Calculus of Variations

Optimal transport of vector measures

Krzysztof J. Ciosmak[1,2] (ORCID)

Abstract

We develop and study a theory of optimal transport for vector measures. We resolve in the negative a conjecture of Klartag, that given a vector measure on Euclidean space with total mass zero, the mass of any transport set is again zero. We provide a counterexample to the conjecture. We generalise the Kantorovich–Rubinstein duality to the vector measures setting. Employing the generalisation, we answer the conjecture in the affirmative provided there exists an optimal transport with absolutely continuous marginals of its total variation.

Mathematics Subject Classification Primary: 49K21 · 49Q20 · 46E30 · 90C25; Secondary: 28A50 · 60D05 · 46E40

1 Introduction

In this note we develop theory of optimal transport of vector measures. Let us first briefly describe the topic of classical optimal transport.

1.1 Optimal transport

In 1781 Gaspard Monge (see [19]) asked the following question: given two probability distributions μ, ν on a metric space (X, d), how to transfer one distribution onto the other in an optimal way. The criterion of optimality was to minimise the average transported distance.

✉ Krzysztof J. Ciosmak
ciosmak@maths.ox.ac.uk; krzysztof.ciosmak@sjc.ox.ac.uk

1   Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter (550), Woodstock Road, Oxford OX2 6GG, UK

2   St John's College, St John's College, University of Oxford, St Giles, Oxford, UK

Since then the topic has been developed extensively and much of this development has been done recently. We refer the reader to the books of Villani (see [25] and [26]) and to the lecture notes of Ambrosio (see [1]) for a thorough discussion, history and applications of the optimal transport problem.

The modern mathematical treatment of the problem has been initiated in 1942 by Kantorovich [14,15]. He proposed to consider a relaxed problem of optimising

$$\int_{X \times X} d(x, y) d\pi(x, y)$$

among all transference plans π between μ and ν, i.e., the set $\dot{}(\mu, \nu)$ of Borel probability measures on $X \times X$ with respective marginal distributions equal to μ and to ν. The existence of an optimal transference plan is a straightforward consequence of the Prokhorov's theorem, provided that X is separable.

The main question that has attracted a lot of attention is whether there exists an optimal transport plan, i.e., a Borel map $T : X \to X$ such that $T_\# \mu = \nu$ and the integral

$$\int_X d(x, T(x)) d\mu(x)$$

is minimal. If we knew that an optimal transference plan is concentrated on a graph of a Borel measurable function then we could infer the existence of an optimal transport plan. The first complete answer on Euclidean space, under regularity assumptions on the considered measures, was presented in a seminal paper [13] of Evans and Gangbo. However, before that, Sudakov in [23] presented a solution of the problem that contained a flaw. The flaw has been remedied by Ambrosio in [1] and later by Trudinger and Wang in [24] for the Euclidean distance and by Caffarelli, Feldman and McCann in [5] for distances induced by norms that satisfy certain smoothness and convexity assumptions. In [6] Caravenna has carried out the original strategy of Sudakov for general strictly convex norms and eventually Bianchini and Daneri in [4] accomplished the plan of a proof of Sudakov for general norms on finite-dimensional normed spaces.

Let us describe briefly the strategy of Sudakov in the context of Euclidean spaces. We assume that the two Borel probability measures μ, ν on $\mathbb{R}^n$ are absolutely continuous with respect to the Lebesgue measure.

Let us recall that the paramount Kantorovich–Rubinstein duality formula tells that

$$\sup \left\{ \int_{\mathbb{R}^n} u d(\mu - \nu) \mid u \text{ is 1-Lipschitz} \right\} \tag{1}$$

is equal to

$$\inf \left\{ \int_{\mathbb{R}^n \times \mathbb{R}^n} \|x - y\| d\pi(x, y) \mid \pi \in \dot{}(\mu, \nu) \right\}. \tag{2}$$

Let us take an optimal u and an optimal π in the two above optimisation problems. We may infer that

$$u(x) - u(y) = \|x - y\| \text{ for } \pi\text{-almost every } (x, y) \in X \times X. \tag{3}$$

Consider the maximal sets on which u is an isometry, called the *transport rays*. We see that all transport has to occur on these sets. Careful analysis of the Lipschitz function u shows that the transport rays form a foliation of the underlying space $\mathbb{R}^n$ into line segments, up to Lebesgue measure zero. Moreover, the so-called mass balance condition holds true. This is to say, for any Borel set A that is a union of some collection of transport rays there is

$\mu(A) = \nu(A)$; see e.g. [13]. Using the mass balance condition, we may construct an optimal transport by gluing together optimal maps for each of the transport rays; see e.g. [1].

This is one of the important observations that is employed in the localisation technique, which allows to reduce the dimension of a considered problem; see a paper of Klartag [17] for application of the technique to weighted Riemannian manifolds satisfying the curvature-dimension condition in the sense of Bakry and Émery [2,3] and papers of Cavalletti, Mondino [7,8] for application in the setting of metric measure spaces. The localisation technique stems from convex geometry, but its generalisations have been employed to prove many novel results concerning functional inequalities, e.g. isoperimetric inequality in the metric measure spaces satisfying the synthetic curvature-dimension condition (see [7,8]). The latter notion was introduced in the foundational papers by Sturm [21,22] and by Lott and Villani [18] and allowed for development of a far-reaching, vast theory of metric measure spaces. We refer the reader to [12] and references therein for a broader description of the localisation technique and its history.

1.2 Optimal transport of vector measures

The purpose of this article is to investigate multi-dimensional generalisation of the optimal transport problem and its connections with the localisation technique, as proposed by Klartag in [17, Chapter 6].

We shall consider finite-dimensional linear spaces equipped with Euclidean norm and 1-Lipschitz maps $u\colon \mathbb{R}^n \to \mathbb{R}^m$. A leaf $\mathcal{S}$ of a 1-Lipschitz map $u\colon \mathbb{R}^n \to \mathbb{R}^m$ is a maximal set, with respect to the order induced by inclusion, such that the restriction $u|_{\mathcal{S}}$ is an isometry. This is to say, $\mathcal{S}$ is a leaf, whenever for any $x, y \in \mathcal{S}$ there is

$$\|u(x) - u(y)\| = \|x - y\|$$

and for any $z \notin \mathcal{S}$ there exists $x \in \mathcal{S}$ such that

$$\|u(x) - u(z)\| < \|x - z\|.$$

The notion of leaves is a multi-dimensional generalisation of the notion of transport rays, see Sect. 1.1, of the one-dimensional optimal transport theory. We refer the reader to [12] for a thorough study of such leaves. Let us mention that such leaves form a convex partition of $\mathbb{R}^n$, up to Lebesgue measure zero. Moreover, any two such leaves may intersect only by their relative boundaries; see [12] for the proofs.

Suppose now that we are given a Borel probability measure μ on $\mathbb{R}^n$, absolutely continuous with respect to the Lebesgue measure, that satisfies m linear constrains. This is to say,

$$\int_{\mathbb{R}^n} f \, d\mu = 0 \tag{4}$$

for some integrable function $f\colon \mathbb{R}^n \to \mathbb{R}^m$ with finite first moments, i.e.,

$$\int_{\mathbb{R}^n} \|f(x)\| \|x\| \, d\mu(x) < \infty.$$

Let $u\colon \mathbb{R}^n \to \mathbb{R}^m$ be a 1-Lipschitz map such that

$$\int_{\mathbb{R}^n} \langle u, f \rangle \, d\mu = \sup \left\{ \int_{\mathbb{R}^n} \langle v, f \rangle \, d\mu \mid v\colon \mathbb{R}^n \to \mathbb{R}^m \text{ is 1-Lipschitz} \right\}. \tag{5}$$

Existence of u follows by the Arzelà–Ascoli theorem.

A Borel subset A of $\mathbb{R}^n$ shall be called a *transport set* associated to u, whenever for any $x \in A$ that belongs to a unique leaf of u and any $y \in \mathbb{R}^n$ such that

$$\|x - y\| = \|u(x) - u(y)\|,$$

there is $y \in A$. In other words, a transport set is a Borel union of a collection of leaves of u. In [17, Chapter 6] it is conjectured that for any transport set A of u

$$\int_A f \, d\mu = 0. \tag{6}$$

This is a generalisation of the mass balance condition, mentioned in Sect. 1.1. The affirmative answer to the conjecture would imply that one may decompose any Borel probability measure μ, satisfying m linear constraints of the form (4), into a mixture of measures, concentrated on pairwise disjoint convex subsets of $\mathbb{R}^n$ of dimension at most m, satisfying the same linear constraints; see also [12] for a discussion of the decomposition.

If $m = 1$ then (5) is precisely the dual problem to the optimal transport problem for measures ρ_1, ρ_2 given by formulae $d\rho_1 = f_+ d\mu$ and $d\rho_2 = f_- d\mu$. As we see in (1), the dual problem, depends merely on the difference of measures, and therefore, it makes sense to consider the optimal transport for signed measures with total mass zero.

Inspired by this observation, in Sect. 2 we develop a theory of optimal transport with metric cost of vector measures of total mass zero and study its basic properties. The rôle of a vector measure in the problem considered above is played by the measure with density f with respect to the measure μ.

The precise formulation of the optimal transport problem for an $\mathbb{R}^m$-valued measure η on a metric space (X, d) that we deal with is as follows:

$$\inf \left\{ \int_{X \times X} d(x, y) d\|\pi\|(x, y) \mid P_1\pi - P_2\pi = \eta \right\}. \tag{7}$$

Here $P_1\pi$ and $P_2\pi$ stand for the first and the second marginal of the $\mathbb{R}^m$-valued measure π respectively. The assumption on η is that

$$\int_X d(x, x_0) d\|\mu\|(x) < \infty \text{ for some } x_0 \in X \text{ and } \eta(X) = 0.$$

The above problem for $m = 1$ simplifies to the original optimal transport problem, as follows readily by the Kantorovich–Rubinstein formula. We prove that for $m > 1$ an analogue of this formula holds with (1) replaced by

$$\sup \left\{ \int_X \langle u, d\eta \rangle \mid u \colon X \to \mathbb{R}^m \text{ is 1-Lipschitz} \right\} \tag{8}$$

and with (2) replaced by (7). This is a content of Theorem 2. We also develop a theory of the Wasserstein space $\mathcal{W}(X, \mathbb{R}^m)$ of vector-valued measures. We identify its dual space as the space of vector-valued Lipschitz maps; see Theorem 1. Theorem 3 provides an analogue of (3) in the new setting.

The conjecture of Klartag (see [17, Chapter 6]) in the language of our theory of optimal transport of vector measures may be restated as follows. Suppose that we are given a vector measure μ on $\mathbb{R}^n$, with $\mu(\mathbb{R}^n) = 0$, which is absolutely continuous with respect to Lebesgue measure. Let $u \colon \mathbb{R}^n \to \mathbb{R}^m$ be a 1-Lipschitz map, with respect to Euclidean norms, that attains the supremum

$$\sup \left\{ \int_{\mathbb{R}^n} \langle v, d\mu \rangle \mid v \colon \mathbb{R}^n \to \mathbb{R}^m \text{ is 1-Lipschitz} \right\}. \tag{9}$$

It is claimed in [17] that the following mass balance condition holds true

$$\mu(A) = 0 \text{ for any Borel set } A \text{ that is a union of a family of leaves of } u. \tag{10}$$

Using the developed theory, in Sect. 3 we resolve the conjecture in the affirmative, provided that there exists an optimal transport with marginals of its total variation that are absolutely continuous with respect to the Lebesgue measure; see Theorem 4. Note that in the one-dimensional setting, the existence of such optimal transport is clear; see (1) and (2).

We provide a counterexample to the conjecture, for the case $m > 1$; see Theorem 5. It shows that, in general, the mass balance condition (10) fails to be true. It follows that it may happen that an optimal transport with absolutely continuous marginals do not exist, unlike in the one-dimensional case.

More generally, let $\mathcal{F}$ be any subset of 1-Lipschitz maps that is locally uniformly closed. We prove that the mass balance condition (10) fails to be true, even when the variational problem (9) is replaced by

$$\sup\left\{ \int_{\mathbb{R}^n} \langle v, d\mu \rangle \mid v \in \mathcal{F} \right\}, \tag{11}$$

unless $\mathcal{F}$ is trivial, i.e. consists merely of affine maps. This is shown for also any norm on $\mathbb{R}^n$ and any strictly convex norm on $\mathbb{R}^m$; see Theorem 6.

Note that the outline of a proof of the conjecture suggested in [17] has a gap, as follows by the results of [11].

Let us mention here that in [12] the generalisation of the localisation technique to multiple constraints is studied. In there, a partition associated to any 1-Lipschitz map $u \colon \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$, is studied thoroughly. It is established that any log-concave measure on $\mathbb{R}^n$ may be disintegrated with respect to this partition and that the resulting conditional measures, associated to leaves of maximal dimension, are again log-concave. This result is also presented in the context of spaces satisfying the curvature-dimension condition $CD(\kappa, N)$, thus partially confirming another conjecture of Klartag [17, Chapter 6].

Let us also mention the existence of another approach to optimal transport of vector measures that differs from ours developed by Chen, Georgiou, Tannenbaum, Tyu, Li, Osher, Haber, Yamamoto (see [9,10] and [20]).

1.3 Outline of the paper

Section 2 is devoted to development and study of the optimal transport theory of vector measures. We define a Wasserstein space and in Theorem 1 we identify its dual. Theorem 2 provides an analogue of the Kantorovich–Rubinstein duality formula.

In Sect. 3 we study the mass balance condition for vector measures. In Theorem 4 we answer in the affirmative the a conjecture of Klartag, provided there exists an optimal transport with absolutely continuous marginals of its total variation. In Theorem 5 we provide a counterexample to the conjecture, in the Euclidean setting. In Theorem 6 we resolve the conjecture in the negative in the general setting.

2 Optimal transport of vector measures

In this section we develop the theory of optimal transport of vector measures.

Let X be a metric space with metric d. Let μ be $\mathbb{R}^m$-valued Borel measure on X. If π is a $\mathbb{R}^m$-valued Borel measure on $X \times X$, we write $\mathrm{P}_1\pi$ for the first *marginal* of π, i.e. the

measure given by

$$P_1\pi(A) = \pi(A \times X),$$

for all Borel $A \subset X$, and $P_2\pi$ for the second *marginal* of π,

$$P_2\pi(B) = \pi(X \times B),$$

for all Borel $B \subset X$. We shall consider a variational problem

$$\mathcal{I}(\mu) = \inf\left\{\int_{X \times X} d(x,y) d\|\pi\|(x,y) \mid \pi \in \Gamma(\mu)\right\}. \tag{12}$$

Here $\Gamma(\mu)$ is the set of all $\mathbb{R}^m$-valued Borel measures π on $X \times X$ such that

$$\mu = P_1\pi - P_2\pi.$$

To check whether (12) defines a meaningful quantity, we have to check whether $\Gamma(\mu)$ is non-empty.

We shall need the following definition.

Definition 1 Let σ be an $\mathbb{R}^m$-valued Borel measure on X and let θ be a Borel signed measure on X. A unique Borel $\mathbb{R}^m$-valued measure $\sigma \otimes \theta$ such that

$$\langle \sigma \otimes \theta, v \rangle = \langle \sigma, v \rangle \otimes \theta$$

for all $v \in \mathbb{R}^m$ we shall call a *product measure*. Here $\langle \sigma, v \rangle \otimes \theta$ is the usual product measure of $\mathbb{R}$-valued measures.

Remark 1 It is clear that the product measure exists. Analogously we define the product measure $\theta \otimes \sigma$ for a Borel signed measure σ and a Borel $\mathbb{R}^m$-valued measure θ.

Proposition 1 $\Gamma(\mu)$ *is non-empty if and only if*

$$\mu(X) = 0. \tag{13}$$

Proof Clearly, if there exists $\pi \in \Gamma(\mu)$, then

$$\mu(X) = P_1\pi(X) - P_2\pi(X) = \pi(X \times X) - \pi(X \times X) = 0,$$

so the condition (13) is satisfied. Conversely, assume that (13) holds true. Let ν be any Borel probability measure on X. Set

$$\pi = \mu \otimes \nu.$$

Here $\mu \otimes \nu$ is the product measure; see Definition 1. Then for any Borel set $A \subset X$ and any vector $v \in \mathbb{R}^m$, we have

$$\langle \pi(A \times X) - \pi(X \times A), v \rangle = \langle \mu(A), v \rangle - \langle \mu(X), v \rangle \nu(A) = \langle \mu(A), v \rangle.$$

This is to say, $P_1\pi - P_2\pi = \mu$. $\qquad\square$

The quantity defined by (12) we shall call the Kantorovich–Rubinstein norm of μ.

Proposition 2 *Assume that $\mu(X) = 0$. Then $\mathcal{I}(\mu) < \infty$ provided that*

$$\int_{\mathbb{R}^n} d(x, x_0) d\|\mu\|(x) < \infty \tag{14}$$

for some (equivalently: any) $x_0 \in X$.

Proof Define

$$\pi = \mu \otimes \delta_{x_0}.$$

Here δ_{x_0} is a probability measure such that $\delta_{x_0}(\{x_0\}) = 1$. Then $\pi \in \Gamma(\mu)$ and

$$\int_{X \times X} d(x, y) d\|\pi\|(x, y) \leq \int_X d(x, x_0) d\|\mu\|(x). \tag{15}$$

This shows that $\mathcal{I}(\mu) < \infty$, provided that (14) is satisfied. The equivalence of finiteness of

$$\int_{\mathbb{R}^n} d(x, y) d\|\mu\|(x) < \infty$$

for any $y \in X$ follows by the triangle inequality. $\qquad\square$

Definition 2 We define the *Wasserstein space* $\mathcal{W}(X, \mathbb{R}^m)$ of all Borel measures μ on X with values in $\mathbb{R}^m$ such that

$$\mu(X) = 0 \text{ and } \int_X d(x, x_0) d\|\mu\|(x) < \infty$$

for some $x_0 \in X$. We endow it with a norm $\|\mu\|_{\mathcal{W}(X, \mathbb{R}^m)} = \mathcal{I}(\mu)$.

Before we proceed let us recall the following definition.

We say that a non-negative Borel measure μ on X is *inner regular* if for any Borel set $B \subset X$ we have

$$\mu(B) = \sup\{\mu(K) \mid K \subset B, K \text{ is a compact set}\}.$$

Let us note that Ulam's lemma tells that any finite Borel measure on a Polish space is inner regular.

Lemma 1 *Suppose that X is a Polish space. Let μ be a $\mathbb{R}^m$-valued Borel measure in $\mathcal{W}(X, \mathbb{R}^m)$. Suppose that for any Lipschitz function $u \colon X \to \mathbb{R}^m$*

$$\int_X \langle u, d\mu \rangle = 0.$$

Then $\mu = 0$.

Proof We may assume that $m = 1$. Let $\mu = \mu_+ - \mu_-$ be the Hahn–Jordan decomposition of μ. There exists two disjoint, Borel sets $A, B \subset X$ with $\mu_+(A^c) = 0$ and $\mu_-(B^c) = 0$. Choose any Borel set $E \subset A$. As any finite measure on X is inner regular, for any $\epsilon > 0$, there exists a compact set $K \subset E$ such that

$$\mu_+(E) \leq \mu_+(K) + \epsilon.$$

Define a function u_ϵ by the formula

$$u_\epsilon(x) = (1 - \frac{1}{\epsilon} \mathrm{dist}(x, K)) \vee 0.$$

Then u_ϵ is Lipschitz, equal to one on K and equal to zero on the complement of

$$K_\epsilon = \{x \in X \mid \mathrm{dist}(x, K) \leq \epsilon\}.$$

Thus

$$0 = \int_X u_\epsilon d\mu = \mu_+(K) + \int_{K_\epsilon \setminus K} u_\epsilon d\mu,$$

Therefore, by the above,

$$\mu_+(E) \le \epsilon + \mu_+(K) \le \epsilon + \mu_-\left(K_\epsilon \setminus K\right).$$

Letting ϵ tend to zero, we get $\mu_+(E) = 0$. It follows that $\mu_+ = 0$. Analogously, $\mu_- = 0$. This is to say, $\mu = 0$. □

Remark 2 In what follows, we shall always assume that underlying space X is a Polish space.

Proposition 3 *The function* $\mathcal{W}(X, \mathbb{R}^m) \ni \mu \mapsto \|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)} \in \mathbb{R}$ *is a norm.*

Proof Let us first check that

$$\|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)} = 0 \text{ if and only if } \mu = 0. \tag{16}$$

If $\mu = 0$, then $\pi = 0$ belongs to $\Gamma(\mu)$, so $\|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)} = 0$. Conversely, assume that $\|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)} = 0$. Choose any L-Lipschitz function

$$u\colon X \to \mathbb{R}^m.$$

Then for any $\pi \in \Gamma(\mu)$ we have

$$\left| \int_X \langle u, d\mu \rangle \right| = \left| \int_{X \times X} \langle u(x) - u(y), d\pi(x, y) \rangle \right| \le L \int_{X \times X} d(x, y) d\|\pi\|(x, y).$$

Therefore if $\|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)} = 0$, then

$$\int_X \langle u, d\mu \rangle = 0.$$

It follows by Lemma 1, that $\mu = 0$. Homogeneity of $\|\cdot\|_{\mathcal{W}(X,\mathbb{R}^m)}$ is clear. Let us show that the triangle inequality holds. For this choose measures $\mu, \nu \in \mathcal{W}(X, \mathbb{R}^m)$ and any measures $\pi \in \Gamma(\mu)$ and $\rho \in \Gamma(\nu)$. Then

$$\mu + \nu = \mathrm{P}_1(\pi + \rho) - \mathrm{P}_2(\pi + \rho),$$

so that $\pi + \rho \in \Gamma(\mu + \nu)$. It follows that

$$\|\mu + \nu\|_{\mathcal{W}(X,\mathbb{R}^m)} \le \int_{X \times X} d(x, y) d\|\pi + \rho\|(x, y) \le$$

$$\le \int_{X \times X} d(x, y) d\|\pi\|(x, y) + \int_{X \times X} d(x, y) d\|\rho\|(x, y).$$

Taking infimum over all π, ρ we see that the triangle inequality holds true. □

Proposition 4 *The linear space* $\mathcal{U}$ *of measures of the form*

$$\sum_{i=1}^{n} \delta_{x_i} v_i$$

for $x_i \in X$ *and* $v_i \in \mathbb{R}^m$, $i = 1, \ldots, n$, *such that* $\sum_{i=1}^{n} v_i = 0$, *is dense in* $\mathcal{W}(X, \mathbb{R}^m)$.

Proof Choose any measure $\mu \in \mathcal{W}(X, \mathbb{R}^m)$. Choose any $\epsilon > 0$. Choose any point $x_0 \in X$ and a compact set K such that

$$\int_{K^c} d(x, x_0) d\|\mu\|(x) \le \epsilon.$$

Choose pairwise disjoint Borel sets $A_1, A_2, \ldots, A_k \subset K$ such that the diameter of each is at most ϵ and

$$K = \bigcup_{i=1}^{k} A_i.$$

Consider the restrictions $\mu_i = \mu|_{A_i}$ of the measure μ to the sets $A_i, i = 1, 2, \ldots, k$. Choose any points $x_i \in A_i$. Then, as

$$\pi_i = \mu_i \otimes \delta_{x_i} \in \Gamma\left(\mu_i - \mu_i(X)\delta_{x_i}\right),$$

we have

$$\|\mu_i - \mu_i(X)\delta_{x_i}\|_{\mathcal{W}(X, \mathbb{R}^m)} \leq \int_X d(y, x_i) \, d\|\mu_i\|(y) \leq \epsilon \|\mu\|(A_i).$$

Let $A_0 = K^c$ and let $\mu_0 = \mu|_{A_0}$. Then

$$\pi_0 = \mu_0 \otimes \delta_{x_0} \in \Gamma\left(\mu_0 - \mu_0(X)\delta_{x_0}\right),$$

so

$$\|\mu_0 - \mu_0(X)\delta_{x_0}\|_{\mathcal{W}(X, \mathbb{R}^m)} \leq \int_X d(x, x_0) \, d\|\mu_0\|(x) \leq \epsilon.$$

Set

$$\nu = \sum_{i=0}^{k} \mu(A_i)\, \delta_{x_i}.$$

Then $\nu \in \mathcal{U}$. By the triangle inequality

$$\|\mu - \nu\|_{\mathcal{W}(X, \mathbb{R}^m)} \leq \sum_{i=0}^{k} \|\mu_i - \mu_i(X)\delta_{x_i}\|_{\mathcal{W}(X, \mathbb{R}^m)} \leq$$

$$\leq \epsilon \sum_{i=1}^{k} \|\mu(A_i)\| + \epsilon \leq \epsilon \left(\|\mu\|(X) + 1\right).$$

This concludes the proof. $\square$

Corollary 1 *If X is separable, then so is the Wasserstein space $\mathcal{W}(X, \mathbb{R}^m)$.*

Proof Fix $n \in \mathbb{N}$. Choose a countable dense subset $A \subset X$ and a set

$$B \subset \left\{ (w_1, \ldots, w_n) \in \mathbb{R}^m \times \ldots \mathbb{R}^m \mid \sum_{i=1}^{n} w_i = 0 \right\} \tag{17}$$

which is countable and dense in the set on the right-hand side of (17). Consider a measure μ given by

$$\mu = \sum_{i=1}^{n} \delta_{x_i} v_i$$

for $x_i \in X$ and $v_i \in \mathbb{R}^m, i = 1, \ldots, n$, such that $\sum_{i=1}^{n} v_i = 0$. Choose $\epsilon > 0$ and $\tilde{x}_i \in A$, $i = 1, \ldots, n$, and $(\tilde{v}_i)_{i=1}^{n} \in B$, such that for $i = 1, \ldots, n$

$$d(x_i, \tilde{x}_i) < \epsilon \text{ and } \|v_i - \tilde{v}_i\| < \epsilon \text{ and } \sum_{i=1}^{n} \tilde{v}_i = 0.$$

Set

$$\tilde{\mu} = \sum_{i=1}^{n} \delta_{\tilde{x}_i} \tilde{v}_i.$$

Then

$$\|\mu - \tilde{\mu}\|_{\mathcal{W}(X, \mathbb{R}^m)} \leq \Big\| \sum_{i=1}^{n} \delta_{x_i} (v_i - \tilde{v}_i) \Big\|_{\mathcal{W}(X, \mathbb{R}^m)} + \Big\| \sum_{i=1}^{n} \left(\delta_{x_i} - \delta_{\tilde{x}_i} \right) v_i \Big\|_{\mathcal{W}(X, \mathbb{R}^m)}$$

Choose any $x_0 \in X$. Taking

$$\pi = \sum_{i=1}^{n} \delta_{x_i} \otimes \delta_{x_0} (v_i - \tilde{v}_i) \text{ and } \rho = \sum_{i=1}^{n} \left(\delta_{x_i} \otimes \delta_{\tilde{x}_i} \right) v_i$$

we see that

$$\Big\| \sum_{i=1}^{n} \delta_{x_i} (v_i - \tilde{v}_i) \Big\|_{\mathcal{W}(X, \mathbb{R}^m)} \leq \epsilon \sum_{i=1}^{n} d(x_i, x_0)$$

and

$$\Big\| \sum_{i=1}^{n} \left(\delta_{x_i} - \delta_{\tilde{x}_i} \right) v_i \Big\|_{\mathcal{W}(X, \mathbb{R}^m)} \leq \epsilon \sum_{i=1}^{n} \|v_i\|.$$

The conclusion follows now from Proposition 4.   □

Definition 3   Choose any $x_0 \in X$. Define

$$\mathcal{L}\left(X, \mathbb{R}^m\right) = \left\{ u \colon X \to \mathbb{R}^m \mid u \text{ is Lipschitz and } u(x_0) = 0 \right\},$$

i.e. the Banach space of $\mathbb{R}^m$-valued Lipschitz functions on X taking value zero at x_0, with norm

$$\|u\|_{\mathcal{L}(X, \mathbb{R}^m)} = \sup \left\{ \frac{\|u(x) - u(y)\|}{d(x, y)} \mid x, y \in X, x \neq y \right\}.$$

Theorem 1   *Define*

$$T \colon \mathcal{L}\left(X, \mathbb{R}^m\right) \to \mathcal{W}\left(X, \mathbb{R}^m\right)^*$$

and

$$S \colon \mathcal{W}\left(X, \mathbb{R}^m\right)^* \to \mathcal{L}\left(X, \mathbb{R}^m\right)$$

by

$$T(u)(\mu) = \int_X \langle u, d\mu \rangle \tag{18}$$

and

$$\langle S(\lambda)(x), w \rangle = \lambda \left(\left(\delta_x - \delta_{x_0} \right) w \right), \tag{19}$$

for any $w \in \mathbb{R}^m$. Then S, T are mutual reciprocals and establish an isometric isomorphism of $\mathcal{L}(X, \mathbb{R}^m)$ and $\mathcal{W}(X, \mathbb{R}^m)^*$.

Proof   Choose any $\pi \in \Gamma(\mu)$. Then $P_1\pi - P_2\pi = \mu$. Thus, if u is a Lipschitz map, then

$$\left| \int_X \langle u, d\mu \rangle \right| = \left| \int_X \langle u(x) - u(y), d\pi(x, y) \rangle \right| \leq \|u\|_{\mathcal{L}(X, \mathbb{R}^m)} \int_X d(x, y) d\|\pi\|(x, y).$$

Taking infimum over all $\pi \in \Gamma(\mu)$, we see that

$$\left| \int_X \langle u, d\mu \rangle \right| \le \|u\|_{\mathcal{L}(X,\mathbb{R}^m)} \|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)}.$$

The above calculation shows that the formula (18) defines a continuous functional of norm at most $\|u\|_{\mathcal{L}(X,\mathbb{R}^m)}$. If $w \in \mathbb{R}^m$ if of norm one and $x, y \in X$, $x \ne y$, then for

$$\mu_{x,y,w} = \frac{\delta_x - \delta_y}{d(x, y)} w \tag{20}$$

we have $\|\mu_{x,y,w}\|_{\mathcal{W}(X,\mathbb{R}^m)} \le 1$ and for any $u \in \mathcal{L}(X, \mathbb{R}^m)$

$$\int_{\mathbb{R}^n} \langle u, d\mu_{x,y,w} \rangle = \frac{\langle w, u(x) - u(y) \rangle}{d(x, y)}.$$

Thus

$$\|u\|_{\mathcal{L}(X,\mathbb{R}^m)} = \|T(u)\|.$$

We shall now show that $T \circ S = \text{Id}$. Take any functional $\lambda \in \mathcal{W}(X, \mathbb{R}^m)^*$. Set

$$\sigma_{x,w} = (\delta_x - \delta_{x_0}) w.$$

Then $S(\lambda) \colon X \to \mathbb{R}^m$ is defined by the formula

$$\langle S(\lambda)(x), w \rangle = \lambda(\sigma_{x,w}).$$

It is clear that the above formula defines $S(\lambda)$ uniquely. Then we claim that map $v = S(\lambda)$ is $\|\lambda\|$-Lipschitz. Indeed

$$\|v(x) - v(y)\| = \sup \left\{ \langle v(x) - v(y), w \rangle \mid w \in \mathbb{R}^m, \|w\| = 1 \right\},$$

and as

$$\langle v(x) - v(y), w \rangle = \lambda(\sigma_{x,w} - \sigma_{y,w}) \le \|\lambda\| \|\sigma_{x,w} - \sigma_{y,w}\|_{\mathcal{W}(X,\mathbb{R}^m)}$$

we see that

$$\|v(x) - v(y)\| \le \|\lambda\| d(x, y), \text{ since } \|\sigma_{x,w} - \sigma_{y,w}\|_{\mathcal{W}(X,\mathbb{R}^m)} \le d(x, y).$$

Suppose that $\nu = (\delta_x - \delta_y)z$. We compute

$$T(v)(\nu) = \int_X \langle v, d\nu \rangle = \int_X \langle v, z \rangle d(\delta_x - \delta_y) = \lambda(\sigma_{x,z} - \sigma_{y,z}) = \lambda(\nu).$$

We see that $T(S(\lambda))$ and λ are equal on the set spanned by $(\delta_x - \delta_y)z$, where $x, y \in X$, $z \in \mathbb{R}^m$. By Proposition 4, we see that $T(S(\lambda))$ and λ are equal on $\mathcal{W}(X, \mathbb{R}^m)$.

Let us show also that $S \circ T = \text{Id}$. Choose any $w \in \mathbb{R}^m$ and any map $u \in \mathcal{L}(X, \mathbb{R}^m)$. Then

$$\langle S(T(u))(x), w \rangle = T(u)((\delta_x - \delta_{x_0}) w) = \int_X \langle u, d(\delta_x - \delta_{x_0}) w \rangle = \langle u(x), w \rangle,$$

as $u(x_0) = 0$. Therefore $S(T(u)) = u$. $\qquad\square$

Theorem 2 *For any* $\mu \in \mathcal{W}(X, \mathbb{R}^m)$

$$\sup \left\{ \int_X \langle u, d\mu \rangle \mid u \colon X \to \mathbb{R}^m \text{ is 1-Lipschitz} \right\} = \|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)}. \tag{21}$$

Moreover, there exists 1-*Lipschitz function* u_0 *such that*

$$\sup\left\{\int_X \langle u, d\mu\rangle \mid u\colon X \to \mathbb{R}^m \text{ is } 1\text{-Lipschitz}\right\} = \int_X \langle u_0, d\mu\rangle. \tag{22}$$

Proof Notice first that the left-hand side of (21) is clearly at most the right-hand side of (21). Take any $\mu \in \mathcal{W}(X, \mathbb{R}^m)$. Then by the Hahn–Banach theorem there exists a continuous linear functional λ of norm one such that

$$\lambda(\mu) = \|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)}.$$

By Theorem 1, we know that λ is of the form

$$\lambda(\mu) = \int_X \langle u_0, d\mu\rangle$$

for some Lipschitz map u_0. The Lipschitz constant of u_0 is equal to one, as

$$\|u_0\|_{\mathcal{L}(X,\mathbb{R}^m)} = \|\lambda\| = 1.$$

This completes the proof. $\square$

Definition 4 Any 1-Lipschitz function $u\colon X \to \mathbb{R}^m$ such that (22) holds we shall call an *optimal potential* of measure μ.

Definition 5 A measure $\pi \in \Gamma(\mu)$ such that

$$\|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)} = \int_{X \times X} d(x, y) d\|\pi\|(x, y)$$

we shall call an *optimal transport* for μ.

Theorem 3 *Let* $\mu \in \mathcal{W}(X, \mathbb{R}^m)$. *Let* $u \in \mathcal{L}(X, \mathbb{R}^m)$ *be a* 1-*Lipschitz map. Let* $\pi \in \Gamma(\mu)$. *The following conditions are equivalent:*

(i)
$$\int_X \langle u, d\mu\rangle = \int_{X \times X} d(x, y) d\|\pi\|(x, y) = \|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)},$$

(ii)
$$\int_A \langle u(x) - u(y), d\pi(x, y)\rangle = \int_A d(x, y) d\|\pi\|(x, y)$$

for any Borel set $A \subset X \times X$,

(iii)
$$\int_X \langle u, d\mu\rangle = \int_{X \times X} d(x, y) d\|\pi\|(x, y),$$

(iv) u *is an optimal potential for* μ *and* π *is an optimal transport for* μ.

Moreover, if the above conditions hold, then

$$\|u(x) - u(y)\| = d(x, y)$$

$\|\pi\|$-*almost everywhere.*

Proof Assume that (iii) holds. Observe that

$$\int_X \langle u, d\mu \rangle = \int_{X \times X} \langle u(x) - u(y), d\pi(x, y) \rangle.$$

As

$$\int_X \langle u, d\mu \rangle \leq \|\mu\|_{\mathcal{W}(X, \mathbb{R}^m)} \leq \int_{X \times X} d(x, y) d\|\pi\|(x, y),$$

then by (iii) we see that in the above inequalities we have equalities. This is to say, (i) holds true.

Suppose now that (i) holds. Clearly

$$\int_A \langle u(x) - u(y), d\pi(x, y) \rangle \leq \int_A d(x, y) d\|\pi\|(x, y).$$

If we had strict inequality in (ii) for some Borel set $A \subset X \times X$, then the above computations show that we would get strict inequality in (i). Condition (iv) is reformulation of (i). The last part of the theorem follows readily from (ii). □

We say that a measure $\mu \in \mathcal{M}(Z, \mathbb{R}^m)$ is *concentrated* on a subset $X \subset Z$ if there is $\|\mu\|(Z \setminus X) = 0$.

Proposition 5 *Assume that* $\mathbb{R}^n, \mathbb{R}^m$ *are equipped with Euclidean norms. Let* $\mu \in \mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$ *be concentrated on a set* $X \subset \mathbb{R}^n$. *Then*

$$\|\mu\|_{\mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)} = \|\mu\|_{\mathcal{W}(X, \mathbb{R}^m)}.$$

Proof The assertion is that

$$\sup \left\{ \int_{\mathbb{R}^n} \langle u, d\mu \rangle \mid u \colon \mathbb{R}^n \to \mathbb{R}^m \text{ is 1-Lipschitz} \right\}$$

is equal to

$$\sup \left\{ \int_X \langle u, d\mu \rangle \mid u \colon X \to \mathbb{R}^m \text{ is 1-Lipschitz} \right\}.$$

By the Kirszbraun theorem (see e.g. [16]) any 1-Lipschitz function $u \colon X \to \mathbb{R}^m$ extends to a 1-Lipschitz function $\tilde{u} \colon \mathbb{R}^n \to \mathbb{R}^m$. Clearly, for any such extension

$$\int_{\mathbb{R}^n} \langle \tilde{u}, d\mu \rangle = \int_X \langle u, d\mu \rangle.$$

The assertion follows. □

3 Mass balance condition

Let us first provide an affirmative answer to the conjecture of Klartag, under the provision of the existence of optimal transport with absolutely continuous marginals of its total variation.

Definition 6 A leaf $\mathcal{S}$ of a 1-Lipschitz map $u \colon \mathbb{R}^n \to \mathbb{R}^m$ is a maximal set, with respect to the order induced by inclusion, such that the restriction $u|_{\mathcal{S}}$ is an isometry. This is to say, $\mathcal{S}$ is a leaf, whenever for any $x, y \in \mathcal{S}$ there is

$$\|u(x) - u(y)\| = \|x - y\|$$

and for any $z \notin S$ there exists $x \in S$ such that

$$\|u(x) - u(z)\| < \|x - z\|.$$

It is proven in [12] that leaves of a map u that is 1-Lipschitz with respect to Euclidean norms are closed and convex sets. Two distinct leaves may intersect at most by their relative boundaries.

Definition 7 Let $u \colon \mathbb{R}^n \to \mathbb{R}^m$ be a 1-Lipschitz map of Euclidean spaces. We say that a Borel set $A \subset \mathbb{R}^n$ is a *transport set* associated with u if it enjoys the following property: if $x \in A$ is contained in a unique leaf of u and $y \in \mathbb{R}^n$ is such that

$$\|u(x) - u(y)\| = \|x - y\|,$$

then $y \in A$.

Let us remark that a Borel set $A \subset \mathbb{R}^n$ that is a union of leaves of u is a transport set.

We shall denote by $B(u)$ the set of all points $x \in \mathbb{R}^n$ such that there exist at least two distinct leaves S_1, S_2 of u such that $x \in S_1 \cap S_2$. In [12, Corollary 2.15] it is proven that $B(u)$ is of Lebesgue measure zero.

Suppose that $\mu \in \mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$. The following theorem shows that if there exists an optimal transport for μ such that its total variation has absolutely continuous marginals, then the conjecture of Klartag holds true. Note that such existence is clear for $m = 1$, whenever μ is absolutely continuous with respect to the Lebesgue measure λ.

Theorem 4 *Assume that $\mathbb{R}^n, \mathbb{R}^m$ are equipped with Euclidean norms. Suppose that $\mu \in \mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$. Let u be an optimal potential for μ. Suppose that there exists an optimal transport π of μ such that*

$$P_1 \|\pi\| \ll, P_2 \|\pi\| \ll \lambda. \tag{23}$$

Then for any transport set A associated with u:

(i) $\mu(A) = 0$,
(ii) $\pi|_{A \times A} \in \Gamma(\mu|_A)$ *is an optimal transport of $\mu|_A$*
(iii) u *is an optimal potential of $\mu|_A$.*

Proof By [12, Corollary 2.15] it follows that

$$\lambda(B(u)) = 0.$$

Suppose that (23) holds true. Then

$$\|\pi\|\big(B(u) \times \mathbb{R}^n\big) = 0 \text{ and } \|\pi\|\big(\mathbb{R}^n \times B(u)\big) = 0.$$

Let

$$I = \big\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mid \|u(x) - u(y)\| = \|x - y\|\big\}.$$

By Theorem 3, $\|\pi\|(I^c) = 0$. Thus π is concentrated on the set

$$C = I \cap \big(B(u)^c \times B(u)^c\big).$$

Suppose that $(x, y) \in C$. Then, as A is a transport set, by the definition of $B(u)$,

$$x \in A \text{ if and only if } y \in A. \tag{24}$$

Let $\eta = \pi|_{A \times A}$. To prove (ii), it is enough to show that η is an optimal transport and that

$$\eta \in \Gamma(\mu|_A).$$

For this, let $D \subset \mathbb{R}^n$ be any Borel set. Using the fact that $\pi \in \Gamma(\mu)$ and the fact that $\|\pi\|(C^c) = 0$ and (24), we have

$$
\begin{aligned}
\mu(A \cap D) &= \int_{\mathbb{R}^n \times \mathbb{R}^n} \left(\mathbf{1}_{A \cap D}(x) - \mathbf{1}_{A \cap D}(y) \right) d\pi(x, y) \\
&= \int_{\mathbb{R}^n \times \mathbb{R}^n} \mathbf{1}_{A \times A}(x, y) \left(\mathbf{1}_D(x) - \mathbf{1}_D(y) \right) d\pi(x, y) \\
&= \int_{\mathbb{R}^n \times \mathbb{R}^n} \left(\mathbf{1}_D(x) - \mathbf{1}_D(y) \right) d\eta(x, y) = \mathrm{P}_1 \eta(D) - \mathrm{P}_2 \eta(D).
\end{aligned}
$$

It follows that $\pi|_{A \times A} \in \Gamma(\mu|_A)$. Then

$$
\int_A \langle u, d\mu \rangle = \int_{\mathbb{R}^n \times \mathbb{R}^n} \mathbf{1}_C(x, y) \Big\langle \mathbf{1}_A(x) u(x) - \mathbf{1}_A(y) u(y), d\pi(x, y) \Big\rangle. \tag{25}
$$

Therefore, by (24),

$$
\int_A \langle u, d\mu \rangle = \int_{\mathbb{R}^n \times \mathbb{R}^n} \mathbf{1}_{A \times A}(x, y) \Big\langle u(x) - u(y), d\pi(x, y) \Big\rangle.
$$

By condition (ii) of Theorem 3 we see that

$$
\int_A \langle u, d\mu \rangle = \int_{A \times A} \|x - y\| d\|\pi\|(x, y).
$$

Theorem 3, condition (iii), tells us that $\pi|_{A \times A}$ is an optimal transport and u is an optimal potential. Also $\mu(A) = 0$, as $\pi|_{A \times A} \in \Gamma(\mu|_A)$. This completes the proof. $\qquad \square$

We shall now provide necessary tools for the aforementioned counterexample to the conjecture of Klartag.

In fact we shall provide a more general theorem for which we shall consider locally uniformly closed subsets subsets $\mathcal{F}$ of 1-Lipschitz maps of $\mathbb{R}^n$ to $\mathbb{R}^m$ endowed with norms which are not necessarily Euclidean. Suppose that a measure μ belongs to $\mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$. We consider supremum of integrals

$$
\int_{\mathbb{R}^n} \langle u, d\mu \rangle \tag{26}
$$

taken over all $u \in \mathcal{F}$. An optimal $u_0 \in \mathcal{F}$, i.e. the map that satisfies

$$
\int_{\mathbb{R}^n} \langle u_0, d\mu \rangle = \sup \left\{ \int_{\mathbb{R}^n} \langle u, d\mu \rangle \mid u \in \mathcal{F} \right\},
$$

we shall call an $\mathcal{F}$-optimal potential of μ.

Lemma 2 *Let $X \subset \mathbb{R}^n$ be a compact set. Suppose that $(\mu_k)_{k=1}^{\infty} \subset \mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$ are all supported on X and converge weakly* to $\mu_0 \in \mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$, i.e. for any continuous and bounded function $g \colon \mathbb{R}^n \to \mathbb{R}^m$ we have*

$$
\lim_{k \to \infty} \int_{\mathbb{R}^n} \langle g, d\mu_k \rangle = \int_{\mathbb{R}^n} \langle g, d\mu_0 \rangle.
$$

Suppose that for $k = 1, 2, \ldots$, $u_k \in \mathcal{F}$ is an $\mathcal{F}$-optimal potential of μ_k and that u_k converge locally uniformly to $u_0 \colon \mathbb{R}^n \to \mathbb{R}^m$. Then u_0 is an $\mathcal{F}$-optimal potential of μ_0.

Proof By the assumption, for any continuous and bounded map $g \colon \mathbb{R}^n \to \mathbb{R}^m$, we have

$$\lim_{k \to \infty} \int_{\mathbb{R}^n} \langle g, d(\mu_k - \mu_0) \rangle = 0.$$

In particular, as μ_k are all supported on X, we have

$$\lim_{k \to \infty} \int_{\mathbb{R}^n} \langle u_0, d(\mu_k - \mu_0) \rangle = 0.$$

By the Banach–Steinhaus theorem, the sequence $(\mu_k)_{k=1}^{\infty}$ is bounded in the total variation norm. Hence, by uniform convergence on X,

$$\lim_{k \to \infty} \int_{\mathbb{R}^n} \langle u_k - u_0, d\mu_k \rangle = 0.$$

It follows that

$$\int_{\mathbb{R}^n} \langle u_k, d\mu_k \rangle = \int_{\mathbb{R}^n} \langle u_0, d\mu_k \rangle + \int_{\mathbb{R}^n} \langle u_k - u_0, d\mu_k \rangle$$

converges to $\int_{\mathbb{R}^n} \langle u_0, d\mu_0 \rangle$. As for any 1-Lipschitz map $h \in \mathcal{F}$ we have

$$\int_{\mathbb{R}^n} \langle h, d\mu_k \rangle \le \int_{\mathbb{R}^n} \langle u_k, d\mu_k \rangle.$$

we also have

$$\int_{\mathbb{R}^n} \langle h, d\mu_0 \rangle \le \int_{\mathbb{R}^n} \langle u_0, d\mu_0 \rangle.$$

The proof is complete. □

Below we shall denote by $B(x, \epsilon)$ an open ball of radius $\epsilon > 0$ centred at $x \in \mathbb{R}^n$.

Lemma 3 *Let $m \le n$. Let $\mu \in \mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$ and let u be an optimal potential of μ. Let A be the union of all leaves of dimension at least one. Then A is Borel measurable. Suppose that there exists an optimal transport π for μ or that any transport set of u is of μ-measure zero. Then*

$$\|\mu\|(A^c) = 0.$$

Proof Observe that

$$A = \bigcup_{n=1}^{\infty} \left\{ x \in \mathbb{R}^n \mid \sup \left\{ \frac{\|u(x) - u(y)\|}{\|x - y\|} \mid y \in \mathrm{cl}B(x, n) \setminus B(x, 1/n) \right\} = 1 \right\}.$$

The function

$$\mathbb{R}^n \ni x \mapsto \sup \left\{ \frac{\|u(x) - u(y)\|}{\|x - y\|} \mid y \in \mathrm{cl}B(x, n) \setminus B(x, 1/n) \right\} \in \mathbb{R}$$

is lower semi-continuous, hence Borel measurable. Thus, A is Borel measurable. Suppose that there exists an optimal transport π for μ. By Theorem 3, π is supported on the set

$$I = \left\{ (x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mid \|u(x) - u(y)\| = \|x - y\| \right\}.$$

As $\mu = \mathrm{P}_1 \pi - \mathrm{P}_2 \pi$, for any Borel set $B \subset A^c$, we have

$$\mu(B) = \pi\left(B \times \mathbb{R}^n\right) - \pi\left(\mathbb{R}^n \times B\right) = 0,$$

for if $B \subset A^c$, then

$$\left(B \times \mathbb{R}^n\right) \cap I \subset \left\{(x, x) \mid x \in \mathbb{R}^n\right\} \text{ and } \left(\mathbb{R}^n \times B\right) \cap I \subset \left\{(x, x) \mid x \in \mathbb{R}^n\right\}.$$

Suppose now that any transport set for u is of μ measure zero. Observe that any Borel set $B \subset A^c$ is a transport set. The conclusion follows. □

In the theorem below we shall provide a counterexample to the conjecture of Klartag.

Theorem 5 *Assume that $m > 1$. There exists an absolutely continuous measure $\mu \in \mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$ for which there exists a transport set associated with an optimal potential of μ with non-zero measure μ.*

In particular, there is no optimal transport π for μ such that

$$P_1 \|\pi\| \ll \lambda \text{ and } P_2 \|\pi\| \ll \lambda.$$

Proof Choose any $v_1, \ldots, v_{m+1} \in \mathbb{R}^m$ such that

$$\sum_{i=1}^{m+1} v_i = 0$$

and that are affinely independent. For $\epsilon > 0$ set

$$\mu_\epsilon = \frac{1}{\lambda(B(0, \epsilon))} \sum_{i=1}^{m+1} \lambda|_{B(x_i, \epsilon)} v_i,$$

where $x_1, \ldots, x_{m+1} \in \mathbb{R}^n$ are pairwise distinct points to be specified later. Here λ denotes the Lebesgue measure on $\mathbb{R}^n$. Then $\mu_\epsilon \in \mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$. Suppose that for some sequence $(\epsilon_k)_{k=1}^{\infty}$ converging to zero there is

$$\mu_{\epsilon_k}(C_k) = 0$$

for any transport set C_k of u_k, where $u_k \colon \mathbb{R}^n \to \mathbb{R}^m$ is an optimal potential of μ_{ϵ_k}. For $k \in \mathbb{N}$ and $i = 1, \ldots, m+1$ consider the union N_{ik} of all non-trivial leaves of u_k – i.e. of dimension at least one – that intersect $\mathrm{cl} B(x_i, \epsilon_k)$. Then N_{ik} is a transport set. Its Borel measurability follows by Lemma 3. Indeed, denote $B = \mathrm{cl} B(x_i, \epsilon_k)$; then the function

$$\mathbb{R}^n \setminus B \ni x \mapsto \sup \left\{ \frac{\|u_k(x) - u_k(y)\|}{\|x - y\|} \mid y \in B \right\} \in \mathbb{R}$$

is lower semi-continuous and therefore

$$N_{ik} = \left\{ x \in \mathbb{R}^n \setminus B \mid \sup \left\{ \frac{\|u_k(x) - u_k(y)\|}{\|x - y\|} \mid y \in B \right\} = 1 \right\} \cup (B \cap A_k)$$

is a Borel set. Here A_k is a set of all leaves of dimension at least one corresponding to u_k, c.f. Lemma 3. Thus $\mu_{\epsilon_k}(N_{ik}) = 0$. Hence,

$$\sum_{j=1}^{m+1} v_j \lambda \left(B\left(x_j, \epsilon_k\right) \cap N_{ik} \right) = 0. \tag{27}$$

As μ_{ϵ_k}, by Lemma 3, is concentrated on non-trivial leaves of u_k, we have for

$$\frac{\lambda \left(B\left(x_i, \epsilon_k\right) \cap N_{ik}\right)}{\lambda \left(B\left(0, \epsilon_k\right)\right)} v_i = \mu_{\epsilon_k} \left(B\left(x_i, \epsilon_k\right) \cap N_{ik}\right) = \mu_{\epsilon_k} \left(B\left(x_i, \epsilon_k\right)\right) = v_i.$$

By (27) and assumption on the vectors $v_1, \ldots, v_{m+1}$

$$\lambda \left(B \left(x_j, \epsilon_k \right) \cap N_{ik} \right) = \lambda \left(B \left(0, \epsilon_k \right) \right) \text{ for all } j = 1, \ldots, m+1.$$

Thus we infer that for any $k \in \mathbb{N}$ and for all $r, s = 1, \ldots, m+1, r \neq s$, there exist points

$$\left(x_{rs}^k, x_{sr}^k \right) \in B \left(x_r, \epsilon_k \right) \times B \left(x_s, \epsilon_k \right)$$

such that

$$\| u_k \left(x_{rs}^k \right) - u_k \left(x_{sr}^k \right) \| = \| x_{rs}^k - x_{sr}^k \|.$$

Using Arzelà–Ascoli theorem and passing to a subsequence we may assume that maps u_k converge locally uniformly to some 1-Lipschitz map u_0. Observe now that

$$x_{rs}^k \text{ converges to } x_r \text{ for all } r, s = 1, \ldots, m+1.$$

Thus, by the locally uniform convergence, u_0 is an isometry on $\{x_1, \ldots, x_{m+1}\}$. Observe that

$$\mu_{\epsilon_k} \text{ converges weakly* to } \mu_0 = \sum_{i=1}^{m+1} \delta_{x_i} v_i.$$

Now, Lemma 2 tells us that u_0 is an optimal potential of μ_0.

Suppose that points $x_1, \ldots, x_{m+1}$ are such that for $i \neq j, i, j = 1, \ldots, m$,

$$\left\langle \frac{x_i - x_{m+1}}{\|x_i - x_{m+1}\|}, \frac{x_j - x_{m+1}}{\|x_j - x_{m+1}\|} \right\rangle < \left\langle \frac{v_i}{\|v_i\|}, \frac{v_j}{\|v_j\|} \right\rangle. \tag{28}$$

Then if we define $h \colon \{x_1, \ldots, x_{m+1}\} \to \mathbb{R}^m$ by

$$h \left(x_{m+1} \right) = 0, h(x_i) = \|x_i - x_{m+1}\| \frac{v_i}{\|v_i\|} \text{ for } i = 1, \ldots, m,$$

then h is 1-Lipschitz. By the Kirszbraun theorem we may assume that h is defined on the entire space. Moreover for

$$\pi = \sum_{i=1}^{m+1} v_i \delta_{(x_i, x_{m+1})}$$

we have

$$P_1 \pi - P_2 \pi = \mu_0$$

and

$$\pi = \sum_{i=1}^m \frac{h \left(x_i \right) - h \left(x_{m+1} \right)}{\|x_i - x_{m+1}\|} \| v_i \| \delta_{(x_i, x_{m+1})}$$

Theorem 3 yields that h is an optimal potential and π is an optimal transport. It follows that

$$\| \mu_0 \|_{\mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)} = \sum_{i=1}^m \| v_i \| \| x_i - x_{m+1} \|.$$

Theorem 3 tells us that also

$$\pi = \sum_{i=1}^m \frac{u_0 \left(x_i \right) - u_0 \left(x_{m+1} \right)}{\|x_i - x_{m+1}\|} \| v_i \| \delta_{(x_i, x_{m+1})}$$

As u_0 is an isometry on $\{x_1, \ldots, x_{m+1}\}$, it follows that for $i, j = 1, \ldots, m$

$$\| h \left(x_i \right) - h \left(x_j \right) \| = \| x_i - x_j \|$$

which is not true, as the inequalities in (28) are strict. The obtained contradiction shows that there is no such sequence $(\epsilon_k)_{k=1}^{\infty}$, i.e. there exists $\epsilon_0 > 0$ such that for all $\epsilon \in (0, \epsilon_0)$ there exists a transport set with non-zero measure μ_ϵ for any optimal potential of μ_ϵ.

By Theorem 4 it follows that for such ϵ there is is no optimal transport with absolutely continuous marginals for μ_ϵ. ∎

The proof of the following theorem is based on the same idea as the proof of Theorem 5. Note that we do not require below that the norms on $\mathbb{R}^n$ and on $\mathbb{R}^m$ are Euclidean. For a 1-Lipschitz map $u \colon \mathbb{R}^n \to \mathbb{R}^m$ a leaf of u is a maximal, with respect to the order induced by inclusion, set $\mathcal{S}$ such that $u|_\mathcal{S}$ is an isometry. A transport set is defined as a set A that enjoys the property that if $x \in A$ belongs to a unique leaf of u, then for any $y \in \mathbb{R}^n$ such that $\|u(y) - u(x)\| = \|y - x\|$ there is $y \in A$. This is to say, the leaves and transport sets are defined as in the Euclidean case.

Theorem 6 *Let $m \leq n$. Suppose that the norm on $\mathbb{R}^m$ is strictly convex. Suppose that $\mathcal{F}$ is a locally uniformly closed subset of 1-Lipschitz maps of $\mathbb{R}^n$ to $\mathbb{R}^m$. Suppose that $\mathcal{F}$ has the property that for any absolutely continuous measure $\mu \in \mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$ and any $\mathcal{F}$-optimal potential u_0 of μ we have $\mu(A) = 0$ for any transport set A of u_0. Then either $m = 1$ or $m > 1$ and*

(i) $m = n$, any $u \in \mathcal{F}$ is affine, and there exists $u \in \mathcal{F}$ that is an isometry of $\mathbb{R}^n$ and of $\mathbb{R}^m$,
(ii) for any absolutely continuous μ, any $\mathcal{F}$-optimal potential of μ is an isometry on a maximal subspace $V \subset \mathbb{R}^n$, so that

$$\mu\left(\{x \in \mathbb{R}^n \mid Px \in A\}\right) = 0 \text{ for any Borel set } A \subset W; \tag{29}$$

here P denotes a projection onto a complement W of V.

Suppose that the norms are Euclidean. Then, if any $\mathcal{F}$-optimal potential is affine and is an isometry on a maximal subspace $V \subset \mathbb{R}^n$ such that (29) holds true, then $\mu(A) = 0$ for any transport set of its $\mathcal{F}$-optimal potential.

Proof Suppose that $m > 1$. Choose any pairwise distinct points $x_1, x_2, x_3 \in \mathbb{R}^n$ and any affinely independent $v_1, v_2, v_3 \in \mathbb{R}^m$ such that $\sum_{i=1}^{3} v_i = 0$. Let

$$v_0 = \sum_{i=1}^{3} v_i \delta_{x_i}.$$

Then $v_0 \in \mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$. For $\epsilon > 0$ let

$$v_\epsilon = \frac{1}{\lambda(B(0, \epsilon))} \sum_{i=1}^{3} v_i \lambda|_{B(x_i, \epsilon)}$$

Choose respective $\mathcal{F}$-optimal potentials u_ϵ for v_ϵ. These exist as $\mathcal{F}$ is locally uniformly closed. Observe that, by the assumption, $v_\epsilon(B_\epsilon) = 0$ for any Borel set B_ϵ consisting of trivial leaves of u_ϵ. Whence, v_ϵ is concentrated on non-trivial leaves of u_ϵ. Let $N_{i\epsilon}$ denote the union of all non-trivial leaves that intersect $B_{i\epsilon} = \mathrm{cl}B(x_i, \epsilon)$ for $i = 1, 2, 3$ and $\epsilon > 0$. The map

$$\mathbb{R}^n \setminus B_{i\epsilon} \ni x \mapsto \sup\left\{\frac{\|u(x) - u(y)\|}{\|x - y\|} \mid y \in B_{i\epsilon}\right\} \in \mathbb{R}$$

is lower semi-continuous. Note that

$$N_{i\epsilon} = \left\{x \in \mathbb{R}^n \setminus B_{i\epsilon} \mid \sup\left\{\frac{\|u(x) - u(y)\|}{\|x - y\|} \mid y \in B_{i\epsilon}\right\} = 1\right\} \cup (B_{i\epsilon} \cap A_\epsilon),$$

where A_ϵ denotes the union of all non-trivial leaves of u_ϵ and is Borel measurable by the argument of Lemma 3. Hence $N_{i\epsilon}$ is Borel measurable.

By the assumption,

$$\nu_\epsilon \left(N_{i\epsilon} \right) = 0,$$

which implies, as in the proof of Theorem 5, that

$$\| u_\epsilon \left(x_{rs}^\epsilon \right) - u_\epsilon \left(x_{sr}^\epsilon \right) \| = \| x_{rs}^\epsilon - x_{sr}^\epsilon \|$$

for some points

$$\left(x_{rs}^\epsilon, x_{sr}^\epsilon \right) \in B \left(x_r, \epsilon \right) \times B \left(x_s, \epsilon \right), r, s = 1, \ldots, 3, r \neq s.$$

By the Arzelà–Ascoli theorem and passing to a subsequence we may assume that u_ϵ converges locally uniformly to some $u_0 \in \mathcal{F}$, which is an $\mathcal{F}$-optimal potential of ν_0 by Lemma 2. By the uniform convergence we infer that u_0 is isometric on $\{x_1, x_2, x_3\}$. Let now $x_2 = t x_1 + (1-t) x_3$ for some $t \in (0, 1)$. Then any map $f : \{x_1, x_2, x_3\} \to \mathbb{R}^m$ that is isometric satisfies

$$f \left(t x_1 + (1 - t) x_3 \right) = t f \left(x_1 \right) + (1 - t) f \left(x_3 \right). \tag{30}$$

Indeed, as f is isometric,

$$\| f \left(x_2 \right) - f \left(x_1 \right) \| = (1 - t) \| x_3 - x_1 \| \text{ and } \| f \left(x_3 \right) - f \left(x_2 \right) \| = t \| x_3 - x_1 \|.$$

As $\| f \left(x_3 \right) - f \left(x_1 \right) \| = \| x_3 - x_1 \|$ it follows that we have equality in the triangle inequality

$$\| f \left(x_3 \right) - f \left(x_1 \right) \| \leq \| f \left(x_2 \right) - f \left(x_1 \right) \| + \| f \left(x_3 \right) - f \left(x_2 \right) \|.$$

By the strict convexity it follows that there is $\lambda > 0$ such that

$$f \left(x_2 \right) - f \left(x_1 \right) = \lambda \left(f \left(x_3 \right) - f \left(x_1 \right) \right).$$

Taking the norms we arrive at (30). A function f that satisfies (30) may be extended to $\mathbb{R}^n$ to an affine map that has derivative of operator norm at most m. This follows by the Hahn–Banach theorem. As u_0 is isometric on $\{x_1, x_2, x_3\}$, we infer that

$$\sum_{i=1}^{3} \langle u_0 \left(x_i \right), v_i \rangle \leq \sup \left\{ \sum_{i=1}^{3} \langle f \left(x_i \right), v_i \rangle \mid f : \mathbb{R}^n \to \mathbb{R}^m \text{ is linear and } \| f \| \leq m \right\}$$

Note now that the set of vectors v_1, v_2, v_3 that sum up to zero and are affinely independent is dense in the set of vectors v_1', v_2', v_3' that sum up to zero. Moreover, u_0 is an $\mathcal{F}$-optimal potential for ν_0. We conclude that for any $u \in \mathcal{F}$ and any vectors v_1, v_2, v_3 that sum up to zero there is

$$\sum_{i=1}^{3} \langle u(x_i), v_i \rangle \leq \sup \left\{ \sum_{i=1}^{3} \langle f(x_i), v_i \rangle \mid f : \mathbb{R}^n \to \mathbb{R}^m \text{ is linear and } \| f \| \leq m \right\}$$

Take now $v_2 = v, v_1 = -t v$ and $v_3 = -(1 - t) v$ with $t \in (0, 1)$ as above and any $v \in \mathbb{R}^m$. We infer that

$$\langle u(x_2) - t u(x_1) - (1 - t) u(x_3), v \rangle \leq 0.$$

As this holds for any v we infer that u is affine. This proves part of i).

If u is affine then there exists a subspace $V \subset \mathbb{R}^n$, possibly trivial, i.e. $V = \{0\}$, such that any set of the form

$$\{x \in \mathbb{R}^n \mid P x \in A\} \tag{31}$$

for a Borel measurable set $A \subset W$ is a transport set of u. Here P denotes a projection onto a complement W of V. Indeed, let $V \subset \mathbb{R}^n$ be a maximal subspace such that $u|_V$ is an isometry. Suppose that V is not a leaf of u. Then there exists $y \notin V$ such that for all $x \in V$

$$\|u(y) - u(x)\| = \|y - x\|.$$

It follows that for all non-zero $\lambda \in \mathbb{R}$

$$\left\| u(y) - u\left(\frac{x}{\lambda}\right) \right\| = \left\| y - \frac{x}{\lambda} \right\|$$

for all $x \in V$. Hence for all $\lambda \in \mathbb{R}$ we have $\|u(\lambda y) - u(x)\| = \|\lambda y - x\|$. As u is affine, it is also an isometry on $V + \mathbb{R}y$. This contradiction shows that V is a leaf of u. Thus ii) is proven.

We shall now provide an example of a vector measure μ such that for any proper subspace V and any x_0 there is $c > 0$ such that

$$\mu\left(\left\{ x \in \mathbb{R}^n \mid \|P\,(x - x_0)\| \le c \right\}\right) \ne 0. \tag{32}$$

Choose any $x_1, \ldots, x_{m+1} \in \mathbb{R}^n$ affinely independent. Let $\epsilon > 0$ be a number such that any set $\{y_1, \ldots, y_{m+1}\}$, with $y_i \in B(x_i, \epsilon)$, $i = 1, \ldots, m+1$, is affinely independent. Choose vectors $v_1, \ldots, v_{m+1} \in \mathbb{R}^m$ that add up to zero and are affinely independent. Let

$$\mu = \sum_{i=1}^{m+1} v_i \lambda|_{B(x_i, \epsilon)},$$

where λ denotes the Lebesgue measure. Choose any proper affine subspace $V \subset \mathbb{R}^n$. Then V intersects at most m of the balls $B(x_i, \epsilon)$, $i = 1, \ldots, m+1$. So does the set

$$\left\{ x \in \mathbb{R}^n \mid \|P\,(x - x_0)\| \le c \right\}$$

provided that $c > 0$ is sufficiently small. Thus (32) follows. It implies, by ii), that $V = \mathbb{R}^n$. We have shown that any $\mathcal{F}$-optimal potential of μ has to be an isometry. Hence $m = n$ and the proof of i) is complete.

To prove the last part of the theorem, it is enough to prove that the translates of V are the only leaves of an affine map. This holds true, since any point in $\mathbb{R}^n$ is covered by a translate of V and the leaves of a map foliate $\mathbb{R}^n$, up to Lebesgue measure zero, if the considered norms are Euclidean, c.f. [12]. □

References

1. Ambrosio, L.: Optimal Transport Problems. Springer, Berlin (2003)

2. Bakry, D.: L'hypercontractivité et son utilisation en théorie des semigroupes. In: Lectures on probability theory (Saint-Flour, 1992), *Lecture Notes in Math.*, vol. 1581, pp. 1–114. Springer, Berlin (1994). https://doi.org/10.1007/BFb0073872

3. Bakry, D., Émery, M.: Diffusions hypercontractives. In: Azéma, J., Yor, M. (eds.) Séminaire de Probabilités XIX 1983(84), pp. 177–206. Springer, Berlin (1985)

4. Bianchini, S., Daneri, S.: On Sudakov's type decomposition of transference plans with norm costs. Mem. Am. Math. Soc. **251** (2013)

5. Caffarelli, L., Feldman, M., McCann, R.J.: Constructing optimal maps for MongeÃ transport problem as a limit of strictly convex costs. J. Am. Math. Soc. **15**(1), 1–26 (2002)

6. Caravenna, L.: A proof of Sudakov theorem with strictly convex norms. Math. Z. **268**(1), 371–407 (2011). https://doi.org/10.1007/s00209-010-0677-6

7. Cavalletti, F., Mondino, A.: Sharp and rigid isoperimetric inequalities in metric-measure spaces with lower Ricci curvature bounds. Invent. Math. **208**(3), 803–849 (2017). https://doi.org/10.1007/s00222-016-0700-6

8. Cavalletti, F., Mondino, A.: Sharp geometric and functional inequalities in metric measure spaces with lower Ricci curvature bounds. Geom. Topol. **21**(1), 603–645 (2017). https://doi.org/10.2140/gt.2017.21.603

9. Chen, Y., Georgiou, T., Tannenbaum, A.: Vector-valued optimal mass transport. SIAM J. Appl. Math. **78**, 1682–1696 (2018)

10. Chen, Y., Haber, E., Yamamoto, K., Georgiou, T.T., Tannenbaum, A.: An efficient algorithm for matrix-valued and vector-valued optimal mass transport. J. Sci. Comput. **77**(1), 79–100 (2018). https://doi.org/10.1007/s10915-018-0696-8

11. Ciosmak, K.J.: Continuity of extensions of Lipschitz maps. To appear in Israel Journal of Mathematics, arXiv e-prints arXiv:1904.02993 (2019)

12. Ciosmak, K.J.: Leaves decompositions in Euclidean spaces. J. Math. Pure. Appl. (2021). https://doi.org/10.1016/j.matpur.2021.08.003

13. Evans, L., Gangbo, W.: Differential equations methods for the Monge-Kantorovich mass transfer problem. Mem. Amer. Math. Soc. **137**(653), viii+66 (1999). https://doi.org/10.1090/memo/0653

14. Kantorovich, L.: On the translocation of masses. C.R. (Doklady) Acad. Sci. URSS (N.S.) (1942)

15. Kantorovich, L.: On the translocation of masses. J. Math. Sci. **133**(4), 1381–1382 (2006). https://doi.org/10.1007/s10958-006-0049-2

16. Kirszbraun, M.: Über die zusammenziehende und Lipschitzsche Transformationen. Fundamenta Mathematicae **22**(1), 77–108 (1934)

17. Klartag, B.: Needle decompositions in Riemannian geometry. Mem. Am. Math. Soc. **249**(1180) (2017)

18. Lott, J., Villani, C.: Ricci curvature for metric-measure spaces via optimal transport. Ann. Math. (2) **169**(3), 903–991 (2009). https://doi.org/10.4007/annals.2009.169.903

19. Monge, G.: Mémoire sur la théorie des déblais et des remblais. In: Histoire de l'Académie Royale de Sciences de Paris, pp. 666–704 (1781)

20. Ryu, E.K., Chen, Y., Li, W., Osher, S.: Vector and matrix optimal mass transport: theory, algorithm, and applications. Preprint (2017)

21. Sturm, K.T.: On the geometry of metric measure spaces. I. Acta Math. **196**(1), 65–131 (2006). https://doi.org/10.1007/s11511-006-0002-8

22. Sturm, K.T.: On the geometry of metric measure spaces. II. Acta Math. **196**(1), 133–177 (2006). https://doi.org/10.1007/s11511-006-0003-7

23. Sudakov, V.: Geometric problems in the theory of infinite-dimensional probability distributions. Trudy Mat. Inst. Steklov. **141** (1976)

24. Trudinger, N.S., Wang, X.J.: On the Monge mass transfer problem. Calc. Var. Partial. Differ. Equ. **13**(1), 19–31 (2001)

25. Villani, C.: Topics in optimal transportation, *Graduate Studies in Mathematics*, vol. 58. American Mathematical Society, Providence, RI (2003). https://doi.org/10.1007/b12016

26. Villani, C.: Optimal transport, *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, vol. 338. Springer-Verlag, Berlin (2009). https://doi.org/10.1007/978-3-540-71050-9. Old and new