARY 4.4. (i) *For $k \to \infty$, the sequence of measures $Q^k := \sum_{z \in \mathbb{Z}^d} \dot{Q}_z^k$, $k \in \mathbb{N}$, converges vaguely to the unique optimal semicoupling Q^{∞}.*

(ii) *For each $z \in \mathbb{Z}^d$ the sequence $(Q_z^k)_{k \in \mathbb{N}}$ converges vaguely to the unique optimal semicoupling Q^{∞}.*

PROOF. (i) A slight extension of the previous Lemma 4.1(iii) + Theorem 4.3 yields that each subsequence $(Q^{k_n})_n$ of the above sequence $(Q^k)_k$ will have a sub-subsequence converging vaguely to an optimal coupling of $\mathfrak{L}$ and $\mu^{\bullet}$. Since the optimal coupling is unique, all these limit points coincide. Hence, the whole sequence $(Q^k)_k$ converges to this limit point; see, for example, [10], Proposition 9.3.1.

(ii) Lemma 4.2(i) implies that for $z, z', u \in \mathbb{Z}^d$ and every measurable $A \subset \mathbb{R}^d \times \mathbb{R}^d \times \Omega$,

$$\begin{aligned}
&\big|Q_z^k\big(A \cap \big(\mathbb{R}^d \times B_0(u) \times \Omega\big)\big) - Q_{z'}^k\big(A \cap \big(\mathbb{R}^d \times B_0(u) \times \Omega\big)\big)\big| \\
&\quad \le \varepsilon_k\big(|z - z'|\big) \cdot \sup_{v \in \mathbb{Z}^d} Q_{B_k(v)}\big(A \cap \big(\mathbb{R}^d \times B_0(u) \times \Omega\big)\big) \\
&\quad \le \varepsilon_k\big(|z - z'|\big) \to 0
\end{aligned}$$

as $k \to \infty$. Hence, for each $f \in \mathcal{C}_c(\mathbb{R}^d \times \mathbb{R}^d \times \Omega)$ and each $z' \in \mathbb{R}^d$,

$$\left|\sum_{z \in \mathbb{Z}^d} \int f(x, y, \omega) 1_{B_0(z)}(y)\, dQ_z^k - \int f(x, y, \omega)\, dQ_{z'}^k\right| \to 0.$$

That is, $|\int f\, dQ^k - \int f\, dQ_{z'}^k| \to 0$ as $k \to \infty$. □

COROLLARY 4.5. *We have* $\mathfrak{c}_\infty = \inf_{q^\bullet \in \Pi_s} \mathfrak{C}_\infty(q^\bullet)$ *where* Π_s *denotes the set of all semicouplings* $q^\bullet$ *of* $\mathfrak{L}$ *and* $\mu^\bullet$. *In particular, the following holds*:

$$\begin{aligned}\inf_{q^\bullet \in \Pi_s} \liminf_{n \to \infty} \frac{1}{\mathfrak{L}(B_n)} \mathbb{E}\left[\int_{\mathbb{R}^d \times B_n} c(x, y)\, dq^\bullet(x, y)\right] \\ = \liminf_{n \to \infty} \inf_{q^\bullet \in \Pi_s} \frac{1}{\mathfrak{L}(B_n)} \mathbb{E}\left[\int_{\mathbb{R}^d \times B_n} c(x, y)\, dq^\bullet(x, y)\right].\end{aligned}$$

PROOF. The optimal coupling Q constructed in the previous theorem has mean asymptotic transportation cost bounded above by $\mathfrak{c}_\infty$. Thus, we have $\inf_{q^\bullet \in \Pi_s} \mathfrak{C}_\infty(q^\bullet) \le \mathfrak{c}_\infty$. Together with Lemma 2.3, this yields the claim. □

4.2. *Quenched limits*. According to Section 3, the unique optimal semicoupling between $d\mathfrak{L}(x)$ and $d\mu^\omega(y)\, d\mathbb{P}(\omega)$ can be represented on $\mathbb{R}^d \times \mathbb{R}^d \times \Omega$ as

$$dQ^\infty(x, y, \omega) = d\delta_{T(x,\omega)}(y)\, d\mathfrak{L}(x)\, d\mathbb{P}(\omega)$$

by means of a measurable map

$$T : \mathbb{R}^d \times \Omega \to \mathbb{R}^d \cup \{\eth\},$$

defined uniquely almost everywhere. Similarly, for each $z \in Z^d$ and $k \in \mathbb{N}$, there exists a measurable map

$$T_{z,k} : \mathbb{R}^d \times \Omega \times \Gamma \to \mathbb{R}^d \cup \{\eth\}$$

such that for each $\gamma \in \Gamma$ the measure

$$dQ_{B_k(z,\gamma)}(x, y, \omega) = d\delta_{T_{z,k}(x,\omega,\gamma)}(y)\, d\mathfrak{L}(x)\, d\mathbb{P}(\omega)$$

on $\mathbb{R}^d \times \mathbb{R}^d \times \Omega$ is the unique optimal semicoupling between $d\mathfrak{L}(x)$ and $1_{B_k(z,\gamma)}(y)\, d\mu^\omega(y)\, d\mathbb{P}(\omega)$.

PROPOSITION 4.6. *For every* $z \in Z^d$,

$$T_{z,k}(x, \omega, \gamma) \to T(x, \omega) \qquad \text{as } k \to \infty \text{ locally in } \mathfrak{L} \otimes \mathbb{P} \otimes \nu\text{-measure}.$$

The claim basically relies on the following lemma which is a slight modification (and extension) of a result in [2].

LEMMA 4.7. *Let X, Y be locally compact Polish spaces, θ a Radon measure on X and ρ a metric on Y compatible with the topology.*

(i) *For all $n \in \mathbb{N}$ let $T_n, T : X \to Y$ be Borel measurable maps. Put $dQ_n(x, y) := d\delta_{T_n(x)}(y)\, d\theta(x)$ and $dQ(x, y) := d\delta_{T(x)}(y)\, d\theta(x)$. Then*

$$T_n \to T \text{ locally in measure on } X \quad \Longleftrightarrow \quad Q_n \to Q \text{ vaguely in } \mathcal{M}(X \times Y).$$

(ii) *More generally, let T and Q be as before whereas*

$$dQ_n(x, y) := \int_{X'} d\delta_{T_n(x,x')}(y)\, d\theta'(x')\, d\theta(x)$$

for some probability space $(X', \mathfrak{A}', \theta')$ and suitable measurable maps $T_n : X \times X' \to Y$. Then

$$Q_n \to Q \text{ vaguely in } \mathcal{M}(X \times Y) \quad \Longrightarrow \quad T_n(x, x') \to T(x) \text{ locally in measure on } X \times X'.$$

PROOF. (i) Assume $T_n \to T$ in θ-measure. Then also $f \circ (\mathrm{id}, T_n) \to f \circ (\mathrm{id}, T)$ in θ-measure for any $f \in C_c(X \times Y)$. Therefore, by the dominated convergence theorem we have

$$\int f(x, y)\, dQ_n = \int f(x, T_n(x))\, d\theta \to \int f(x, T(x))\, d\theta = \int f(x, y)\, dQ.$$

This proves the vague convergence of Q_n toward Q.

For the opposite direction, fix $\tilde{K} \subset X$ compact and $\varepsilon > 0$. By Lusin's theorem there is a compact set $K \subset \tilde{K}$ such that $T|_K$ is continuous and $\theta(\tilde{K} \setminus K) < \varepsilon$. Put $\eta : \mathbb{R}_+ \to \mathbb{R}_+, t \mapsto 1 \wedge |t|/\varepsilon$. The function

$$\phi(x, y) = 1_K(x)\eta(\rho(y, T(x)))$$

is upper semicontinuous, nonnegative and compactly supported. Thus, there exist $\phi_l \in C_c(X \times Y)$ with $\phi_l \searrow \phi$. By assumption, we have for each l

$$\int \phi(x, y)\, dQ_n(x, y) \le \int \phi_l(x, y)\, dQ_n(x, y) \overset{n\to\infty}{\to} \int \phi_l(x, y)\, dQ(x, y).$$

Moreover,

$$\int \phi_l(x, y)\, dQ(x, y) \overset{l\to\infty}{\to} \int \phi(x, y)\, dQ(x, y) = 0.$$

Therefore, $\lim_{n\to\infty} \int \phi(x, y)\, dQ_n(x, y) = 0$. In other words,

$$\lim_{n\to\infty} \int 1_K(x)\eta(\rho(T_n(x), T(x)))\, d\theta(x) = 0.$$

This implies $\lim_{n\to\infty} \theta(\{x \in K : \rho(T_n(x), T(x)) \ge \varepsilon\}) = 0$ and then in turn

$$\lim_{n\to\infty} \theta(\{x \in \tilde{K} : \rho(T_n(x), T(x)) \ge 2\varepsilon\}) = 0.$$

(ii) Given any compact $\tilde{K} \subset X$ and any $\varepsilon > 0$, choose ϕ as before. Then vague convergence again implies $\lim_{n\to\infty} \int \phi(x, y)\, dQ_n(x, y) = 0$. This, in other words, now reads as

$$\lim_{n\to\infty} \int_X \int_{X'} 1_K(x)\eta\bigl(\rho\bigl(T_n\bigl(x, x'\bigr), T(x)\bigr)\bigr)\, d\theta'\bigl(x'\bigr)\, d\theta(x) = 0.$$

Therefore,

$$\lim_{n\to\infty} (\theta \otimes \theta')\bigl(\bigl\{\bigl(x, x'\bigr) \in \tilde{K} \times X' : \rho\bigl(T_n\bigl(x, x'\bigr), T(x)\bigr) \geq 2\varepsilon\bigr\}\bigr) = 0.$$

This is the claim. □

PROOF OF PROPOSITION 4.6. Fix $z \in Z^d$ and recall that

$$Q_z^k \to Q^\infty \qquad \text{vaguely on } \mathbb{R}^d \times \mathbb{R}^d,$$

where

$$dQ^\infty(x, y, \omega) = d\delta_{T(x,\omega)}(y)\, d\mathfrak{L}(x)\, d\mathbb{P}(\omega)$$

and

$$\begin{aligned} dQ_z^k(x, y, \omega) &= \int_\Gamma dQ_{B_k(z,\gamma)}(x, y, \omega)\, d\nu(\gamma) \\ &= \int_\Gamma d\delta_{T_{z,k}(x,\omega,\gamma)}(y)\, d\mathfrak{L}(x)\, d\mathbb{P}(\omega)\, d\nu(\gamma) \end{aligned}$$

with transport maps $T : \mathbb{R}^d \times \Omega \to \mathbb{R}^d \cup \{\eth\}$ and $T_{z,k} : \mathbb{R}^d \times \Omega \times \Gamma \to \mathbb{R}^d \cup \{\eth\}$ as above. Apply assertion (ii) of the previous lemma with $X := \mathbb{R}^d \times \Omega$, $X' = \Gamma$, $Y = \mathbb{R}^d \cup \{\eth\}$ and $\theta = \mathfrak{L} \otimes \mathbb{P}$, $\theta' = \nu$. □

Actually, this convergence result can significantly be improved.

THEOREM 4.8. *For every $z \in Z^d$ and every bounded Borel set $M \subset \mathbb{R}^d$,*

$$\lim_{k\to\infty} (\mathfrak{L} \otimes \mathbb{P} \otimes \nu)\bigl(\bigl\{(x, \omega, \gamma) \in M \times \Omega \times \Gamma : T_{z,k}(x, \omega, \gamma) \neq T(x, \omega)\bigr\}\bigr) = 0.$$

PROOF. Let M as above and $\varepsilon > 0$ be given. Finiteness of the asymptotic mean transportation cost implies that there exists a bounded set $M' \subset \mathbb{R}^d$ such that

$$(\mathfrak{L} \otimes \mathbb{P})\bigl(\bigl\{(x, \omega) \in M \times \Omega : T(x, \omega) \notin M'\bigr\}\bigr) \leq \varepsilon.$$

Given the bounded set M' there exists $\delta > 0$ such that the probability to find two distinct particles of the point process at distance $< \delta$, at least one of them within M', is less than ε, that is,

$$\mathbb{P}\bigl(\bigl\{\omega : \exists \bigl(y, y'\bigr) \in M' \times \mathbb{R}^d : 0 < \bigl|y - y'\bigr| < \delta, \mu^\omega\bigl(\{y\}\bigr) > 0, \mu^\omega\bigl(\{y'\}\bigr) > 0\bigr\}\bigr) \leq \varepsilon.$$

On the other hand, Proposition 4.6 states that with high probability the maps T and $T_{z,k}$ have distance less than δ. More precisely, for each $\delta > 0$ there exists k_0 such that for all $k \geq k_0$,

$$(\mathfrak{L} \otimes \mathbb{P} \otimes \nu)\big(\{(x, \omega, \gamma) \in M \times \Omega \times \Gamma : \big|T_{z,k}(x, \omega, \gamma) - T(x, \omega)\big| \geq \delta\}\big) \leq \varepsilon.$$

Since all the maps T and $T_{z,k}$ take values in the support of the point process (plus the point $\eth$) it follows that

$$(\mathfrak{L} \otimes \mathbb{P} \otimes \nu)\big(\{(x, \omega, \gamma) \in M \times \Omega \times \Gamma : T_{z,k}(x, \omega, \gamma) \neq T(x, \omega)\}\big) \leq 3\varepsilon$$

for all $k \geq k_0$. □

COROLLARY 4.9. *There exists a subsequence $(k_l)_l$ such that*

$$T_{z,k_l}(x, \omega, \gamma) \to T(x, \omega) \qquad \text{as } l \to \infty$$

for almost every $x \in \mathbb{R}^d$, $\omega \in \Omega$, $\gamma \in \Gamma$ and every $z \in Z^d$. Indeed, the sequence $(T_{z,k_l})_l$ is finally stationary. That is, there exists a random variable $l_z : \mathbb{R}^d \times \Omega \times \Gamma \to \mathbb{N}$ such that almost surely

$$T_{z,k_l}(x, \omega, \gamma) = T(x, \omega) \qquad \text{for all } l \geq l_z(x, \omega, \gamma).$$

COROLLARY 4.10. *There is a measurable map $\Upsilon : \mathcal{M}(\mathbb{R}^d) \to \mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d)$ s.t. $q^\omega := \Upsilon(\mu^\omega)$ denotes the unique optimal semicoupling between $\mathfrak{L}$ and μ^ω. In particular the optimal semicoupling is a factor coupling.*

PROOF. By Theorem 2.1, the maps $T_{z,k}$ are measurable with respect to the sigma algebra generated by $\mu^\bullet$. By Theorem 4.8, the optimal transportation map T is also measurable with respect to the sigma algebra generated by $\mu^\bullet$. Because the optimal semicoupling $q^\bullet$ is given by $q^\omega = (\mathrm{id}, T^\omega)_* \mathfrak{L}$, it is also measurable with respect to the sigma algebra generated by $\mu^\bullet$. Thus there is a measurable map Υ such that $q^\bullet = \Upsilon(\mu^\bullet)$. □

5. Estimates for the asymptotic mean transportation cost of a Poisson process. Throughout this section, $\mu^\bullet$ will be a Poisson point process of intensity $\beta \leq 1$. The asymptotic mean transportation cost for $\mu^\bullet$ will be denoted by

$$\mathfrak{c}_\infty = \mathfrak{c}_\infty(\vartheta, d, \beta)$$

or, if $\vartheta(r) = r^p$, by $\mathfrak{c}_\infty(p, d, \beta)$. We will present sufficient as well as necessary conditions for finiteness of $\mathfrak{c}_\infty$. These criteria will be quite sharp. Moreover, in the case of L^p-cost, we also present explicit sharp estimates for $\mathfrak{c}_\infty$.

To begin with, let us summarize some elementary monotonicity properties of $\mathfrak{c}_\infty(\vartheta, d, \beta)$.

LEMMA 5.1. (i) $\vartheta \le \overline{\vartheta}$ *implies* $\mathfrak{c}_\infty(\vartheta, d, \beta) \le \mathfrak{c}_\infty(\overline{\vartheta}, d, \beta)$.
More generally, $\limsup_{r\to\infty} \frac{\overline{\vartheta}(r)}{\vartheta(r)} < \infty$ *and* $\mathfrak{c}_\infty(\vartheta, d, \beta) < \infty$ *imply* $\mathfrak{c}_\infty(\overline{\vartheta}, d, \beta) < \infty$.

(ii) *If* $\overline{\vartheta} = \varphi \circ \vartheta$ *for some convex increasing* $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$, *then*

$$\varphi\bigl(\beta^{-1}\mathfrak{c}_\infty(\vartheta, d, \beta)\bigr) \le \beta^{-1}\mathfrak{c}_\infty(\overline{\vartheta}, d, \beta).$$

(iii) $\beta \le \overline{\beta}$ *implies* $\mathfrak{c}_\infty(\vartheta, d, \beta) \le \mathfrak{c}_\infty(\vartheta, d, \overline{\beta})$.

PROOF. (i) Is obvious. (ii) If $\overline{q}$ denotes the optimal semicoupling for $\overline{\vartheta}$, then Jensen's inequality implies

$$\begin{aligned}
&\beta^{-1}\mathfrak{c}_\infty(\overline{\vartheta}, d, \beta)\\
&\quad = \beta^{-1}\mathbb{E}\int_{\mathbb{R}^d\times[0,1)^d} \varphi\bigl(\vartheta(|x-y|)\bigr)\,d\overline{q}(x,y)\\
&\quad \ge \varphi\left(\beta^{-1}\mathbb{E}\int_{\mathbb{R}^d\times[0,1)^d} \vartheta(|x-y|)\,d\overline{q}(x,y)\right) \ge \varphi\bigl(\beta^{-1}\mathfrak{c}_\infty(\vartheta, d, \beta)\bigr).
\end{aligned}$$

(iii) Given a realization $\overline{\mu}^\omega$ of a Poisson point process with intensity $\overline{\beta}$. Delete each point $\xi \in \text{supp}[\overline{\mu}^\omega]$ with probability $1 - \beta/\overline{\beta}$, independently of each other. Then the remaining point process μ^ω is a Poisson point process with intensity β. Hence, each semicoupling $\overline{q}^\omega$ between $\mathfrak{L}$ and $\overline{\mu}^\omega$ leads to a semicoupling q^ω between $\mathfrak{L}$ and μ^ω with less or equal transportation cost. The centers which survive are coupled with the same cells as before. □

5.1. *Lower estimates.*

THEOREM 5.2 ([14]). *Assume* $\beta = 1$ *and* $d \le 2$. *Then for all translation invariant couplings of Lebesgue and Poisson*

$$\mathbb{E}\left[\int_{\mathbb{R}^d\times[0,1)^d} |x-y|^{d/2}\,dq^\bullet(x,y)\right] = \infty.$$

THEOREM 5.3. *For all* $\beta \le 1$ *and* $d \ge 1$ *there exists a constant* $\kappa' = \kappa'(d, \beta)$ *such that for all translation invariant semicouplings of Lebesgue and Poisson*

$$\mathbb{E}\left[\int_{\mathbb{R}^d\times[0,1)^d} \exp\bigl(\kappa'|x-y|^d\bigr)\,dq^\bullet(x,y)\right] = \infty.$$

The result is well known in the case $\beta = 1$. In this case, it is based on a lower bound for the event "no Poisson particle in the cube $[-r, r)^d$" and on a lower estimate for the cost of transporting the Lebesgue measure in $[-r/2, r/2)^d$ to some distribution on $\mathbb{R}^d \setminus [-r, r)^d$,

$$\mathfrak{c}_\infty \ge \exp\bigl(-(2r)^d\bigr)\cdot\vartheta\left(\frac{r}{2}\right)\cdot 2^{-d}.$$

Hence, $\mathfrak{c}_\infty \to \infty$ as $r \to \infty$ if $\vartheta(r) = \exp(\kappa' r^d)$ with $\kappa' > 2^{2d}$.

However, this argument breaks down in the case $\beta < 1$. We will present a different argument which works for all $\beta \le 1$.

PROOF. Consider the event "more than $(3r)^d$ Poisson particles in the box $[-r/2, r/2)^d$" or, formally,

$$\Omega(r) = \{\mu^\bullet([-r/2, r/2)^d) \ge (3r)^d\}.$$

Note that $\mathbb{E}\mu^\bullet([-r/2, r/2)^d) = \beta r^d$ with $\beta \le 1$. For $\omega \in \Omega(r)$, the cost of a semi-coupling between $\mathfrak{L}$ and $1_{[-r/2,r/2)^d}\mu^\omega$ is bounded from below by

$$\vartheta(r/2) \cdot r^d$$

(since r^d Poisson points—or more—must be transported at least a distance $r/2$). The large deviation result formulated in the next lemma allows us to estimate

$$\mathbb{P}(\Omega(r_n)) \ge e^{-k \cdot r_n{}^d}$$

for any $k > I_\beta(3^d)$ and suitable $r_n \to \infty$. Hence, if $\vartheta(r) \ge \exp(\kappa' r^d)$ with $\kappa' > 2^d \cdot k$, then

$$\mathfrak{c}_\infty \ge \mathbb{P}(\Omega(r_n)) \cdot \vartheta(r/2) \ge \exp((\kappa' 2^{-d} - k) r^d) \to \infty$$

as $r \to \infty$. $\square$

LEMMA 5.4. *Given any nested sequence of boxes $B_n(z, \gamma) \subset \mathbb{R}^d$ and $t \ge \beta$*

$$\lim_{n\to\infty} \frac{-1}{2^{nd}} \log \mathbb{P}\left[\frac{1}{2^{nd}} \mu^\bullet(B_n(z, \gamma)) \ge t\right] = I_\beta(t)$$

with $I_\beta(t) = t \log(t/\beta) - t + \beta$.

PROOF. For a fixed sequence $B_n(z, \gamma)$, $n \in \mathbb{N}$, consider the sequence of random variables $Z_n(\cdot) = \mu^\bullet(B_n(z, \gamma))$. For each $n \in \mathbb{N}$,

$$Z_n = \sum_{i \in B_n(z,\gamma) \cap \mathbb{Z}^d} X_i$$

with $X_i = \mu^\bullet(B_0(i))$. The X_i are i.i.d. Poisson random variables with mean β. Hence, Cramér's theorem states that for all $t \ge \beta$,

$$\liminf_{n\to\infty} \frac{-1}{2^{nd}} \log \mathbb{P}\left[\frac{1}{2^{nd}} Z_n \ge t\right] \ge I_\beta(t)$$

with

$$I_\beta(t) = \sup_x [tx - \log \hat{\mu}(x)] = t \log(t/\beta) - t + \beta. \qquad \square$$

5.2. *Upper estimates for concave cost.* In this section we treat the case of a concave scale function ϑ. In particular this implies that the cost function $c(x,y)=\vartheta(|x-y|)$ defines a metric on $\mathbb{R}^d$. The results of this section will be mainly of interest in the case $d\le 2$; in particular, they will prove assertion (ii) of Theorem 1.3. It suffices to consider the case $\beta=1$. Similar to the early work of Ajtai, Komlós and Tusnády [1], our approach will be based on iterated transports between cuboids of doubled edge length.

We put

$$\Theta(r):=\int_0^r \vartheta(s)\,ds \quad \text{and} \quad \varepsilon(r):=\sup_{s\ge r}\frac{\vartheta(s)}{s^{d/2}}. \tag{7}$$

5.2.1. *Modified cost.* In order to prove the finiteness of the asymptotic mean transportation cost, we will estimate the cost of a semicoupling between $\mathfrak{L}$ and $1_A\mu^\bullet$ from above in terms of the cost of another, related coupling.

Given two measure-valued random variables $\nu_1^\bullet,\nu_2^\bullet:\Omega\to\mathcal{M}(\mathbb{R}^d)$ with $\nu_1^\omega(\mathbb{R}^d)=\nu_2^\omega(\mathbb{R}^d)$ for a.e. $\omega\in\Omega$, we define their transportation distance by

$$\mathbb{W}_\vartheta(\nu_1,\nu_2):=\int_\Omega W_\vartheta\big(\nu_1^\omega,\nu_2^\omega\big)\,d\mathbb{P}(\omega),$$

where

$$W_\vartheta(\eta_1,\eta_2)=\inf\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d}\vartheta\big(|x-y|\big)\,dq(x,y): q \text{ is coupling of } \eta_1,\eta_2\right\}$$

denotes the usual L^1-Wasserstein distance—w.r.t. the distance $\vartheta(|x-y|)$—between (not necessarily normalized) measures $\eta_1,\eta_2\in\mathcal{M}(\mathbb{R}^d)$ of equal total mass.

LEMMA 5.5. (i) *For any triple of random measures* $\nu_1^\bullet,\nu_2^\bullet,\nu_3^\bullet:\Omega\to\mathcal{M}(\mathbb{R}^d)$ *with* $\nu_1^\omega(\mathbb{R}^d)=\nu_2^\omega(\mathbb{R}^d)=\nu_3^\omega(\mathbb{R}^d)$ *for a.e.* $\omega\in\Omega$, *we have the triangle inequality*

$$\mathbb{W}_\vartheta(\nu_1,\nu_3)\le\mathbb{W}_\vartheta(\nu_1,\nu_2)+\mathbb{W}_\vartheta(\nu_2,\nu_3).$$

(ii) *For each countable family of pairs of measure-valued random variables* $\nu_{1,k}^\bullet,\nu_{2,k}^\bullet:\Omega\to\mathcal{M}(\mathbb{R}^d)$ *with* $\nu_{1,k}^\omega(\mathbb{R}^d)=\nu_{2,k}^\omega(\mathbb{R}^d)$ *for a.e.* $\omega\in\Omega$ *and all* k *we have*

$$\mathbb{W}_\vartheta\bigg(\sum_k\nu_{1,k}^\bullet,\sum_k\nu_{2,k}^\bullet\bigg)\le\sum_k\mathbb{W}_\vartheta\big(\nu_{1,k}^\bullet,\nu_{2,k}^\bullet\big).$$

PROOF. Gluing lemma (cf. [10] or [32], Chapter 1) plus Minkowski inequality yield (i); (ii) is obvious. □

For each bounded measurable $A\subset\mathbb{R}^d$ let us now define a random measure $\nu_A^\bullet:\Omega\to\mathcal{M}(\mathbb{R}^d)$ by

$$\nu_A^\omega:=\frac{\mu^\omega(A)}{\mathfrak{L}(A)}\cdot 1_A\mathfrak{L}.$$

Note that—by construction—the measures ν_A^ω and $1_A\mu^\omega$ have the same total mass. The *modified transportation cost* is defined as

$$\widehat{\mathsf{C}}_A(\omega) = \inf\left\{\int c(x,y)\,d\widehat{q}(x,y) : \widehat{q} \text{ is coupling of } \nu_A^\omega \text{ and } 1_A\mu^\omega\right\}$$
$$= W_\vartheta\big(\nu_A^\omega, 1_A\mu^\omega\big).$$

Put

$$\widehat{\mathfrak{c}}_n = 2^{-nd}\cdot\mathbb{E}[\widehat{\mathsf{C}}_{B_n}]$$

with $B_n = [0, 2^n)^d$ as usual.

5.2.2. *Semi-subadditivity of modified cost.* The crucial advantage of this modified cost function $\widehat{\mathsf{C}}_A$ is that it is semi-subadditive (i.e., subadditive up to correction terms) on suitable classes of *cuboids* which we are going to introduce now. For $n \in \mathbb{N}_0$, $k \in \{1,\dots,d\}$ and $i \in \{0,1\}^k$, put

$$B_{n+1}^i := \big[0, 2^n\big)^k \times \big[0, 2^{n+1}\big)^{d-k} + 2^n\cdot(i_1,\dots,i_k,0,\dots,0).$$

These cuboids can be constructed by iterated subdivision of the standard cube B_{n+1} as follows: We start with $B_{n+1} = [0, 2^{n+1})^d$ and subdivide it (along the first coordinate) into two disjoint congruent pieces $B_{n+1}^{(0)} = [0, 2^n) \times [0, 2^{n+1})^{d-1}$ and $B_{n+1}^{(1)} = B_{n+1}^{(0)} + 2^n\cdot(1,0,\dots,0)$. In the kth step, we subdivide each of the $B_{n+1}^i = B_{n+1}^{(i_1,\dots,i_{k-1})}$ for $i \in \{0,1\}^{k-1}$ along the kth coordinate into two disjoint congruent pieces $B_{n+1}^{(i_1,\dots,i_{k-1},0)}$ and $B_{n+1}^{(i_1,\dots,i_{k-1},1)}$. After d steps we are done. Each of the B_{n+1}^i for $i \in \{0,1\}^d$ is a copy of the standard cube B_n, more precisely,

$$B_{n+1}^i = B_n + 2^n\cdot i.$$

LEMMA 5.6. *Given $n \in \mathbb{N}_0, k \in \{1,\dots,d\}$ and $i \in \{0,1\}^k$ put $D_0 = B_{n+1}^{(i_1,\dots,i_{k-1},0)}$, $D_1 = B_{n+1}^{(i_1,\dots,i_{k-1},1)}$ and $D = D_0 \cup D_1 = B_{n+1}^{(i_1,\dots,i_{k-1})}$. Then*

$$\mathbb{W}_\vartheta(\nu_{D_0} + \nu_{D_1}, \nu_D) \le 2^{-(n+1)}\Theta\big(2^{n+1}\big)2^{d/2(n+1)-k/2},$$

with Θ as defined in (7).

PROOF. Put $Z_j(\omega) := \mu^\omega(D_j)$ for $j \in \{0,1\}$. Then Z_0, Z_1 are independent Poisson random variables with parameter $\alpha_0 = \alpha_1 = \mathfrak{L}(D_j) = 2^{d(n+1)-k}$, and $Z := \mu(D) = Z_0 + Z_1$ is a Poisson random variable with parameter $\alpha = 2^{d(n+1)-k+1}$.

The measure ν_D has density $\frac{Z}{\alpha}$ on D whereas the measure $\tilde{\nu}_D := \nu_{D_0} + \nu_{D_1}$ has density $\frac{2Z_0}{\alpha}$ on the part $D_0 \subset D$ and it has density $\frac{2Z_1}{\alpha}$ on the remaining part $D_1 \subset D$. If $Z = 0$ nothing has to be transported since $\tilde{\nu}$ already coincides with ν. Hence, for the sequel we may assume $Z > 0$.

Assume that $Z_0 > Z_1$. Then a total amount of mass $\frac{Z_0 - Z_1}{2}$, uniformly distributed over D_0, will be transported with the map

$$T : (x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_d) \mapsto \big(x_1, \dots, x_{k-1}, 2^{n+1} - x_k, x_{k+1}, \dots, x_d\big)$$

from D_0 to D_1. The rest of the mass remains where it is. Hence, the cost of this transport is

$$\frac{|Z_0 - Z_1|}{2} \cdot 2^{-n} \int_0^{2^n} \vartheta\big(2^{n+1} - 2x_k\big)\, dx_k = 2^{-(n+2)} \Theta\big(2^{n+1}\big) \cdot |Z_0 - Z_1|.$$

Hence, we get

$$\begin{aligned} \mathbb{W}_\vartheta(\tilde{\nu}_D, \nu_D) &= 2^{-(n+2)} \Theta\big(2^{n+1}\big) \cdot \mathbb{E}\big[|Z_0 - Z_1|\big] \\ &\le 2^{-(n+1)} \Theta\big(2^{n+1}\big) \cdot \mathbb{E}\big[|Z_0 - \alpha_0|\big] \\ &\le 2^{-(n+1)} \Theta\big(2^{n+1}\big) \cdot \alpha_0^{1/2} = 2^{-(n+1)} \Theta\big(2^{n+1}\big) 2^{d/2(n+1) - k/2}. \qquad \square \end{aligned}$$

PROPOSITION 5.7. *For all $n \in \mathbb{N}$ and arbitrary dimension d the following holds*:

$$\widehat{\mathfrak{c}}_{n+1} \le \widehat{\mathfrak{c}}_n + 2^{d/2+1} \cdot 2^{-(n+1)(d/2+1)} \Theta\big(2^{n+1}\big).$$

PROOF. By definition

$$\mathbb{W}_\vartheta(1_{B_{n+1}}\mu, \nu_{B_{n+1}}) = 2^{d(n+1)} \cdot \widehat{\mathfrak{c}}_{n+1},$$

and it is easily observed that

$$\begin{aligned} \mathbb{W}_\vartheta\Bigg(1_{B_{n+1}}\mu, \sum_{i \in \{0,1\}^d} \nu_{B_n^i}\Bigg) &\le \sum_{i \in \{0,1\}^d} \mathbb{W}_\vartheta(1_{B_n^i}\mu, \nu_{B_n^i}) \\ &= 2^d \cdot \mathbb{W}_\vartheta(1_{B_n}\mu, \nu_{B_n}) = 2^{d(n+1)} \cdot \widehat{\mathfrak{c}}_n. \end{aligned}$$

Hence, by the triangle inequality for $\mathbb{W}_\vartheta$ an upper estimate for $\widehat{\mathfrak{c}}_{n+1} - \widehat{\mathfrak{c}}_n$ will follow from an upper bound for $\mathbb{W}_\vartheta(\sum_{i \in \{0,1\}^d} \nu_{B_n^i}, \nu_{B_{n+1}})$.

In order to estimate the cost of transportation from $\nu_{(d)} := \sum_{i \in \{0,1\}^d} \nu_{B_n^i}$ to $\nu_{(0)} := \nu_{B_{n+1}}$ for fixed $n \in \mathbb{N}_0$, we introduce $(d-1)$ further ("intermediate") measures

$$\nu_{(k)} = \sum_{i \in \{0,1\}^k} \nu_{B_{n+1}^i}$$

and estimate the cost of transportation from $\nu_{(k)}$ to $\nu_{(k-1)}$ for $k \in \{1, \dots, d\}$. For each k, these cost arise from *merging* 2^{k-1} *pairs of cuboids* into 2^{k-1} cuboids of twice the size. More precisely, from moving mass within pairs of adjacent cuboids in order to obtain equilibrium in the unified cuboid of twice the size. These costs—for each of the 2^{k-1} pairs involved—have been estimated in the previous lemma,

$$\begin{aligned}\mathbb{W}_\vartheta(\nu_{(k)}, \nu_{(k-1)}) &\le 2^{k-1} \cdot \mathbb{W}_\vartheta(\nu_{B^{i,0}_{n+1}} + \nu_{B^{i,1}_{n+1}}, \nu_{B^i_{n+1}}) \\ &\le 2^{k-1} \cdot 2^{-(n+1)} \Theta(2^{n+1}) 2^{d/2(n+1)-k/2}\end{aligned}$$

for $k \in \{1, \dots, d\}$ (and arbitrary $i \in \{0, 1\}^{k-1}$). Thus

$$\begin{aligned}2^{d(n+1)} \cdot [\widehat{\mathfrak{c}}_{n+1} - \widehat{\mathfrak{c}}_n] &\le \mathbb{W}_\vartheta(1_{B_{n+1}}\mu, \nu_{(0)}) - \mathbb{W}_\vartheta(1_{B_{n+1}}\mu, \nu_{(d)}) \\ &\le \sum_{k=1}^d \mathbb{W}_\vartheta(\nu_{(k-1)}, \nu_{(k)}) \\ &\le \sum_{k=1}^d 2^{k/2} \cdot 2^{-(n+2)} \Theta(2^{n+1}) 2^{d/2(n+1)} \\ &\le 4 \cdot 2^{(n+2)(d/2-1)} \cdot \Theta(2^{n+1}),\end{aligned}$$

which yields the claim. □

COROLLARY 5.8. *If* $\sum_{n\ge 1} 2^{-(n+1)(d/2+1)} \Theta(2^{n+1}) < \infty$, *we have*

$$\widehat{\mathfrak{c}}_\infty := \lim_{n\to\infty} \widehat{\mathfrak{c}}_n$$

exists and is finite.

PROOF. According to the previous proposition,

$$\lim_{n\to\infty} \widehat{\mathfrak{c}}_n \le \widehat{\mathfrak{c}}_N + \sum_{m\ge N} 2^{-(m+1)(d/2+1)} \Theta(2^{m+1}) \tag{8}$$

for each $N \in \mathbb{N}$. As the sum was assumed to converge, the claim follows. □

5.2.3. *Comparison of costs.* Recall the definition of $\mathfrak{c}_n$ from Section 2.7.

PROPOSITION 5.9. *For all* $d \in \mathbb{N}$ *and for all* $n \in \mathbb{N}_0$,

$$\mathfrak{c}_n \le \widehat{\mathfrak{c}}_n + \sqrt{2d} \cdot \varepsilon(2^n).$$

PROOF. Let a box $B = B_n = [0, 2^n)^d$ for some fixed $n \in \mathbb{N}_0$ be given. We define a measure-valued random variable $\lambda^\bullet_B : \Omega \to \mathcal{M}(\mathbb{R}^d)$ by

$$\lambda^\omega_B = 1_{\widehat{B}(\omega)} \cdot \mathfrak{L}$$

with a randomly scaled box $\widehat{B}(\omega) = [0, Z(\omega)^{1/d})^d \subset \mathbb{R}^d$ and $Z(\omega) = \mu^\omega(B)$. Recall that Z is a Poisson random variable with parameter $\alpha = 2^{nd}$. Moreover, note that

$$\lambda^\omega_B(\mathbb{R}^d) = \mu^\omega(B) = \nu^\omega_B(\mathbb{R}^d)$$

and that $\lambda_B^\omega \le \mathfrak{L}$ for each $\omega \in \Omega$. Each coupling of λ_B^ω of $1_B\mu^\omega$, therefore, is also a semicoupling of $\mathfrak{L}$ and $1_B\mu^\omega$. Hence,

$$2^{nd} \cdot \mathfrak{c}_n \le \mathbb{W}_\vartheta(\lambda_B, 1_B\mu).$$

On the other hand, obviously,

$$2^{nd} \cdot \widehat{\mathfrak{c}}_n = \mathbb{W}_\vartheta(\nu_B, 1_B\mu)$$

and thus

$$2^{nd} \cdot (\mathfrak{c}_n - \widehat{\mathfrak{c}}_n) \le \mathbb{W}_\vartheta(\nu_B, \lambda_B).$$

If $Z > \alpha$ a transport $T_*\nu_B = \lambda_B$ can be constructed as follows: at each point of B the portion $\frac{\alpha}{Z}$ of ν_B remains where it is; the rest is transported from B into $\widehat{B} \setminus B$. The maximal transportation distance is $\sqrt{d} \cdot Z^{1/d}$. Hence, the cost can be estimated by

$$\vartheta\big(\sqrt{d} \cdot Z^{1/d}\big) \cdot (Z - \alpha).$$

On the other hand, if $Z < \alpha$ in a similar manner, a transport $T'_*\lambda_B = \nu_B$ can be constructed with cost bounded from above by

$$\vartheta\big(\sqrt{d} \cdot \alpha^{1/d}\big) \cdot (\alpha - Z).$$

Therefore, by definition of the function $\varepsilon(\cdot)$,

$$\begin{aligned}
\mathbb{W}_\vartheta(\nu_B, \lambda_B) &\le \mathbb{E}\big[\vartheta\big(\sqrt{d}(Z \vee \alpha)^{1/d}\big) \cdot |Z - \alpha|\big] \\
&\le \varepsilon\big(\alpha^{1/d}\big) \cdot \sqrt{d} \cdot \mathbb{E}\big[(Z \vee \alpha)^{1/2} \cdot |Z - \alpha|\big] \\
&\le \varepsilon\big(\alpha^{1/d}\big) \cdot \sqrt{d} \cdot \mathbb{E}[Z + \alpha]^{1/2} \cdot \mathbb{E}\big[|Z - \alpha|^2\big]^{1/2} \\
&= \varepsilon\big(2^n\big) \cdot \sqrt{d} \cdot \big[2 \cdot 2^{nd} \cdot 2^{nd}\big]^{1/2}.
\end{aligned}$$

This finally yields

$$\mathfrak{c}_n - \widehat{\mathfrak{c}}_n \le 2^{-nd} \cdot \mathbb{W}_\vartheta(\nu_B, \lambda_B) \le \varepsilon\big(2^n\big) \cdot \sqrt{2d}. \qquad \square$$

THEOREM 5.10. *Assume that*

$$\int_1^\infty \frac{\vartheta(r)}{r^{1+d/2}}\, dr < \infty \tag{9}$$

then

$$\mathfrak{c}_\infty \le \widehat{\mathfrak{c}}_\infty < \infty.$$

PROOF. Since

$$\int_1^\infty \frac{\vartheta(r)}{r^{1+d/2}}\, dr < \infty \quad \Longleftrightarrow \quad \sum_{n=1}^\infty \frac{\Theta(2^n)}{2^{n(1+d/2)}} < \infty,$$

Corollary 5.8 applies and yields $\widehat{\mathfrak{c}}_\infty < \infty$. Moreover, since ϑ is increasing, the integrability condition (9) implies that

$$\varepsilon(r) = \sup_{s \geq r} \frac{\vartheta(s)}{s^{d/2}} \to 0$$

as $r \to \infty$. Hence, $\mathfrak{c}_\infty \leq \widehat{\mathfrak{c}}_\infty$ by Proposition 5.9. □

The previous theorem essentially says that $\mathfrak{c}_\infty < \infty$ if ϑ grows "slightly" slower than $r^{d/2}$. This criterion is quite sharp in dimensions 1 and 2. Indeed, according to Theorem 5.2 in these two cases we also know that $\mathfrak{c}_\infty = \infty$ if ϑ grows like $r^{d/2}$ or faster.

5.3. *Estimates for L^p-cost.* The results of the previous section in particular apply to L^p-cost for $p < d/2$ in $d \leq 2$ and to L^p-cost for $p \leq 1$ in $d \geq 3$. A slight modification of these arguments will allow us to deduce cost estimates for L^p cost for arbitrary $p \geq 1$ in the case $d \geq 3$.

In this case, the finiteness of $\mathfrak{c}_\infty$ will also be covered by the more general results of [16]; see Theorem 1.3(i). However, using the idea of modified cost we get reasonably good quantitative estimates on $\mathfrak{c}_\infty$. Throughout this section we assume $\beta = 1$.

5.3.1. *Some moment estimates for Poisson random variables.* For $p \in \mathbb{R}$ let us denote by $\lceil p \rceil$ the smallest integer $\geq p$.

LEMMA 5.11. *For each $p \in (0, \infty)$ there exist constants $C_1(p), C_2(p)$ and $C_3(p)$ such that for every Poisson random variable Z with parameter $\alpha \geq 1$:*

(i) $\mathbb{E}[Z^p] \leq C_1(p) \cdot \alpha^p$, *where one can choose $C_1(1) = 1, C_1(2) = 4$. For general p one may choose $C_1(p) = \lceil p \rceil^p$ or $C_1(p) = 2^{p-1} \cdot (\lceil p \rceil - 1)!$.*
(ii) $\mathbb{E}[Z^{-p} \cdot 1_{\{Z>0\}}] \leq C_2(p) \cdot \alpha^{-p}$. *For general p one may choose $C_2(p) = (\lceil p \rceil + 1)!$.*
(iii) $\mathbb{E}[(Z - \alpha)^p] \leq C_3(p) \cdot \alpha^{p/2}$, *where one can choose $C_3(2) = 1, C_1(4) = 2$. For general p one may choose $C_3 = 2^{p-1} \cdot (2\lceil \frac{p}{2} \rceil - 1)!$.*

PROOF. In all cases, by Hölder's inequality it suffices to prove the claim for integer $p \in \mathbb{N}$.

(i) The moment generating function of Z is

$$M(t) := \mathbb{E}[e^{tZ}] = \exp(\alpha(e^t - 1)).$$

For integer p, the pth moment of Z is given by the pth derivative of M at the point $t = 0$, that is, $\mathbb{E}[Z^p] = M^{(p)}(0)$. As a function of α, the pth derivative of M is a polynomial of order p (with coefficients depending on t). As $\alpha \geq 1$ we are done.

To get quantitative estimates for C_1, observe that differentiating $M(t)$ p times yields at most 2^{p-1} terms, each of them having a coefficient $\leq (p-1)!$ (if we do not merge terms of the same order). Thus, we can take $C_1 = 2^{p-1} \cdot (p-1)!$.

Alternatively, we may use the recursive formula

$$T_{n+1}(\alpha) = \alpha \sum_{k=0}^{n} \binom{n}{k} T_k(\alpha)$$

for the *Touchard polynomials* $T_n(\alpha) := \mathbb{E}[Z^n]$; see, for example, [30]. Assuming that $T_k(\alpha) \leq (k\alpha)^k$ for all $k = 1, \ldots, n$ leads to the corresponding estimate for $k = n+1$.

(iii) Put $p = 2k$ with integer k. The moment generating function of $(Z - \alpha)$ is

$$\begin{aligned} N(t) &:= \exp\bigl(\alpha\bigl(e^t - 1 - t\bigr)\bigr) = \exp\left(\frac{\alpha}{2} t^2 h(t)\right) \\ &= 1 + \frac{\alpha}{2} t^2 h(t) + \frac{1}{2}\left(\frac{\alpha}{2}\right)^2 t^4 h^2(t) + \frac{1}{6}\left(\frac{\alpha}{2}\right)^3 t^6 h^3(t) + \cdots \end{aligned}$$

with $h(t) = \frac{2}{t^2}(e^t - 1 - t)$. Hence, the $2k$th derivative of N at the point $t = 0$ is a polynomial of order k in α. Since $\alpha \geq 1$ by assumption, $\mathbb{E}[(Z-\alpha)^{2k}] = N^{(2k)}(0) \leq C_3 \cdot \alpha^k$ for some C_3. To estimate C_3, again observe that differentiating $N(t)$ (2k) times yields at most 2^{2k-1} terms. Each of these terms has a coefficient $\leq (2k-1)!$ (if we do not merge terms). Hence we can take $C_3(2k) = 2^{2k-1} \cdot (2k-1)!$.

(ii) The result follows from the inequality

$$\frac{1}{x^k} \leq \frac{(k+1)!x!}{(k+x)!}$$

for positive integers k and x. The inequality is equivalent to

$$\binom{x+k}{x-1} \leq x^{k+1}.$$

For fixed k the latter inequality holds for $x = 1$. If x increases from x to $x+1$ the right-hand side grows by a factor of $(\frac{x+1}{x})^{k+1}$ and the left-hand side by a factor of $\frac{x+k+1}{x}$. As $(x+k+1)x^k \leq (x+1)^{k+1}$, the inequality holds. Then we can estimate

$$\begin{aligned} \mathbb{E}\left[\frac{1}{Z^k} \cdot 1_{Z>0}\right] &\leq \mathbb{E}\left[\frac{(k+1)!}{(Z+1)\cdots(Z+k)} \cdot 1_{Z>0}\right] \\ &= e^{-\alpha} \cdot \sum_{j=1}^{\infty} \frac{\alpha^j}{j!} \cdot \frac{(k+1)!}{(j+1)\cdots(j+k)} \\ &= \frac{(k+1)!}{\alpha^k} \cdot e^{-\alpha} \cdot \sum_{j=1}^{\infty} \frac{\alpha^{j+k}}{(j+k)!} \leq \frac{(k+1)!}{\alpha^k}. \end{aligned}$$

If we choose $k = \lceil p \rceil$, this yields the claim. $\square$

5.3.2. *L^p-cost for $p \geq 1$ in $d \geq 3$.* Given two measure valued random variables $\nu_1^\bullet, \nu_2^\bullet : \Omega \to \mathcal{M}(\mathbb{R}^d)$ with $\nu_1^\omega(\mathbb{R}^d) = \nu_2^\omega(\mathbb{R}^d)$ for a.e. $\omega \in \Omega$, we define their L^p-transportation distance by

$$\mathbb{W}_p(\nu_1, \nu_2) := \left[\int_\Omega W_p^p(\nu_1^\omega, \nu_2^\omega)\, d\mathbb{P}(\omega)\right]^{1/p},$$

where

$$W_p(\eta_1, \eta_2) = \inf\left\{\left[\int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^p \, d\theta(x,y)\right]^{1/p} : \theta \text{ is coupling of } \eta_1, \eta_2\right\}$$

denotes the usual L^p-Wasserstein distance between (not necessarily normalized) measures $\eta_1, \eta_2 \in \mathcal{M}(\mathbb{R}^d)$ of equal total mass. Note that $\mathbb{W}_p(\nu_1, \nu_2)$ is *not* the L^p-Wasserstein distance between the distributions of $\nu_1^\bullet$ and $\nu_2^\bullet$. The latter in general is smaller. Similar to the concave case the triangle inequality holds, and we define the *modified transportation cost* as

$$\begin{aligned}\widehat{\mathsf{C}}_A(\omega) &= \inf\left\{\int |x-y|^p \, d\widehat{q}(x,y) : \widehat{q} \text{ is coupling of } \nu_A^\omega \text{ and } 1_A\mu^\omega\right\}\\ &= W_p^p(\nu_A^\omega, 1_A\mu^\omega).\end{aligned}$$

Put

$$\widehat{\mathfrak{c}}_n = 2^{-nd} \cdot \mathbb{E}[\widehat{\mathsf{C}}_{B_n}] = \mathbb{W}_p^p(\nu_{B_n}^\bullet, 1_{B_n}\mu^\bullet)$$

with $B_n = [0, 2^n)^d$ as usual.

LEMMA 5.12. *Given $n \in \mathbb{N}_0, k \in \{1, \dots, d\}$ and $i \in \{0,1\}^k$ put $D_0 = B_{n+1}^{(i_1,\dots,i_{k-1},0)}$, $D_1 = B_{n+1}^{(i_1,\dots,i_{k-1},1)}$ and $D = D_0 \cup D_1 = B_{n+1}^{(i_1,\dots,i_{k-1})}$. Then for some constant κ_1 depending only on p,*

$$\mathbb{W}_p^p(\nu_{D_0} + \nu_{D_1}, \nu_D) \leq \kappa_1 \cdot 2^{(n+1)(p+d-pd/2)} \cdot 2^{k(p/2-1)+1}.$$

One may choose $\kappa_1(p) = \frac{1}{p+1} 2^{-p} \cdot C_3(2p) \cdot C_2(2(p-1))$.

PROOF. The proof will be a modification of the proof of Lemma 5.6. An optimal transport map $T : D \to D$ with $T_* \tilde{\nu}_D = \nu_D$ is now given by

$$T : (x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_d) \mapsto \left(x_1, \dots, x_{k-1}, \frac{2Z_0}{Z} \cdot x_k, x_{k+1}, \dots, x_d\right)$$

on D_0 and

$$\begin{aligned}&T : (x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_d)\\ &\qquad \mapsto \left(x_1, \dots, x_{k-1}, 2^{n+1} - (2^{n+1} - x_k) \cdot \frac{2Z_1}{Z}, x_{k+1}, \dots, x_d\right)\end{aligned}$$

on D_1. As before, we put $Z_j(\omega) = \mu^\omega(D_j)$ for $j = 0, 1$ and $Z = Z_0 + Z_1$. (If $p > 1$ this is indeed the only optimal transport map.) The cost of this transport can easily be calculated,

$$\int_{D_0} |T(x) - x|^p \, d\tilde{\nu}(x) = Z_0 \cdot 2^{-n} \int_0^{2^n} \left| \frac{2Z_0}{Z} \cdot x_k - x_k \right|^p dx_k$$
$$= \frac{2^{np}}{p+1} \cdot Z_0 \cdot \left| \frac{Z_0 - Z_1}{Z} \right|^p$$

and analogously

$$\int_{D_1} |T(x) - x|^p \, d\tilde{\nu}(x) = \frac{2^{np}}{p+1} \cdot Z_1 \cdot \left| \frac{Z_0 - Z_1}{Z} \right|^p.$$

Hence, together with the estimates from Lemma 5.11 this yields

$$\begin{aligned}
\mathbb{W}_p^p(\tilde{\nu}_D, \nu_D) &= \frac{2^{np}}{p+1} \cdot \mathbb{E}\left[\frac{|Z_0 - Z_1|^p}{Z^{p-1}} \cdot 1_{\{Z>0\}} \right] \\
&\le \frac{2^{np}}{p+1} \cdot \mathbb{E}\left[|Z_0 - Z_1|^{2p}\right]^{1/2} \cdot \mathbb{E}\left[Z^{-2(p-1)} \cdot 1_{\{Z>0\}}\right]^{1/2} \\
&\le \frac{2^{(n+1)p}}{p+1} \cdot \mathbb{E}\left[|Z_0 - \alpha_0|^{2p}\right]^{1/2} \cdot \mathbb{E}\left[Z^{-2(p-1)} \cdot 1_{\{Z>0\}}\right]^{1/2} \\
&\le \frac{2^{(n+1)p}}{p+1} \cdot C_3 \cdot \alpha_0^{p/2} \cdot C_2 \cdot \alpha^{1-p} \\
&\le \kappa_1 \cdot 2^{(n+1)(p+d-pd/2)} \cdot 2^{k(p/2-1)+1},
\end{aligned}$$

which is the claim. $\square$

With the very same proof as before (Proposition 5.7), by inserting different results, we get the following:

PROPOSITION 5.13. *For all $d \in \mathbb{N}$ and all $p \ge 1$, there is a constant $\kappa_2 = \kappa_2(p, d)$ such that for all $n \in \mathbb{N}_0$,*

$$\widehat{\mathfrak{c}}_{n+1}^{1/p} \le \widehat{\mathfrak{c}}_n^{1/p} + \kappa_2 \cdot 2^{(n+1)(1-d/2)}.$$

One may choose $\kappa_2(p, d) = \kappa_1(p)^{1/p} \cdot \sum_{k=1}^d 2^{k/2} \le \kappa_1(p)^{1/p} \cdot 2^{d/2+2}$, where κ_1 is the constant from the previous lemma.

COROLLARY 5.14. *For all $d \ge 3$ and all $p \ge 1$,*

$$\widehat{\mathfrak{c}}_\infty := \lim_{n \to \infty} \widehat{\mathfrak{c}}_n < \infty.$$

More precisely, for all $n \in \mathbb{N}_0$,

$$\widehat{\mathfrak{c}}_\infty^{1/p} \le \widehat{\mathfrak{c}}_n^{1/p} + \kappa_2 \cdot \frac{2^{-(n+1)(d/2-1)}}{1-2^{-(d/2-1)}}.$$

In particular,

$$\widehat{\mathfrak{c}}_\infty^{1/p} \le \widehat{\mathfrak{c}}_0^{1/p} + \frac{4\kappa_1(p)^{1/p}}{2^{-1}-2^{-d/2}}.$$

Recall the definition of $\mathfrak{c}_n$ from Section 2.7. Comparison of costs $\widehat{\mathfrak{c}}_n$ and $\mathfrak{c}_n$ now yields the following:

PROPOSITION 5.15. *For all* $d \ge 3$ *and all* $p \ge 1$, *there is a constant* κ_3 *such that for all* $n \in \mathbb{N}_0$,

$$\mathfrak{c}_n^{1/p} \le \widehat{\mathfrak{c}}_n^{1/p} + \kappa_3 \cdot 2^{n(1-d/2)}.$$

PROOF. It is a modification of the proof of Proposition 5.9. This time, the map $T : B \mapsto \widehat{B}$

$$T : x \mapsto \left(\frac{Z}{\alpha}\right)^{1/d} \cdot x$$

defines an optimal transport $T_*\nu_B = \lambda_B$. Put $\tau' = \tau'(d,p) = \int_{[0,1)^d} |x|^p\,dx$. (This can easily be estimated, e.g., by $\tau' \le \frac{1}{p+1} d^{p/2}$ if $p \ge 2$.) The cost of the transport T is

$$\begin{aligned}\int_B |T(x)-x|^p\,d\nu_B(x) &= \tau' \cdot 2^{np} \cdot Z \cdot \left|\left(\frac{Z}{\alpha}\right)^{1/d} - 1\right|^p \\ &\le \tau' \cdot 2^{np} \cdot Z \cdot \left|\frac{Z}{\alpha} - 1\right|^p.\end{aligned}$$

The inequality in the above estimation follows from the fact that $|t-1| \le |t-1| \cdot (t^{d-1} + \cdots + t + 1| = |t^d - 1|$ for each real $t > 0$. The previous cost estimates hold true for each fixed ω (which for simplicity we had suppressed in the notation). Integrating w.r.t. $d\mathbb{P}(\omega)$ yields

$$\begin{aligned}\mathbb{W}_p^p(\nu_B, \lambda_B) &\le \tau' \cdot 2^{np} \cdot \mathbb{E}\left[Z \cdot \left|\frac{Z}{\alpha} - 1\right|^p\right] \\ &\le \tau' \cdot 2^{np} \cdot \alpha^{-p} \cdot \mathbb{E}[Z^2]^{1/2} \cdot \mathbb{E}[|Z-\alpha|^{2p}]^{1/2} \\ &\le \tau' \cdot 2^{np} \cdot \alpha^{-p} \cdot \alpha \cdot C_3 \cdot \alpha^{p/2} = \kappa_3^p \cdot 2^{n(d+p-dp/2)}\end{aligned}$$

and thus

$$\mathfrak{c}_n^{1/p^*} - \widehat{\mathfrak{c}}_n^{1/p^*} \le \kappa_3 \cdot 2^{n(1-d/2)}.$$ □

COROLLARY 5.16. *For all* $d \ge 3$ *and all* $p \ge 1$,

$$\mathfrak{c}_\infty \le \widehat{\mathfrak{c}}_\infty < \infty.$$

5.3.3. *Quantitative estimates.* Throughout this section, we assume that $\vartheta(r) = r^p$ with $p < \overline{p}(d)$ where

$$p < \overline{p}(d) := \begin{cases} \infty, & \text{for } d \geq 3, \\ 1, & \text{for } d = 2, \\ \frac{1}{2}, & \text{for } d = 1. \end{cases}$$

PROPOSITION 5.17. *Put $\tau(p,d) = \frac{d}{d+p} \cdot (\Gamma(\frac{d}{2}+1)^{1/d} \cdot \pi^{-1/2})^p$. Then*

$$\mathfrak{c}_\infty \geq \mathfrak{c}_0 \geq \tau(p,d).$$

PROOF. The number τ as defined above is the minimal cost of a semicoupling between $\mathfrak{L}$ and a single Dirac mass, say δ_0. Indeed, this Dirac mass will be transported onto the d-dimensional ball $K_r = \{x \in \mathbb{R}^d : |x| < r\}$ of unit volume, that is, with radius r chosen s.t. $\mathfrak{L}(K_r) = 1$. The cost of this transport is $\int_{K_r} |x|^p \, dx = \frac{d}{d+p} r^p = \tau$.

For each integer $Z \geq 2$, the minimal cost of a semicoupling between $\mathfrak{L}$ and a sum of Z Dirac masses will be $\geq Z \cdot \tau$. Hence, if Z is Poisson distributed with parameter 1,

$$\mathfrak{c}_0 \geq \mathbb{E}[Z] \cdot \tau = \tau. \qquad \square$$

REMARK 5.18. Explicit calculations yield

$$\tau(p,1) = \frac{1}{1+p} \cdot 2^{-p}, \qquad \tau(p,2) = \frac{2}{2+p} \cdot \pi^{-p/2},$$
$$\tau(p,3) = \frac{3}{3+p} \cdot \left(\frac{3}{4\pi}\right)^{p/3}$$

whereas Stirling's formula yields a uniform lower bound, valid for all $d \in \mathbb{N}$ (which indeed is a quite good approximation for large d)

$$\tau(p,d) \geq \frac{d}{d+p} \cdot \left(\frac{d}{2\pi e}\right)^{p/2}.$$

PROPOSITION 5.19. *Put $\widehat{\tau} = \widehat{\tau}(d,p) = \int_{[0,1)^d} \int_{[0,1)^d} |x-y|^p \, dy \, dx$. Then*

$$e^{-1} \cdot \widehat{\tau} \leq \widehat{\mathfrak{c}}_0 \leq \widehat{\tau}.$$

Moreover, $\widehat{\tau} \leq \frac{1}{(1+p)(1+p/2)} \cdot d^{p/2}$ for all $p \geq 2$ and $\widehat{\tau} \leq (\frac{d}{6})^{p/2}$ for all $0 < p \leq 2$.

PROOF. If there is exactly one Poisson particle in $B_0 = [0,1)^d$—which then is uniformly distributed– then the transportation cost is exactly $\widehat{\tau}(d,p)$. If there are $N > 1$ particles in B_0, the cost per particle is by definition of $\widehat{\mathfrak{c}}_0$ bounded by $\widehat{\tau}(d,p)$. Hence, we can bound $\widehat{\mathfrak{c}}_0$ by the expected number of particles in B_0

times $\widehat{\tau}(d,p)$ which is precisely $\widehat{\tau}(d,p)$. The number of particles will be Poisson distributed with parameter 1. The lower estimate for the cost follows from the fact that with probability e^{-1} there is exactly one Poisson particle in $B_0=[0,1)^d$.

Using the inequality $(x_1^2+\cdots+x_d^2)^{p/2}\le d^{p/2-1}\cdot(x_1^p+\cdots+x_d^p)$—valid for all $p\ge 2$—the upper estimate for $\widehat{\tau}$ can be derived as follows:

$$\begin{aligned}\int_{[0,1)^d}\int_{[0,1)^d}|x-y|^p\,dy\,dx &\le d^{p/2-1}\sum_{i=1}^d\int_{[0,1]^d}\int_{[0,1]^d}|x_i-y_i|^p\,dy\,dx\\ &= d^{p/2}\int_0^1\int_0^1|s-t|^p\,ds\,dt\\ &= \frac{1}{(1+p)(1+p/2)}\cdot d^{p/2}.\end{aligned}$$

Applying Hölder's inequality to the inequality for $p=2$ yields the claim for all $p\le 2$. □

THEOREM 5.20. *For all $p\le 1$ and $d>2p$,*

$$\frac{d}{d+p}\cdot\left(\frac{d}{2\pi e}\right)^{p/2}\le\mathfrak{c}_\infty\le\left(\frac{d}{6}\right)^{p/2}+\frac{1}{(p+1)(2^{d/2-p}-1)}$$

whereas for all $p\ge 1$ and $d\ge 3$,

$$\left(\frac{d}{d+p}\right)^{1/p}\cdot\left(\frac{d}{2\pi e}\right)^{1/2}\le\mathfrak{c}_\infty^{1/p}\le\frac{d^{1/2}}{6^{1/2}\wedge[(1+p)(1+p/2)]^{1/p}}+28\cdot\kappa_1^{1/p}.$$

PROOF. Proposition 5.17 and the subsequent remark imply the lower bound

$$\frac{d}{d+p}\cdot\left(\frac{d}{2\pi e}\right)^{p/2}\le\tau\le\mathfrak{c}_\infty,$$

valid for all d and p. In the case $p\ge 1$ the upper bound follows from Proposition 5.19 and Corollary 5.14 by

$$\mathfrak{c}_\infty^{1/p}\le\widehat{\tau}^{1/p}+\frac{4\kappa_1^{1/p}}{2^{-1}-2^{-d/2}}\le\frac{d^{1/2}}{6^{1/2}\wedge[(1+p)(1+p/2)]^{1/p}}+28\cdot\kappa_1^{1/p}.$$

In the case $p\le 1$, estimate (8) with $\Theta(r)=\frac{1}{p+1}r^{p+1}$ yields

$$\widehat{\mathfrak{c}}_\infty\le\widehat{\mathfrak{c}}_0+\sum_{m=0}^\infty 2^{-(m+1)(d/2+1)}\cdot\frac{1}{p+1}2^{(m+1)(p+1)}=\widehat{\mathfrak{c}}_0+\frac{1}{(p+1)(2^{d/2-p}-1)},$$

provided $p<d/2$. Together with Proposition 5.9 this yields the claim. □

COROLLARY 5.21. (i) *For all* $p \in (0, \infty)$,

$$\frac{1}{\sqrt{2\pi e}} \le \liminf_{d\to\infty} \frac{\mathfrak{c}_\infty^{1/p}}{d^{1/2}} \le \limsup_{d\to\infty} \frac{\mathfrak{c}_\infty^{1/p}}{d^{1/2}} \le \frac{1}{\sqrt{6} \wedge [(1+p)(1+p/2)]^{1/p}}.$$

Note that the ratio of right and left-hand sides is less than 5, *and for* $p \le 2$ *even less than* 2.

(ii) *For all* $p \in (0, \infty)$ *there exist constants* k, k' *such that for all* $d > 2(p \wedge 1)$,

$$k \cdot d^{p/2} \le \mathfrak{c}_\infty \le k' \cdot d^{p/2}.$$

6. Optimal semicouplings with bounded second marginal. The goal of this chapter is to prove Theorem 2.1 (= Theorem 6.6), the crucial existence and uniqueness result for optimal semicouplings between the Lebesgue measure and the point process restricted to a bounded set.

Throughout this chapter, we fix the cost function $c(x, y) = \vartheta(|x - y|)$ with ϑ—as before—being a strictly increasing, continuous function from $\mathbb{R}_+$ to $\mathbb{R}_+$ with $\vartheta(0) = 0$ and $\lim_{r\to\infty} \vartheta(r) = \infty$. In dimension one we exclude the case $\vartheta(r) = r$.

LEMMA 6.1. *Suppose there is given a finite set* $\Xi = \{\xi_1, \dots, \xi_k\} \subset \mathbb{R}^d$ *and a probability density* $\rho \in L^1(\mathbb{R}^d, \mathfrak{L})$.

(i) *There exists a unique coupling* q *of* $\rho\mathfrak{L}$ *and* $\sigma = \frac{1}{k}\sum_{\xi\in\Xi} \delta_\xi$ *which minimizes the cost function* $\mathsf{Cost}(\cdot)$.

(ii) *There exists a* ($\mathfrak{L}$-*a.e. unique*) *map* $T : \{\rho > 0\} \to \Xi$ *with* $T_*(\rho\mathfrak{L}) = \sigma$ *which minimizes* $\int c(x, T(x))\rho(x)\, d\mathfrak{L}(x)$.

(iii) *There exists a* ($\mathfrak{L}$-*a.e. unique*) *map* $T : \{\rho > 0\} \to \Xi$ *with* $T_*(\rho\mathfrak{L}) = \sigma$ *which is* c-*monotone* (*in the sense that the closure of* $\{(x, T(x)) : \rho(x) > 0\}$ *is a* c-*cyclically monotone set*).

(iv) *The minimizers in* (i), (ii) *and* (iii) *are related by* $q = (\mathrm{id}, T)_*(\rho\mathfrak{L})$ *or, in other words*,

$$dq(x, y) = d\delta_{T(x)}(y)\rho(x)\, d\mathfrak{L}(x).$$

PROOF. We prove the lemma in three steps.

(a) By compactness of $\Pi(\rho\mathfrak{L}, \sigma)$ w.r.t. weak convergence and continuity of $c(\cdot, \cdot)$, there is a coupling q minimizing the cost function $\mathsf{Cost}(\cdot)$; see also [32], Theorem 4.1.

(b) Write $\rho\mathfrak{L} =: \lambda = \sum_{i=1}^k \lambda_i$ where $\lambda_i(\cdot) := q(\cdot \times \{\xi_i\})$ for each $i = 1, \dots, k$. We claim that the measures $(\lambda_i)_i$ are mutually singular. Assuming that there is a Borel set N such that for some $i \neq j$ we have $\lambda_i(N) = \alpha > 0$ and $\lambda_j(N) = \beta > 0$, we will redistribute the mass on N being transported to ξ_i and ξ_j in a cheaper way. This will show that the measures $(\lambda_i)_i$ are mutually singular. In particular, the proof implies the existence of a measurable c-monotone map T such that $q = (\mathrm{id}, T)_*(\rho\mathfrak{L})$.

We may assume w.l.o.g. that $(\rho\mathfrak{L})(N) = \alpha + \beta$. Otherwise write $\rho = \rho_1 + \rho_2$ such that on N $d\lambda_i(x) + d\lambda_j(x) = d(\rho_1\mathfrak{L})(x)$, and just work with the density ρ_1.

Put $f(x) := c(x, \xi_i) - c(x, \xi_j)$. As $c(\cdot, \cdot)$ is continuous, f is continuous. The function $c(x, y)$ is a strictly increasing function of the distance $|x - y|$. Thus, the level sets $\{f \equiv b\}$ define (locally) $(d - 1)$ dimensional submanifolds (e.g., use implicit function theorem for non smooth functions, see Corollary 10.52 in [32]) changing continuously with b. Choose b_0 such that $\rho\mathfrak{L}(\{f < b_0\} \cap N) = \alpha$ [which implies $\rho\mathfrak{L}(\{f > b_0\} \cap N) = \beta$] and set $N_i := \{f < b_0\} \cap N$ and $N_j := \{f \geq b_0\} \cap N$.

For $l = i, j$,

$$d\tilde{\lambda}_l(x) := d\lambda_l(x) - 1_N(x)\, d\lambda_l(x) + 1_{N_l}(x)\, d(\rho\mathfrak{L})(x).$$

For $l \neq i, j$ set $\tilde{\lambda}_l = \lambda_l$. By construction, $\tilde{q} = \sum_{l=1}^k \tilde{\lambda}_l \otimes \delta_{\xi_l}$ is a coupling of $\rho\mathfrak{L}$ and σ. Moreover, $\tilde{q}$ is c-cyclically monotone on N, that is, $\forall x_i \in N_i, x_j \in N_j$ we have

$$c(x_i, \xi_i) + c(x_j, \xi_j) \leq c(x_j, \xi_i) + c(x_i, \xi_j).$$

Furthermore, the set where equality holds is a null set because $c(x, y)$ is a strictly increasing function of the distance. Then we have

$$\begin{aligned} \mathsf{Cost}(q) - \mathsf{Cost}(\tilde{q}) = & \int_N c(x, \xi_i)\, d\lambda_i(x) + c(x, \xi_j)\, d\lambda_j(x) \\ & - \int_{N_i} c(x, \xi_i)\, d\tilde{\lambda}_i(x) - \int_{N_j} c(x, \xi_j)\, d\tilde{\lambda}_j(x) > 0, \end{aligned}$$

by cyclical monotonicity. This proves that λ_i and λ_j are singular to each other.

Hence, the family $(\lambda_i)_{i=1,\dots,k}$ is mutually singular which in turn implies that there exist Borel sets $S_i \subset \mathbb{R}^d$ with $\dot{\bigcup}_i S_i = \mathbb{R}^d$ and $\lambda_i(S_j) = 0$ for all $i \neq j$. Define the map $T : \mathbb{R}^d \to \Xi$ by $T(x) := \xi_i$ for all $x \in S_i$. Then $q = (\mathrm{id}, T)_*(\rho\mathfrak{L})$.

(c) Assume there are two minimizers of the cost function Cost, say q_1 and q_2. Then $q_3 := \frac{1}{2}(q_1 + q_2)$ is a minimizer as well. By step (b) we have $q_i = (\mathrm{id}, T_i)_*\rho\mathfrak{L}$ for $i = 1, 2, 3$. This implies

$$\begin{aligned} d\delta_{T_3(x)}(y)\, d\rho\mathfrak{L}(x) &= dq_3(x, y) = d\big(\tfrac{1}{2}q_1(x, y) + \tfrac{1}{2}q_2(x, y)\big) \\ &= d\big(\tfrac{1}{2}\delta_{T_1(x)}(y) + \tfrac{1}{2}\delta_{T_2(x)}(y)\big)\, d\rho\mathfrak{L}(x). \end{aligned}$$

This, however, implies $T_1(x) = T_2(x)$ for $\rho\mathfrak{L}$ a.e. $x \in \mathbb{R}^d$ and thus $q_1 = q_2$. □

REMARK 6.2. (1) In dimension one we exclude the case $c(x, y) = |x - y|$ because the optimal coupling between an absolutely continuous measure and a discrete measure need not be unique. In higher dimensions it is unique, as we get strict inequalities in the triangle inequalities. A counterexample for one dimension is the following. Take λ to be the Lebesgue measure on [0, 1] and put $\mu = \frac{1}{3}\delta_0 + \frac{2}{3}\delta_{1/16}$. Then, for any $a \in [1/16, 1/3]$,

$$q_a(dx, dy) = 1_{[0,a)}(x)\delta_0(dy)\lambda(dx) + 1_{[a,2/3+a)}(x)\delta_{1/16}(dy)\lambda(dx) + 1_{[a+2/3,1]}(x)\delta_0(dy)\lambda(dx)$$

is an optimal coupling of λ and μ with $\mathsf{Cost}(q_a) = 11/24$.

(2) In the case $\vartheta(r) = r^2$, there exists a convex function $\varphi : \{\rho > 0\} \to \mathbb{R}$ such that

$$T(x) = \nabla\varphi(x) \qquad \text{for } \mathfrak{L}\text{-a.e. } x.$$

More generally, if $\vartheta(r) = r^p$ with $p > 1$, then the map T is given as $T(x) = x + |\nabla\psi(x)|^{(2-p)/(p-1)} \cdot \nabla\psi(x)$ for some $|\cdot|^p$-convex function $\psi : \{\rho > 0\} \to \mathbb{R}$.

PROPOSITION 6.3. *For each finite set $\Xi \subset \mathbb{R}^d$ there exists a unique semicoupling q of $\mathfrak{L}$ and $\sigma = \sum_{\xi\in\Xi}\delta_\xi$ which minimizes the cost functional* $\mathsf{Cost}(\cdot)$.

PROOF. (i) The functional $\mathsf{Cost}(\cdot)$ on $\mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d)$ is lower semicontinuous w.r.t. weak topology. Indeed, if $\eta_n \to \eta$ weakly, then with $c_k(x, y) := \min\{\vartheta(|x - y|), k\}$

$$\liminf_n \mathsf{Cost}(\eta_n) \geq \sup_k \left[\lim_n \int c_k \, d\eta_n\right] = \sup_k \int c_k \, d\eta = \mathsf{Cost}(\eta).$$

(ii) Let $\mathfrak{Q}$ denote the set of all semicouplings of $\mathfrak{L}$ and σ and $\mathfrak{Q}_1$ the subset of those $q \in \mathfrak{Q}$ which satisfy $\frac{1}{2}\mathsf{Cost}(q) \leq \inf_{q'\in\mathfrak{Q}} \mathsf{Cost}(q') =: c$. Then $\mathfrak{Q}_1$ is relatively compact w.r.t. the weak topology. Indeed, $q(\mathbb{R}^d \times \complement\Xi) = 0$ for all $q \in \mathfrak{Q}_1$ and

$$q\big(\complement K_r(\Xi) \times \Xi\big) \leq \frac{1}{\vartheta(r)} \cdot \mathsf{Cost}(q) \leq \frac{2}{\vartheta(r)} c$$

for each $r > 0$ where $K_r(\Xi)$ denotes the closed r-neighborhood of Ξ in $\mathbb{R}^d$. Thus for any $\varepsilon > 0$ there exists a compact set $K = K_r(\Xi) \times \Xi$ in $\mathbb{R}^d \times \mathbb{R}^d$ such that $q(\complement K) \leq \varepsilon$ uniformly in $q \in \mathfrak{Q}_1$.

(iii) The set $\mathfrak{Q}$ is closed w.r.t. weak convergence. Indeed, if $q_n \to q$, then $(\pi_1)_*q_n \to (\pi_1)_*q$ and $(\pi_2)_*q_n \to (\pi_2)_*q$.

Thus, $\mathfrak{Q}_1$ is compact and $\mathsf{Cost}(\cdot)$ attains its minimum on $\mathfrak{Q}$ (or equivalently on $\mathfrak{Q}_1$).

(iv) Now let a minimizer q of $\mathsf{Cost}(\cdot)$ on $\mathfrak{Q}$ be given, and let $\lambda = (\pi_1)_*q$ denote its first marginal. Then $\lambda = \rho \cdot \mathfrak{L}$ for some density $0 \leq \rho \leq 1$ on $\mathbb{R}^d$. Our first claim will be that ρ *only attains values* 0 *and* 1.

Indeed, put $U = \{\rho > 0\}$. According to the previous Lemma 6.1, there exists an a.e. unique "transport map" $T : U \to \Xi$ s.t.

$$q = (\mathrm{id}, T)_*\lambda.$$

For a given "target point" $\xi \in \Xi$, $U_\xi := U \cap T^{-1}(\xi)$ is the set of points which under the map T will be transported to the point ξ. Within this set, the density ρ has values between 0 and 1 and its integral is 1. If the density is not already equal to 1 we can replace it by another one which gives maximal mass to the points which are closest to the target ξ. Indeed, put $r(\xi) := \inf\{r > 0 : \mathfrak{L}(K_r(\xi) \cap U_\xi) \geq 1\}$ and $\tilde{\lambda} := \tilde{\rho} \cdot \mathfrak{L}$ with

$$\tilde{\rho}(x) = 1_{\bigcup_{\xi \in \Xi} K_{r(\xi)}(\xi) \cap U_\xi}(x).$$

Then

$$\tilde{q} := (\mathrm{id}, T)_* \tilde{\lambda}$$

defines a semicoupling of $\mathfrak{L}$ and σ with $\mathsf{Cost}(\tilde{q}) \leq \mathsf{Cost}(q)$. Moreover, it holds that $\mathsf{Cost}(\tilde{q}) = \mathsf{Cost}(q)$ if and only if $\tilde{\rho} = \rho$ a.e. on $\mathbb{R}^d$. The latter is equivalent to $\rho \in \{0, 1\}$ a.e.

(v) Assume there are two optimal semicouplings q_1 and q_2 whose first marginals have density 1_{U_1} and 1_{U_2}, respectively. Then $q := \frac{1}{2}(q_1 + q_2)$ is optimal as well and its first marginal has density $\frac{1}{2}(1_{U_1} + 1_{U_2})$. By the previous part (iv) of this proof the density can attain only values 0 or 1. Therefore, we have $U_1 = U_2$ (up to measure zero sets) and $q_1 = q_2$. $\square$

LEMMA 6.4. *Given a bounded Borel set $A \subset \mathbb{R}^d$, let $\mathcal{M}_{\mathrm{count}}(A) = \{\sigma \in \mathcal{M}_{\mathrm{count}}(\mathbb{R}^d) : \sigma(\mathbb{R}^d \setminus A) = 0\}$ denote the set of finite counting measures which are concentrated on A. Define $\Upsilon : \mathcal{M}_{\mathrm{count}}(A) \to \mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d)$ the map which assigns to each $\sigma \in \mathcal{M}_{\mathrm{count}}(A)$ the unique $q \in \Pi_s(\mathfrak{L}, \sigma)$ which minimizes the cost functional* $\mathsf{Cost}(\cdot)$. *Then Υ is continuous (w.r.t. weak convergence on the respective spaces).*

PROOF. (i) Take a sequence $(\sigma_n)_n \subset \mathcal{M}_{\mathrm{count}}(A)$ converging weakly to some $\sigma \in \mathcal{M}_{\mathrm{count}}(A)$. Put $q_n := \Upsilon(\sigma_n)$ for $n \in \mathbb{N}$ and $q = \Upsilon(\sigma)$. We have to prove that $q_n \to q$.

(ii) The weak convergence $\sigma_n \to \sigma$ implies that finally all the measures σ_n have the same total mass as σ, say k. Hence, for each sufficiently large $n \in \mathbb{N}$ there exist points $x_1^n, \ldots, x_k^n$ and Borel sets $S_1^n, \ldots, S_k^n$ such that

$$\sigma_n = \sum_{i=1}^k \delta_{x_i^n}, \qquad q_n = \sum_{i=1}^k 1_{S_i^n} \mathfrak{L} \otimes \delta_{x_i^n}.$$

Similarly $\sigma = \sum_{i=1}^k \delta_{x_i}$ and $q = \sum_{i=1}^k 1_{S_i} \mathfrak{L} \otimes \delta_{x_i}$ with suitable points $x_1, \ldots, x_k$ and Borel sets $S_1, \ldots, S_k$. Weak convergence moreover implies that for each $i = 1, \ldots, k$,

$$x_i^n \to x_i \qquad \text{as } n \to \infty.$$

(iii) Based on the representations of q and σ_n, we can construct a semicoupling $\hat{q}_n$ of $\mathfrak{L}$ and σ_n as follows:

$$\hat{q}_n = \sum_{i=1}^{k} 1_{S_i} \mathfrak{L} \otimes \delta_{x_i^n}.$$

Then by continuity of ϑ and dominated convergence theorem,

$$\limsup_n \mathsf{Cost}(\hat{q}_n) = \limsup_n \sum_{i=1}^{k} \int_{S_i} \vartheta(|y - x_i^n|)\, dy$$
$$= \sum_{i=1}^{k} \int_{S_i} \vartheta(|y - x_i|)\, dy = \mathsf{Cost}(q).$$

And of course $\mathsf{Cost}(q_n) \le \mathsf{Cost}(\hat{q}_n)$. Thus

$$\limsup_n \mathsf{Cost}(q_n) \le \mathsf{Cost}(q).$$

(iv) The sequence $(q_n)_n$ is relatively compact in the weak topology of $\mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d)$. Therefore, there is a subsequence, denoted again by $(q_n)_n$, converging weakly to some measure $\tilde{q} \in \mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d)$. It follows that $(\pi_2)_* q_n \to (\pi_2)_* \tilde{q}$ and thus $(\pi_2)_* \tilde{q} = \sigma$. Similarly, $(\pi_1)_* \tilde{q} \le \mathfrak{L}$. Thus $\tilde{q} \in \Pi_s(\mathfrak{L}, \sigma)$. Lower semicontinuity of the cost functional implies

$$\mathsf{Cost}(\tilde{q}) \le \liminf_{n\to\infty} \mathsf{Cost}(q_n).$$

(v) Summarizing, we have proven that $\tilde{q}$ is a semicoupling of $\mathfrak{L}$ and σ with

$$\mathsf{Cost}(\tilde{q}) \le \mathsf{Cost}(q).$$

Since q is the unique minimizer of the cost functional among all these semicouplings, it follows that $\tilde{q} = q$. In other words,

$$\lim_{n\to\infty} \Upsilon(\sigma_n) = \Upsilon\Big(\lim_{n\to\infty} \sigma_n\Big).$$

This proves the continuity of Υ. □

For a given ω let us apply the previous results to the measure

$$\sigma = 1_A \mu^\omega = \sum_{\xi \in \Xi(\omega) \cap A} \delta_\xi$$

for a realization μ^ω of the point process. Then, there is a unique minimizer—in the sequel denoted by q_A^ω—of the cost functional Cost among all semicouplings of $\mathfrak{L}$ and $1_A \mu^\omega$.

LEMMA 6.5. *For each bounded Borel set $A \subset \mathbb{R}^d$ the map $\omega \to q_A^\omega$ is measurable.*

PROOF. We saw that the map $\Upsilon : \mathcal{M}_{\text{count}}(A) \to \mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d)$, $\sigma \mapsto \Upsilon(\sigma)$ assigning to each counting measure σ its unique minimizer of $\mathsf{Cost}(\cdot)$ is continuous. By definition of the point process, $\omega \mapsto \mu^\omega$ is measurable. Hence, the map

$$\omega \mapsto q_A^\omega = \Upsilon\Bigg(\sum_{\xi \in A \cap \Xi(\omega)} \delta_\xi\Bigg)$$

is measurable. □

THEOREM 6.6. (i) *For each bounded Borel set $A \subset \mathbb{R}^d$ there exists a unique semicoupling Q_A of $\mathfrak{L}$ and $(1_A\mu^\bullet)\mathbb{P}$ which minimizes the mean cost functional $\mathfrak{Cost}(\cdot)$.*

(ii) *Q_A can be disintegrated as $dQ_A(x,y,\omega) := dq_A^\omega(x,y)\,d\mathbb{P}(\omega)$ where for $\mathbb{P}$-a.e. ω the measure q_A^ω is the unique minimizer of the cost functional* $\mathsf{Cost}(\cdot)$ *among the semicouplings of $\mathfrak{L}$ and $1_A\mu^\omega$.*

(iii) $\mathfrak{Cost}(Q_A) = \int_\Omega \mathsf{Cost}(q_A^\omega)\,d\mathbb{P}(\omega)$.

PROOF. The existence of a minimizer is proven along the same lines as in the previous proposition: We choose an approximating sequence Q_n in $\mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d \times \Omega)$—instead of a sequence q_n in $\mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d)$—minimizing the lower semicontinuous functional $\mathfrak{Cost}(\cdot)$. Existence of a limit follows as before from tightness of the set of all semicouplings Q with $\mathfrak{Cost}(Q) \le 2 \inf_{\tilde{Q}} \mathfrak{Cost}(\tilde{Q})$.

For each semicoupling Q of $\mathfrak{L}$ and $\mu^\bullet\mathbb{P}$ with disintegration as $q^\bullet\mathbb{P}$, we obviously have

$$\mathfrak{Cost}(Q) = \int_\Omega \mathsf{Cost}(q^\omega)\,d\mathbb{P}(\omega).$$

Hence, Q is a minimizer of the functional $\mathfrak{Cost}(\cdot)$ (among all semicouplings of $\mathfrak{L}$ and $\mu^\bullet\mathbb{P}$) if and only if for $\mathbb{P}$-a.e. $\omega \in \Omega$ the measure q^ω is a minimizer of the functional $\mathsf{Cost}(\cdot)$ (among all semicouplings of $\mathfrak{L}$ and μ^ω).

Uniqueness of the minimizer of $\mathsf{Cost}(\cdot)$ therefore implies uniqueness of the minimizer of $\mathfrak{Cost}(\cdot)$. □

COROLLARY 6.7. *For each $z \in \mathbb{R}^d$ and each bounded Borel set $A \subset \mathbb{R}^d$, the measure Q_A satisfies*

$$Q_A(B, C, \omega) = Q_{A+z}(B+z, C+z, \omega+z)$$

for all Borel sets $B, C \in \mathcal{B}(\mathbb{R}^d)$.

PROOF. Since $\mathfrak{L}$ is equivariant and $\mu^\bullet$ is equivariant, the claim follows from the uniqueness of the minimizer of the cost functional $\mathfrak{Cost}(\cdot)$. □

REMARK 6.8. As before, for a finite set $\Xi \subset \mathbb{R}^d$ put $\sigma = \sum_{\xi \in \Xi} \delta_\xi$. Let q be a semicoupling of $\mathfrak{L}$ and σ. Then q minimizes $\mathsf{Cost}(\cdot)$ if and only if the support of q is c-cyclically monotone and q is *c-sequentially monotone* in the following sense:

$$\sum_{i=1}^{n} c(x_i, \xi_i) \le \sum_{i=1}^{n} c(x_{i+1}, \xi_i)$$

for all $n \in \mathbb{N}$, $\{(x_i, \xi_i)\}_{i=1}^n \in \operatorname{supp}(q)$, $\forall x_{n+1} \notin \operatorname{supp}((\pi_1)_* q)$.

PROOF. Let q be the unique minimizing semicoupling. The cyclical monotonicity follows from the general theory of optimal transportation; cf. Section 2.5. Put $U := \operatorname{supp}((\pi_1)_* q)$. Assume that q is not sequentially monotone. Then there are $n \in \mathbb{N}$, $x = x_{n+1} \in \complement U$, $\{(x_i, \xi_i)\}_{i=1}^n \in \operatorname{supp}(q)$ such that

$$\sum_{i=1}^{n} c(x_i, \xi_i) > \sum_{i=1}^{n} c(x_{i+1}, \xi_i).$$

By continuity of the cost function, there are (compact) neighborhoods U_i of x_i and V_i of ξ_i such that $U_{n+1} \cap U = \varnothing$ and

$$\sum_{i=1}^{n} c(u_i, v_i) > \sum_{i=1}^{n} c(u_{i+1}, v_i),$$

whenever $u_i \in U_i$ and $v_j \in V_j$. Moreover, as $\operatorname{supp}(\sigma)$ is discrete, we can assume (by shrinking V_j slightly if necessary) that $V_j \cap \operatorname{supp}(\sigma) = \{\xi_j\}$. As $(x_i, \xi_i) \in \operatorname{supp}(q)$ for $1 \le i \le n$, we have $\inf_i q(U_i \times \{\xi_i\}) > 0$. Set $\lambda := \inf\{q(U_1 \times \{\xi_1\}), \dots, q(U_n \times \{\xi_n\}), \mathfrak{L}(U_{n+1})\}$. Then we can reallocate mass to define a new measure with less cost. Indeed, we can choose subsets $\tilde{U}_i \subset U_i, \tilde{U}_i \times \{\xi\}_i \subset \operatorname{supp}(q)$ with $\mathfrak{L}(\tilde{U}_i) = \lambda$ and define a new measure $\tilde{q}$ by

$$d\tilde{q}(x, y) = dq(x, y) - \frac{1}{n} \sum_{i=1}^{n} 1_{\tilde{U}_i \times \{\xi_i\}}(x, y)\, d\mathfrak{L}(x) + \frac{1}{n} \sum_{i=1}^{n} 1_{\tilde{U}_{i+1} \times \{\xi_i\}}(x, y)\, d\mathfrak{L}(x).$$

By assumption, we have $\mathsf{Cost}(\tilde{q}) < \mathsf{Cost}(q)$. Hence, q is not minimizing Cost.

For the other direction let us assume that q is cyclically monotone and sequentially monotone but not minimizing $\mathsf{Cost}(\cdot)$. Then there is a Borel set $\tilde{U} \ne U (= \operatorname{supp}((\pi_1)_* q))$ (by uniqueness of optimal transportation of fixed measures) and a unique Cost minimizing coupling $\tilde{q}$ of $1_{\tilde{U}}\mathfrak{L}$ and σ such that $\mathsf{Cost}(\tilde{q}) \le \mathsf{Cost}(q)$, and the support of $\tilde{q}$ is cyclically monotone. As $\tilde{U} \ne U$ there is some $z \in \tilde{U} \setminus U$ which is transported by $\tilde{q}$ to ξ_0, say. For $\xi \in \Xi$ set $S_\xi := \{x \in \mathbb{R}^d : (x, \xi) \in \operatorname{supp}(q)\}$ and similarly $\tilde{S}_\xi$ for $\tilde{q}$. By sequential monotonicity of q for all $x_0 \in S_{\xi_0}$, we must have $c(x_0, \xi_0) \le c(z, \xi_0)$. Moreover, the set $\{x \in S_{\xi_0} : c(x, \xi_0) = c(z, \xi_0)\}$

is a $\mathfrak{L}$ null set. Thus there is a set $\hat{S}_{\xi_0} \subset S_{\xi_0}$ of Lebesgue measure one such that for all $x \in \hat{S}_{\xi_0}$, we have $c(x, \xi_0) < c(z, \xi_0)$. By the first part, we know that a minimizing semicoupling is sequentially monotone. Thus $\hat{S}_{\xi_0} \subset \tilde{U}$ and also $S_{\xi_0} \subset \tilde{U}$ (in particular if $\Xi = \{\xi_0\}$ we are done).

Moreover, by assumption there is some $x_1 \in S_{\xi_0} \setminus \tilde{S}_{\xi_0}$ which is transported by $\tilde{q}$ to some $\xi_1 \in \Xi$. Then $S_{\xi_1} \setminus \tilde{S}_{\xi_1}$ is not empty. If $S_{\xi_1} \cap \complement\tilde{U} \neq \varnothing$, we choose $x_2 \in S_{\xi_1} \cap \complement\tilde{U}$ and stop. If $S_{\xi_1} \subset \tilde{U}$, there is $x_2 \in S_{\xi_1} \setminus \tilde{S}_{\xi_1}$ which is transported by $\tilde{q}$ to some ξ_2. If $\xi_2 \in \{\xi_0, \xi_1\}$ (i.e., $\xi_2 = \xi_0$), we choose $x_2 \in \tilde{S}_{\xi_2} \cap S_{\xi_1}$ and stop. Otherwise we proceed in the same manner until either $S_{\xi_k} \cap \complement\tilde{U} \neq \varnothing$ or $\xi_k \in \{\xi_0, \dots, \xi_{k-2}\}$. By this procedure, we construct a sequence $x_0, \dots, x_k$ such that $x_j \in \tilde{S}_{\xi_j} \cap S_{\xi_{j-1}}$ for $1 \leq j \leq k-1$, $x_0 \in \tilde{S}_{\xi_0} \setminus U$, and either $x_k \in S_{\xi_k} \setminus \tilde{U}$ or $x_k \in \tilde{S}_{\xi_k} \cap S_{y_{k-1}} = \tilde{S}_{\xi_j} \cap S_{y_{k-1}}$ for some $0 \leq j \leq k-2$. In the latter case, we have by cyclical monotonicity for $\tilde{q}$ and q,

$$\sum_{i=j}^{k} c(x_i, \xi_i) \leq \sum_{i=j}^{k} c(x_{i+1}, \xi_i) \leq \sum_{i=j}^{k} c(x_i, \xi_i),$$

where $\xi_k = \xi_j$ and $x_{k+1} = x_j$. Hence we have equality everywhere. However, we can move the x_i slightly to get a contradiction. Thus, we need to have $x_k \in S_{\xi_k} \setminus \tilde{U}$. Then we have by the sequential monotonicity of $\tilde{q}$ and q

$$\sum_{i=0}^{k-1} c(x_i, \xi_i) \leq \sum_{i=0}^{k-1} c(x_{i+1}, \xi_i) \leq \sum_{i=0}^{k-1} c(x_i, \xi_i).$$

Hence we need to have equality and therefore a contradiction as before. Hence $\tilde{q} = q$. □

Acknowledgments. The first author would like to thank Alexander Holroyd for pointing out the challenges of $p < d/2$ in dimensions $d \leq 2$. Both authors would like to thank Matthias Erbar for the nice pictures.

REFERENCES

[1] AJTAI, M., KOMLÓS, J. and TUSNÁDY, G. (1984). On optimal matchings. *Combinatorica* **4** 259–264. MR0779885

[2] AMBROSIO, L. (2003). Lecture notes on optimal transport problems. In *Mathematical Aspects of Evolving Interfaces* (*Funchal*, 2000). *Lecture Notes in Math.* **1812** 1–52. Springer, Berlin. MR2011032

[3] AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed. Birkhäuser, Basel. MR2401600

[4] AMBROSIO, L., SAVARÉ, G. and ZAMBOTTI, L. (2009). Existence and stability for Fokker–Planck equations with log-concave reference measure. *Probab. Theory Related Fields* **145** 517–564. MR2529438

[5] AURENHAMMER, F. (1991). Voronoi diagrams—A survey of a fundamental geometric data structure. *ACM Computing Surveys* (*CSUR*) **23** 345–405.

[6] BEIGLBÖCK, M., GOLDSTERN, M., MARESCH, G. and SCHACHERMAYER, W. (2009). Optimal and better transport plans. *J. Funct. Anal.* **256** 1907–1927. MR2498564
[7] BRENIER, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44** 375–417. MR1100809
[8] CHATTERJEE, S., PELED, R., PERES, Y. and ROMIK, D. (2010). Phase transitions in gravitational allocation. *Geom. Funct. Anal.* **20** 870–917. MR2729280
[9] CHATTERJEE, S., PELED, R., PERES, Y. and ROMIK, D. (2010). Gravitational allocation to Poisson points. *Ann. of Math.* (2) **172** 617–671. MR2680428
[10] DUDLEY, R. M. (2002). *Real Analysis and Probability*. *Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358
[11] FIGALLI, A. (2010). The optimal partial transport problem. *Arch. Ration. Mech. Anal.* **195** 533–560. MR2592287
[12] GANGBO, W. and MCCANN, R. J. (1996). The geometry of optimal transportation. *Acta Math.* **177** 113–161. MR1440931
[13] HOFFMAN, C., HOLROYD, A. E. and PERES, Y. (2006). A stable marriage of Poisson and Lebesgue. *Ann. Probab.* **34** 1241–1272. MR2257646
[14] HOLROYD, A. E. and LIGGETT, T. M. (2001). How to find an extra head: Optimal random shifts of Bernoulli and Poisson random fields. *Ann. Probab.* **29** 1405–1425. MR1880225
[15] HOLROYD, A. E., PEMANTLE, R., PERES, Y. and SCHRAMM, O. (2009). Poisson matching. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** 266–287. MR2500239
[16] HOLROYD, A. E. and PERES, Y. (2005). Extra heads and invariant allocations. *Ann. Probab.* **33** 31–52. MR2118858
[17] LAST, G. and THORISSON, H. (2009). Invariant transports of stationary random measures and mass-stationarity. *Ann. Probab.* **37** 790–813. MR2510024
[18] LAUTENSACK, C. and ZUYEV, S. (2008). Random Laguerre tessellations. *Adv. in Appl. Probab.* **40** 630–650. MR2454026
[19] LOTT, J. and VILLANI, C. (2009). Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math.* (2) **169** 903–991. MR2480619
[20] MARKÓ, R. and TIMÁR, Á. (2011). A poisson allocation of optimal tail. Available at arXiv:1103.5259.
[21] OHTA, S.-I. (2009). Uniform convexity and smoothness, and their applications in Finsler geometry. *Math. Ann.* **343** 669–699. MR2480707
[22] OHTA, S.-I. and STURM, K.-T. (2009). Heat flow on Finsler manifolds. *Comm. Pure Appl. Math.* **62** 1386–1433. MR2547978
[23] OTTO, F. (2001). The geometry of dissipative evolution equations: The porous medium equation. *Comm. Partial Differential Equations* **26** 101–174. MR1842429
[24] OTTO, F. and VILLANI, C. (2000). Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173** 361–400. MR1760620
[25] RACHEV, S. T. and RÜSCHENDORF, L. (1998). *Mass Transportation Problems. Vol. I*. Springer, New York. MR1619170
[26] STURM, K.-T. (2006). On the geometry of metric measure spaces. I. *Acta Math.* **196** 65–131. MR2237206
[27] STURM, K.-T. (2006). On the geometry of metric measure spaces. II. *Acta Math.* **196** 133–177. MR2237207
[28] TALAGRAND, M. (1994). The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3. *Ann. Probab.* **22** 919–959. MR1288137
[29] TIMAR, A. (2009). Invariant matchings of exponential tail on coin flips in Z^d. Available at arXiv:0909.1090.
[30] TOUCHARD, J. (1956). Nombres exponentiels et nombres de Bernoulli. *Canad. J. Math.* **8** 305–320. MR0079021

[31] VILLANI, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483

[32] VILLANI, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454

UNIVERSITY OF BONN
ENDENICHER ALLEE 60
53115 BONN
GERMANY
E-MAIL: huesmann@iam.uni-bonn.de
sturm@iam.uni-bonn.de