Available online at www.sciencedirect.com

ScienceDirect

Ann. I. H. Poincaré – AN 34 (2017) 1669–1685

ANNALES DE L'INSTITUT HENRI POINCARÉ
ANALYSE NON LINÉAIRE

www.elsevier.com/locate/anihpc

From optimal transportation to optimal teleportation

G. Wolansky

Department of Mathematics, Technion, Haifa 32000, Israel

Received 23 March 2016; accepted 1 December 2016

Available online 15 December 2016

Abstract

The object of this paper is to study estimates of $\epsilon^{-q} W_p(\mu + \epsilon\nu, \mu)$ for small $\epsilon > 0$. Here W_p is the Wasserstein metric on positive measures, $p > 1$, μ is a probability measure and ν a signed, neutral measure ($\int d\nu = 0$). In [16] we proved uniform (in ϵ) estimates for $q = 1$ provided $\int \phi d\nu$ can be controlled in terms of $\int |\nabla\phi|^{p/(p-1)} d\mu$, for any smooth function ϕ.

In this paper we extend the results to the case where such a control fails. This is the case where, e.g., μ has a disconnected support, or the dimension d of μ (to be defined) is larger or equal to $p/(p-1)$.

In the latter case we get such an estimate provided $1/p + 1/d \neq 1$ for $q = \min(1, 1/p + 1/d)$. If $1/p + 1/d = 1$ we get a log-Lipschitz estimate.

As an application we obtain Hölder estimates in W_p for curves of probability measures which are absolutely continuous in the total variation norm.

In case the support of μ is disconnected (corresponding to $d = \infty$) we obtain sharp estimates for $q = 1/p$ ("optimal teleportation"):

$$\lim_{\epsilon \to 0} \epsilon^{-1/p} W_p(\mu, \mu + \epsilon\nu) = \|\nu\|_\mu$$

where $\|\nu\|_\mu$ is expressed in terms of optimal transport on a metric graph, determined only by the relative distances between the connected components of the support of μ, and the weights of the measure ν in each connected component of this support.

Keywords: Optimal transport; Monge–Kantorovich; Wasserstein metric

1. Introduction

1.1. Notation

- $\Omega \subset \mathbb{R}^k$ is a compact set, equal to the closure of its interior.
- $\mathcal{M} := \mathcal{M}(\Omega)$ is the set of Borel measures on Ω. $\mathcal{M}_+$ is the set of non-negative measures in $\mathcal{M}$. $\mathcal{M}_1$ is the set of probability (normalized) measures in $\mathcal{M}_+$.

E-mail address: gershonw@math.technion.ac.il.

http://dx.doi.org/10.1016/j.anihpc.2016.12.003

- The duality between $\mathcal{M}(\Omega)$ and $C(\Omega)$ (continuous functions) is denoted by $\langle \mu, \phi \rangle$, where $\mu \in \mathcal{M}$ and $\phi \in C(\Omega)$. This duality implies an order relation on $\mathcal{M}$: $\mu_1 \geq \mu_2$ iff $\langle \mu_1, \phi \rangle \geq \langle \mu_2, \phi \rangle$ for any non-negative $\phi \in C(\Omega)$.
- For $\mu \in \mathcal{M}_+$, supp(μ) is the minimal closed set $A \subset \Omega$ such that $\mu(A) = \mu(\Omega)$.
- If $\mu \in \mathcal{M}_+$ then $|\mu| :=< \mu, 1 >$ (the "mass" of μ).
- For $\nu \in \mathcal{M}$, $\nu_\pm \in \mathcal{M}_+$ is the factorization of ν into positive and negative parts, namely $\nu = \nu_+ - \nu_-$ such that $\|\nu\|_{TV} := |\nu_+| + |\nu_-|$ is the total variation norm of ν (in particular, $\nu_\pm$ are mutually singular).
- $\mathcal{M}_0$ is the set of measures $\nu = \nu_+ - \nu_-$ where $\nu_\pm \in \mathcal{M}_+$ and $|\nu_-| = |\nu_+|$. In particular, for any $\nu \in \mathcal{M}_0$ there exists a single factorization $\nu_\pm$.

1.2. Background

Recall the definition of the p-Wasserstein metric ($p > 1$) on $\mathcal{M}_1(\Omega)$:

$$W_p(\mu_1, \mu_2) := \left(\inf_{\pi \in \Pi(\mu_1, \mu_2)} \int_\Omega \int_\Omega |x - y|^p \pi(dxdy) \right)^{1/p} \tag{1}$$

where $\mu_1, \mu_2 \in \mathcal{M}_1$,

$$\Pi(\mu_1, \mu_2) := \{\pi \in \mathcal{M}_1(\Omega \times \Omega)\ ;\ \ \pi_{\#1} = \mu_1;\ \ \pi_{\#2} = \mu_2\} \tag{2}$$

Here $\pi_{\#,1,2}$ represents the first and second marginals of π on Ω, respectively.

The $(C(\Omega))^*$ topology restricted to $\mathcal{M}_1$ can be metrized by W_p with $p \geq 1$ [14, Theorem 6.9]. See also [8,9,13,4,15,11,5,12] among many other sources for this and related metrics.

W_p can be trivially extended to any pair $\mu_1, \mu_2 \in \mathcal{M}_+$ provided $|\mu_1| = |\mu_2|$. This extension is defined naturally by the homogeneity relation

$$W_p(\alpha\mu_1, \alpha\mu_2) = \alpha^{1/p} W_p(\mu_1, \mu_2) \tag{3}$$

for $\alpha > 0$.

Note that the total variation of $\nu = \nu^+ - \nu_- \in \mathcal{M}_0$ is given by

$$\|\nu\|_{TV} = \inf_{\pi \in \Pi(\nu_+, \nu_-)} \int_\Omega \int_\Omega d(x, y)\pi(dxdy)$$

where d is the discrete metric ($d(x, y) = 1$ if $x \neq y$, $d(x, x) = 0$), see [14], Theorem 6.15. Since $|x - y|^p < Diam^p(\Omega)d(x, y)$ for any x, y in the compact set Ω, then

$$W_p(\nu_+, \nu_-) \leq Diam(\Omega)\, \|\nu\|_{TV}^{1/p}\ , \tag{4}$$

hence, by the *principle of monotone additivity* (see Proposition 3.2 below) and (3),

$$\epsilon^{-1/p} W_p(\mu + \epsilon\nu_+, \mu + \epsilon\nu_-) \leq Diam(\Omega)\|\nu\|_{TV}^{1/p} \tag{5}$$

for any $\epsilon > 0$, provided $\mu \in \mathcal{M}_+$.

Lemma 5.6 in [16] (see also Theorem 7.26 in [13]) implies that for any $\nu = \nu_+ - \nu_- \in \mathcal{M}_0$, $\nu_\pm \in \mathcal{M}_+$ and any probability measure μ

$$\liminf_{\epsilon \searrow 0} \epsilon^{-1} W_p(\mu + \epsilon\nu_+, \mu + \epsilon\nu_-) \geq \sup_{\phi \in \mathcal{B}_p(\mu)} \langle \nu, \phi \rangle \tag{6}$$

where, if $p > 1$,

$$\mathcal{B}_p(\mu) := \left\{ \phi \in C^1(\Omega);\ \int_\Omega |\nabla\phi|^{p/(p-1)} d\mu \leq 1 \right\} \tag{7}$$

while Lemma 5.7 establishes the opposite inequality for lim sup in (6) (in particular, the existence of a limit), if ν is absolutely continuous with respect to μ and both measures are regular enough.

Remark 1.1. Note that for $p = 1$ an equality

$$\epsilon^{-1} W_1(\mu + \epsilon\nu_+, \mu + \epsilon\nu_-) = \sup_{\phi \in \mathcal{B}_1} \langle \nu, \phi \rangle$$

holds for any $\epsilon > 0$ where $\mathcal{B}_1$ is the set of 1-Lipschitz functions on Ω.

In the cases where there is equality in (6) we obtain

$$\liminf_{\epsilon \searrow 0} \epsilon^{-1} W_p(\mu + \epsilon\nu_+, \mu + \epsilon\nu_-) \leq D_p(\mu) \|\nu\|_{TV} \tag{8}$$

where

$$D_p(\mu) := \sup_{\phi \in \mathcal{B}_p(\mu)} \left(\sup_{x \in \mathrm{supp}(\mu)} \phi(x) - \inf_{x \in \mathrm{supp}(\mu)} \phi(x) \right)$$

is the maximal oscillation of functions in $\mathcal{B}_p(\mu)$ restricted to $\mathrm{supp}(\mu)$ and is, of course, independent of ν.

In this paper we consider the case $D_p(\mu) = \infty$.

1.3. Measures of connected support

Suppose μ is a uniform (Lebesgue) measure on a "nice" domain $\Omega \subset \mathbb{R}^d$ (e.g. a ball in $\mathbb{R}^d$). Then $\mathcal{B}_p(\mu)$ is dense in the unit ball of the Sobolev space $\mathbb{W}^{1,p\prime}(\Omega)/\mathbb{R}$ (with respect to that norm) where $p\prime := p/(p-1)$. Sobolev embedding theorem then implies that $D_p(\mu) < \infty$ if $d < p\prime$ (where $\mathbb{W}^{1,p}(\Omega)$ is embedded in $C(\Omega)$), while $D_p(\mu) = \infty$ if $p\prime \leq d$ (see Remark 2.4).

The first result (Theorem 1) deals with measures μ of connected support. We introduce the notion of dimensionality of measure and define d-connected property of such measures in Definitions 2.1 and 2.2.

For strong d-connected measure μ and under the assumption that the support of ν is contained in the support of μ we state the existence of a constant C depending only on μ, and an exponent $q \in [1/p, 1]$ for which

$$\sup_{\epsilon > 0} \epsilon^{-q} W_p(\mu + \epsilon\nu_+, \mu + \epsilon\nu_-) \leq C \|\nu\|_{TV} \quad , \tag{9}$$

where

$$q = \min(1, 1/d + 1/p) \ \text{ if } \ 1/d + 1/p \neq 1 \, . \tag{10}$$

The second case $d = p/(p-1)$ (i.e. $1/d + 1/p = 1$) corresponds to the critical Sobolev embedding $\mathbb{W}^{1,p\prime}(\mathbb{R}^d)$ and leads to a log Lipschitz estimate

$$\sup_{\epsilon > 0} \frac{1}{\epsilon \ln^{1/p}(1/\epsilon + 1)} W_p(\mu + \epsilon\nu_+, \mu + \epsilon\nu_-) \leq C \|\nu\|_{TV} \quad . \tag{11}$$

1.4. Application: curves of measures

Let $I \subset \mathbb{R}$ be an interval and $\vec{\mu} \in \mathcal{M}_+(\Omega \times I)$ such that its t marginal $\mu_{(t)}$ is a probability measure on Ω for any $t \in I$ [3]. Then

$$\mathbb{R} \supset I \ni t \mapsto \mu_{(t)} \in \mathcal{M}_1$$

can be viewed as a curve in $\mathcal{M}_1 := \mathcal{M}_1(\Omega)$ parameterized in I. We say that $\vec{\mu} \in AC^r(I, \mathcal{M}_1; TV)$ for some $\infty \geq r \geq 1$ if there exists a non-negative $m \in \mathbb{L}^r(I)$ such that

$$\|\mu_{(t)} - \mu_{(\tau)}\|_{TV} \leq \int_\tau^t m(s) ds$$

for any $t > \tau \in I$.

The *metric derivative* [2,1] of $\vec{\mu}$ with respect to the TV norm is

$$\vec{\mu}'_{(t)}(t) := \lim_{\tau\to t} \frac{\|\mu_{(t)} - \mu_{(\tau)}\|_{TV}}{|t-\tau|} .$$

By Theorem 1.1.2 in [1], the metric derivative exists t a.s. in I for $\vec{\mu} \in AC^r(I, \mathcal{M}_1; TV)$.

On the other hand, $\mathcal{M}_1$ can also be considered as a metric space with respect to the Wasserstein metric W_p. Recalling (4), we observe that, if $\vec{\mu} \in AC^r(I, \mathcal{M}_1; TV)$ then

$$W_p(\mu_{(t)}, \mu_{(\tau)}) \le Diam(\Omega)\|\mu_{(t)} - \mu_{(\tau)}\|_{TV}^{1/p} \le Diam(\Omega)(\int_\tau^t m)^{1/p} \le Diam(\Omega)\|m\|_r^{1/p}|t-\tau|^{\frac{r-1}{r}\frac{1}{p}} .$$

So, we cannot expect that such a curve $\vec{\mu} \in AC^r(I, \mathcal{M}_1; TV)$ is more than $(r-1)/rp$-Hölder with respect to the Wasserstein metric W_p.

In Theorem 2 we state that if the support of $\vec{\mu}$ is monotone non-increasing, namely $\mathrm{supp}(\mu_{(t)}) \subseteq \mathrm{supp}(\mu_{(\tau)})$ for any $t > \tau$, and $\mathrm{supp}(\mu_{(t)})$ is strongly d-connected for any $t \in I$ then we can improve this estimate: Under the above conditions, $\vec{\mu}$ is $q(r-1)/r$-Hölder on I in $(\mathcal{M}_1, W_p)$, (q-Hölder if $r = \infty$) where q given by (10).

Moreover, if $1/p + 1/d > 1$ then $q = 1$ (10) and $\vec{\mu} \in AC^r(I, \mathcal{M}_1; W_p)$ as well. This implies the existence of a Borel vector field $v \in \mathbb{L}^r\left(I, \mathbb{L}^p(\mu_{(t)})\right)$ for which the continuity equation

$$\partial_t \mu + \nabla_x \cdot (\mu v) = 0 \tag{12}$$

holds as a distribution [1].

To illustrate the above results, consider

$$\mu_{(t)} = m(t)\delta_{x_0} + (1 - m(t))\delta_{x_1} \tag{13}$$

where $x_0 \neq x_1$ and $t \mapsto m(t) \in (0,1)$ is a non-constant smooth function. Then $\dot{\mu}_{(t)} = \dot{m}(t)(\delta_{x_0} - \delta_{x_1}) \in \mathcal{M}$ and $\|\dot{\mu}_{(t)}\|_{TV} = 2|\dot{m}(t)|$.

If we consider the above curve in $(\mathcal{M}_1, W_p)$ where $p > 1$ then the metric derivative does not exist.

Indeed, since $W_p^p(\mu_{(t)}, \mu_{(\tau)}) = |m(t) - m(\tau)| \times |x - x_0|^p$, all we can obtain is $1/p$ Hölder estimate:

$$\lim_{\tau\to t} \frac{W_p(\mu_{(t)}, \mu_{(\tau)})}{|t-\tau|^{1/p}} = \lim_{\tau\to t} \frac{|m(t) - m(\tau)|^{1/p}}{|t-\tau|^{1/p}} |x - x_0| = |\dot{m}|^{1/p}(t)|x - x_0| .$$

Now, replace (13) by

$$\mu_{(t)} = m(t)\delta_{x_0} + (1 - m(t))\delta_{x_1} + \bar{\mu} \tag{14}$$

(recall (3)) where $\bar{\mu} \in \mathcal{M}_+$ a stationary (independent of t) positive, strongly d-connected measure whose support contains x_0, x_1. Even though $\dot{\mu} = \dot{m}(\delta_{x_0} - \delta_{x_1})$ is the same for both (13) and (14)), we can find out that for μ given by (14)

$$\frac{W_p(\mu_{(t)}, \mu_{(\tau)})}{|t-\tau|^q} \le C(\bar{\mu})|\dot{m}|$$

for $q = \min[1, 1/p + 1/d]$ (provided $1/p + 1/d \neq 1$), or the Log-Lipschitz estimate

$$\frac{W_p(\mu_{(t)}, \mu_{(\tau)})}{|t-\tau| \ln^{1/p}(1/|t-\tau|)} \le C(\bar{\mu})$$

if $1/p + 1/d = 1$. In particular (14) is uniformly Lipschitz if $1/p + 1/d > 1$. If this is the case, it is absolutely continuous in W_p. Hence the continuity equation (12) is satisfied for some Borel vectorfield v [1].

To elaborate further, let us consider the case where $\mu_{(t)}$ is supported in an interval $J \subset \mathbb{R}$ and $[x_0, x_1] \subset J$:

$$\mu_{(t)} := \beta 1_J(dx) + m(t)\delta_{x_1} + (1 - m(t))\delta_{x_0}$$

where $\beta > 0$ is a constant. Then $\mu_{(\cdot)}$ is strongly 1-connected (see Definition 2.2). Hence for any $p > 1$, $\mu_{(\cdot)}$ is Lipschitz in W_p. In particular, it satisfies (12). It can be verified that the transporting vector field is nothing but

$$v(x,t) = \beta^{-1}\dot{m}(t) \ \text{ if } \ x_0 < x < x_1, \ , \quad v(x_0, t) = v(x_1, t) = 0 \ ; \ \ \forall t \in I$$

and v is arbitrary otherwise.

The case $q < 1$ corresponds, in this context, to a "teleportation": No vector field v exists for which an orbit $\mu_{(t)}$ is transported via the continuity equation (12). In particular, if the support of μ is disconnected (e.g. $\beta = 0$ above).

1.5. Disconnected support

In the last part of the paper we discuss the case of disconnected support of $\mu \in \mathcal{M}_+$, corresponding to $d = \infty$. In that case $q = 1/p$. Under appropriate condition we state in Theorem 3 that there exists a *sharp* limit

$$\lim_{\epsilon \searrow 0} \epsilon^{-1/p} W_p(\mu + \epsilon \nu_+, \mu + \epsilon \nu_-) = \lim_{\epsilon \searrow 0} \epsilon^{-1/p} W_p(\mu + \epsilon \nu, \mu) := \|\nu\|_\mu^{1/p}$$

where $\|\nu\|_\mu$ is defined in terms of an optimal transport on a *finite, metric graph*. This is the rational behind the title "optimal teleportation".

To describe the nature of $\|\nu\|_\mu$, consider a finite graph whose vertices are identified with the connected components A_i of the support of μ. The length of an edge connecting two vertices is defined as the p power of the distance between the corresponding supports. We then consider the discrete metric space composed of these vertices, subjected to the geodesic distance corresponding the edge's length defined above.

At each vertex i of this graph let $\bar{\nu}_i \in \mathbb{R}$ be the weight of the measure ν restricted to corresponding component A_i. By neutrality $\sum_i \bar{\nu}_i = 0$.

Then $\|\nu\|_\mu$ is just the optimal transport cost of $\{\bar{\nu}_i > 0\}$ to $\{\bar{\nu}_i < 0\}$ for the above defined metric (cf. Fig. 2).

2. Detailed description of main results

We start by posing some assumptions on a measure $\mu \in \mathcal{M}_1$:

Definition 2.1. μ is d-connected if supp(μ) is arc-connected and there exists $K, \delta > 0$ such that for any $x \in \text{supp}(\mu)$ and any $0 < r < \delta$

$$\mu(B_r(x)) \geq K r^d \; . \tag{15}$$

Remark 2.1. Condition (15) states, in fact, that μ is *d-Ahlfors regular* from below on its support. See e.g. [6] for more general definitions. Note also that if μ is d-connected then μ is d^*-connected for any $d^* \geq d$.

Actually, we need a stronger definition for d-connected measure:

Definition 2.2. μ is strongly d-connected if there exist $L, K > 0$, $2 \leq N \in \mathbb{N}$ and a measure space (D, β) such that for any $x_0, x_1 \in \text{supp}(\mu)$ there are $k \leq N$ points $y_1 = x_0, y_2, \dots y_k = x_1$ in supp(μ) and $k - 1$ measurable mappings $\Phi_j : J = [0, 1] \times D \to \text{supp}(\mu)$, $j = 1, \dots k - 1$ such that

i) $\Phi_j(\cdot, b) : [0, 1] \to \Omega$ is L-Lipschitz on $[0, 1]$ for any $b \in D$.
ii) Φ_j is injective on $(0, 1) \times D$, $\Phi_j(0, b) = y_j$, $\Phi_j(1, b) = y_{j+1}$ for any $b \in D$.
iii) $\Phi_{j,\#}(\rho) \leq \mu$ where $\rho \in \mathcal{M}_+(J)$ given by the density $\rho(ds, d\beta) = K s^{d-1}(1 - s)^{d-1} ds d\beta$.

See Fig. 1.

Remark 2.2. We conjecture that d-connectedness should be enough for the main results of this paper. Unfortunately we had to adopt the stronger definition for proving these results. Note that strong d-connected set is also d-connected. In fact, supp(μ) is arc connected and $\Phi_1([0, r/L] \times D) \subset B_r(x_0)$ by (i, ii). By (iii), $\mu(B_r(x_0)) \geq K\beta(D) \int_0^{r/L} s^{d-1}(1 - s)^{d-1} ds$, hence if, say, $r < L/2$ then $\mu(B_r(x_0)) \geq K\beta(D) r^d / (d L^d 2^d)$.

Examples:

- Let Y be a convex subset of an m $(\leq k)$ dimensional hyperplane in $\mathbb{R}^k$. Let $\mu \geq c\mathcal{H}^m(T)$, $c > 0$, $\mathcal{H}^m(Y)$ being the m-Hausdorff measure on Y. Then μ is strongly m-connected $(N = 2)$. The same for a starshaped Y $(N = 3)$.

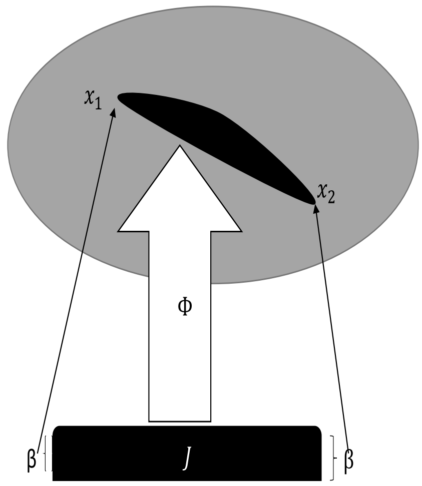

Fig. 1. Mapping of J to $\text{supp}(\mu)$ via Φ.

- μ is uniformly distributed on the wedge
$$\{(x,y)\in\mathbb{R}^k\ ;\ 0\le x\le 1,\ y\in\mathbb{R}^{k-1},\ |y|\le x^\beta\}$$
where $k>1$, $\beta\ge 0$. μ is strongly $\beta(k-1)+1$ connected if $\beta\ge 1$ and strongly k connected if $0\le\beta\le 1$ $(N=2)$.
- $\Omega=[0,3]\subset\mathbb{R}$ and μ has a density proportional to $x\mapsto x(x-1)^2(x-2)^2$. In that case μ is strongly 3-connected and $N=3$.

2.1. Connected support

Theorem 1. *Suppose μ is strongly d-connected ($d\ge 1$) and $\nu=\nu_+-\nu_-\in\mathcal{M}_0$ such that* $\text{supp}(\nu_\pm)\subset\text{supp}(\mu)$. *Then there exists C depending only on μ such that*

$$\sup_{\epsilon>0}\epsilon^{-q}W_p(\mu+\epsilon\nu_+,\mu+\epsilon\nu_-)<C\|\nu\|_{TV} \tag{16}$$

where $q=\min(1,1/d+1/p)$ provided $p\neq d/(d-1)$.

In the critical case $p=d/(d-1)$ (where $q=1$)

$$\sup_{\epsilon>0}\frac{1}{\epsilon\ln^{1/p}(1/\epsilon+1)}W_p(\mu+\epsilon\nu_+,\mu+\epsilon\nu_-)<C\|\nu\|_{TV}\ .$$

In particular there exists $C=C(\mu)$ for which

$$W_p(\mu+\nu_+,\mu+\nu_-)\le C(\mu)\|\nu\|^q_{TV} \tag{17}$$

if $p\neq d/(d-1)$, while if $p=d/(d-1)$,

$$W_p(\mu+\nu_+,\mu+\nu_-)\le C(\mu)\|\nu\|_{TV}\ln^{1/p}\left(\|\nu\|^{-1}_{TV}+1\right) \tag{18}$$

holds for any balanced pair $\nu=\nu_+-\nu_-$.

Remark 2.3. By Proposition 3.2 below we can observe that the *optimal* $C(\mu)$ in (17), (18) is monotone non-increasing in μ, that is $C(\mu_1)\ge C(\mu_2)$ if $\mu_1\le\mu_2$. By the same Proposition we can also assume that $\nu_\pm$ is a factorization of ν, namely $\|\nu\|_{TV}=|\nu_+|+|\nu_-|$.

Remark 2.4. We may now make a connection between (6), (7), Theorem 1 and the Sobolev embedding Theorem. Consider the Sobolev space

$$\mathbb{W}^{1,p\prime}(\Omega) := \left\{\phi \in \mathbb{L}^{p\prime}(\Omega)\ ;\ \nabla\phi \in \mathbb{L}^{p\prime}\right\}$$

where $p > 1$, $p\prime := p/(p-1)$ and $\Omega \subset \mathbb{R}^d$. If $p\prime > d$ then $\mathbb{W}^{1,p\prime}(\Omega)$ is embedded in the space of bounded continuous functions $C(\Omega)$. Suppose μ is the Lebesgue measure on a convex set $\Omega \subset \mathbb{R}^d$ (so, in particular, d-connected). This implies that the $\mathbb{W}^{1,p'}$ closure of $\mathcal{B}_p(\mu)$ is embedded in $C(\Omega)$. Let $\nu = \delta_{x_0} - \delta_{x_1}$ where $x_0, x_1 \in \text{supp}(\mu)$. Then the right hand side of (6) is finite. On the other hand, the case $p\prime > d$ corresponds to the case $q = 1$ so (16) is consistent with (6) in that case.

Recall that the case $p\prime = d$ corresponds to the *critical Sobolev embedding* where $\mathbb{W}^{1,d}$ (or $\mathcal{B}_p(\mu)$) just fails to be embedded in the space of continuous functions. In that case $D_p(\mu) = \infty$ (see (8)). The bound of (18) suggests that a Log-Lipschitz estimate corresponds to a critical Sobolev embedding in the context of Wasserstein metric.

2.2. Curves of probability measures

Let $\vec{\mu} := \{\mu_{(t)}\}$, $t \in I$ be a curve of probability measures

$$\mathbb{R} \supset I \ni t \mapsto \mu_{(t)} \in \mathcal{M}_1\ .$$

Recall that $\vec{\mu} \in AC^r(I, \mathcal{M}_1; TV)$ for some $\infty \geq r \geq 1$ if $\exists m \in \mathbb{L}^r(I)$ such that

$$\|\mu_{(t)} - \mu_{(\tau)}\|_{TV} \leq \int_\tau^t m(s)ds$$

for any $t > \tau \in I$.

Theorem 2. *Suppose $\vec{\mu} \in AC^r(I, \mathcal{M}_1; TV)$ for some $\infty \geq r > 1$. Assume also that the support of $\vec{\mu}$ is non-increasing, namely* $\text{supp}(\mu_{(t)}) \subseteq \text{supp}(\mu_{(\tau)})$ *for any $\tau < t \in I$, and* $\text{supp}(\mu_{(t)})$ *is uniformly d-connected with respect to t (that is, N, K, L can be chosen independently of t in Definition 2.2).*

Then

i) *For any $p > 1$, $p/(p-1) \neq d$, μ is uniformly $q(r-1)/r$-Hölder (q-Hölder if $r = \infty$) in the Wasserstein metric W_p where $q = \min(1, 1/d + 1/p)$, namely*

$$W_p(\mu_{(t)}, \mu_{(\tau)}) \leq C|t-\tau|^{q(r-1)/r}$$

where C is independent of $t \in I$.
If $r = \infty$ and $p/(p-1) = d$ then μ is uniformly log-Lipschitz, that is,

$$W_p(\mu_{(t)}, \mu_{(\tau)}) \leq C|t-\tau| \left[\ln^{1/p}\left(\frac{1}{|t-\tau|}\right) + 1\right]$$

for some C independent of $t, \tau \in I$.

ii) *If $1 < p < d/(d-1)$ then there exists a Borel vector field $v \in \mathbb{L}^r\left(I, \mathbb{L}^p(\Omega; \mu_{(t)})\right)$ such that the continuity equation*

$$\partial_t\mu + \nabla_x \cdot (v\mu) = 0 \tag{19}$$

is satisfied in the sense of distributions in $I \times \Omega$.

3. Proofs for the case of a connected support

In this section we introduce the proofs of Theorems 1–2.

3.1. Proof of Theorem 1

Proposition 3.1. *Suppose μ is strongly d-connected ($d \geq 1$) and $x_0, x_1 \in \text{supp}(\mu)$. Then there exists $C = C(\mu)$ depending only on μ such that*

$$\sup_{\epsilon>0} \epsilon^{-q} W_p\left(\mu+\epsilon\delta_{x_0}, \mu+\epsilon\delta_{x_1}\right) < C(\mu) \tag{20}$$

where $q=\min(1, 1/d+1/p)$ *provided* $p \neq d/(d-1)$.

In the critical case $p=d/(d-1)$ *(where* $q=1$*)*

$$\sup_{\epsilon>0} \frac{1}{\epsilon \ln^{1/p}(1/\epsilon+1)} W_p\left(\mu+\epsilon\delta_{x_0}, \mu+\epsilon\delta_{x_1}\right) < C(\mu) . \tag{21}$$

Lemma 3.1. *Proposition 3.1 and Theorem 1 are equivalent.*

Proof. Obviously Theorem 1 implies Proposition 3.1. To see the opposite direction recall (see, e.g. [10])

$$W_p^p(\mu_1, \mu_2) = \sup_{(\phi,\psi)\in\mathcal{C}_p(\Omega)} \langle \mu_1, \phi\rangle - \langle \mu_2, \psi\rangle \tag{22}$$

where

$$\mathcal{C}_p(\Omega) := \left\{(\phi, \psi) \in C(\Omega)\times C(\Omega);\ \phi(x)-\psi(y) \le |x-y|^p \quad \forall (x,y)\in\Omega\times\Omega\right\} \tag{23}$$

Without limiting the generality we may assume $|\nu_+|=|\nu_-|=1$). Let $\delta>0$ and $(\bar\phi_\delta, \bar\psi_\delta) \in \mathcal{C}_p(\Omega)$ such that

$$W_p^p(\mu_1,\mu_2) \le \langle \mu_1, \bar\phi_\delta\rangle - \langle \mu_2, \bar\psi_\delta\rangle + \delta$$

where $\mu_1=\mu+\epsilon\nu_+$, $\mu_2=\mu+\epsilon\nu_-$. Let x_0 be a maximizer of $\bar\phi_\delta$ and x_1 a minimizer of $\bar\psi_\delta$. Then

$$\begin{aligned} W_p^p(\mu+\epsilon\nu_+, \mu+\epsilon\nu_-) &\le \langle \mu+\epsilon\nu_+, \bar\phi_\delta\rangle - \langle \mu+\epsilon\nu_-, \bar\psi_\delta\rangle + \delta \\ &\le \langle \mu+\epsilon\delta_{x_0}, \bar\phi_\delta\rangle - \langle \mu+\epsilon\delta_{x_1}, \bar\psi_\delta\rangle + \delta \le W_p^p(\mu+\epsilon\delta_{x_0}, \mu+\epsilon\delta_{x_1}) + \delta . \end{aligned} \tag{24}$$

Since $\delta>0$ is arbitrary (and independent of ϵ) we obtain the result. □

The following result is very easy but useful. For completeness we introduce the proof:

Proposition 3.2. Principle of monotone additivity*: Let* $\mu_1, \mu_2, \lambda \in \mathcal{M}_+$, $|\mu_1|=|\mu_2|$. *Then* $W_p(\mu_1,\mu_2) \ge W_p(\mu_1+\lambda, \mu_2+\lambda)$.

Proof. Let $\delta>0$. By (22), (23) there exists $(\phi,\psi)\in\mathcal{C}_p$ for which

$$\begin{aligned} W_p^p(\mu_1+\lambda, \mu_2+\lambda) &\le \langle \mu_1+\lambda, \phi\rangle - \langle \mu_2+\lambda, \psi\rangle + \delta \\ &= \langle \mu_1, \phi\rangle - \langle \mu_2, \psi\rangle + \langle \lambda, \phi-\psi\rangle + \delta \le \langle \mu_1, \phi\rangle - \langle \mu_2, \psi\rangle + \delta \le W_p^p(\mu_1,\mu_2) + \delta . \end{aligned} \tag{25}$$

The first inequality follows from $\phi(x)-\psi(x)\le 0$ for any $x\in\Omega$ by (23). The third one from (22). Again, we obtain the desired result since $\delta>0$ is arbitrary. □

3.2. *Proof of Proposition 3.1*

To illustrate the proof we start by stating some simplifying assumptions:

Ω is one dimensional, e.g. $\Omega=[0,1]$, $x_0=0$, $x_1=1$ and

$$\mu(ds) = \frac{\rho(s)ds}{\int_0^1 \rho(t)dt} \quad \text{where } \rho(s) = Ks^{d-1}(1-s)^{d-1} . \tag{26}$$

For $\mu_1, \mu_2 \in \mathcal{M}_1[0,1]$, let $M_i(s) := \mu_i[0,s]$ be the cumulative distribution function (CDF) of μ_i for $i=1,2$ respectively. Let $S^{(i)}$ be the generalized inverses of M_i. Then (cf. Theorem 2.18 in [13] for the case $p=2$ and Remark 2.19 there for the general case)

$$W_p^p(\mu_1,\mu_2) = \int_0^1 |S^{(1)}(m) - S^{(2)}(m)|^p dm . \tag{27}$$

In our case M_1 is the CDF of $\mu+\epsilon\delta_0$ while M_2 the CDF of $\mu+\epsilon\delta_1$. Setting $M=M(s)$ the CDF of μ and $S=S(m)$ its generalized inverse, then $M_1(s)=M(s)+\epsilon$ on $(0,1]$ and $M_2(s)=M(s)$ on $s\in[0,1)$, $M_2(1)=1+\epsilon$. The corresponding inverses are

i) $S^{(1)}(m)=0$ for $m\in[0,\epsilon]$, $S^{(1)}(m)=S(m-\epsilon)$ for $\epsilon\le m\le 1+\epsilon$.
ii) $S^{(2)}(m)=S(m)$ for $m\in[0,1]$ and $S^{(2)}(m)=1$ for $m\in[1,1+\epsilon]$.

Then (27) implies $W_p^p(\mu+\epsilon\delta_0,\mu+\epsilon\delta_1)=$

$$\int_0^\epsilon |S(m)|^p dm+\int_\epsilon^1 |S(m)-S(m-\epsilon)|^p dm+\int_1^{1+\epsilon}|S(m-\epsilon)-1|^p dm\ . \tag{28}$$

Since S is monotone non decreasing:

$$\int_0^\epsilon |S(m)|^p dm+\int_1^{1+\epsilon}|S(m-\epsilon)-1|^p dm\le\epsilon\left[S^p(\epsilon)+|1-S(1-\epsilon)|^p\right] \tag{29}$$

while

$$\int_\epsilon^1 |S(m)-S(m-\epsilon)|^p dm=\epsilon^p\int_\epsilon^{1-\epsilon}\left|\frac{dS}{dm}\right|^p dm\,(1+o(1))\ . \tag{30}$$

By the simplifying assumptions (26)

$$\kappa_1 s^d\le M(s)\le\kappa_2 s^d\quad,\quad \kappa_1(1-s)^d\le 1-M(s)\le\kappa_2(1-s)^d$$

for some $0<\kappa_1<\kappa_2$ where $s\in[0,1]$. Hence

$$\kappa_2^{-1/d}m^{1/d}\le S(m)\le\kappa_1^{-1/d}m^{1/d}\quad,\quad \kappa_2^{-1/d}(1-m)^{1/d}\le 1-S(m)\le\kappa_1^{-1/d}(1-m)^{1/d}$$

for $m\in[0,1]$. From this and $S'(m):=dS/dm=1/\rho(S(m))$

$$S'(m)=\frac{1}{\rho(S(m))}\le\kappa\min\{m^{1/d-1},(1-m)^{1/d-1}\}$$

for some $\kappa>0$ and $m\in[0,1]$. It follows from (27)–(30) that

i) If $p<d/(d-1)$ then $W_p^p(\mu+\epsilon\delta_0,\mu+\epsilon\delta_1)\le O(\epsilon^p)$.
ii) if $p=d/(d-1)$ then $W_p^p(\mu+\epsilon\delta_0,\mu+\epsilon\delta_1)\le O\left(\epsilon^p\ln(1/\epsilon+1)\right)$.
iii) if $p>d/(d-1)$ then $W_p^p(\mu+\epsilon\delta_0,\mu+\epsilon\delta_1)\le O\left(\epsilon^{p/d+1}\right)$.

In the general case, we provide the estimate (i–iii) for $W_p^p(\mu+\epsilon\delta_{y_j},\mu+\epsilon\delta_{y_{j+1}})$ for $j=1,\dots k-1$ (see Definition 2.2). Indeed, since W_p is a metric we get by the triangle inequality

$$W_p^p(\mu+\epsilon\delta_{x_0},\mu+\epsilon\delta_{x_1})\le\left(\sum_{j=1}^k W_p(\mu+\epsilon\delta_{y_j},\mu+\epsilon\delta_{y_{j+1}})\right)^p\ .$$

Consider J, Φ_j as in Definition 2.2. We my replace μ by the measure $\hat\mu:=K\Phi_{j,\#}\rho$. Indeed, by assumption, $\hat\mu\le\mu$ and the inequality

$$W_p(\mu+\epsilon\delta_{y_j},\mu+\epsilon\delta_{y_{j+1}})\le W_p(\hat\mu+\epsilon\delta_{y_j},\hat\mu+\epsilon\delta_{y_{j+1}}) \tag{31}$$

is evident by monotone additivity (Proposition 3.2).

Let (X_ϵ,σ) be a reference measure space such that $\int_{X_\epsilon}d\sigma=\epsilon+\int_\Omega d\hat\mu$. If $T^{(i)}:X_\epsilon\to\Omega$, $i=1,2$, is a pair of Borel mappings such that $T^{(1)}_\#\sigma=\hat\mu+\epsilon\delta_{y_j}$, $T^{(2)}_\#\sigma=\hat\mu+\epsilon\delta_{y_{j+1}}$ then

$$W_p^p(\hat{\mu}+\epsilon\delta_{y_j}, \hat{\mu}+\epsilon\delta_{y_{j+1}}) \leq \int_{X_\epsilon} \left|T^{(1)}(x)-T^{(2)}(x)\right|^p \sigma(dx) . \quad (32)$$

We now construct (X_ϵ, σ) as follows:

Let $M = M(s)$ be the CDF of ρ (cf. (26)). Set $\bar{M} := M(1)$. Then

$$X_\epsilon := \left\{(m, \beta) \in [0, \bar{M}+\epsilon] \times D\right\}$$

and σ is a multiple $dmd\beta$ on X_ϵ, normalized according to $\int_{X_\epsilon} d\sigma = \int_\Omega d\hat{\mu} + \epsilon$.

Let $S = S(m)$ the generalized inverse of $M = M(s)$, and extend it to X_ϵ by $S(m, \beta) = S(m)$. In analogy with one-dimensional case above, set

i) $S^{(1)}(m, \beta) = 0$ for $m \in [0, \epsilon]$, $S^{(1)}(m, \beta) = S(m-\epsilon, \beta)$ for $\epsilon \leq m \leq \bar{M}+\epsilon$.
ii) $S^{(2)}(m, \beta) = S(m, \beta)$ for $m \in [0, \bar{M}]$ and $S^{(2)}(m, \beta) = \bar{M}$ for $m \in [\bar{M}, \bar{M}+\epsilon]$.

By construction, $S^{(i)} : X_\epsilon \to J$ satisfy

$$S_\#^{(1)}\sigma = \rho ds + \epsilon\delta_{s=0} d\beta \ ; \quad S_\#^{(2)}\sigma = \rho ds + \epsilon\delta_{s=1} d\beta \ .$$

From Definition 2.2 it follows that $T^{(1,2)} := \Phi_j \circ S^{(1,2)}$ satisfy $T_\#^{(1)}\sigma = \tilde{\mu} + \epsilon\delta_{y_j}$, $T_\#^{(2)}\sigma = \tilde{\mu}+\epsilon\delta_{y_{j+1}}$. Then Definition 2.2-(i') yields

$$\int_{X_\epsilon} \left|T^{(1)}(m,\beta)-T^{(2)}(m,\beta)\right|^p dmd\beta \leq L^p \int_{X_\epsilon} \left|S^{(1)}(m,\beta)-S^{(2)}(m,\beta)\right|^p dmd\beta \ .$$

We now proceed as in the one-dimensional case to obtain the proof by (31), (32) via (28)–(30), in the general case.

3.3. *Proof of Theorem 2*

Proposition 3.3. *Suppose $\mu \in \mathcal{M}_+$, $\nu \in \mathcal{M}_0$ and $\mu+\nu \in \mathcal{M}_+$. Under the assumptions of Theorem 1, there exists $\bar{C} = \bar{C}(\mu)$ for which*

$$W_p(\mu+\nu, \mu) < \bar{C}\|\nu\|_{TV}^q \quad (33)$$

where $q = \min(1, 1/d+1/p)$ provided $p \neq d/(d-1)$.

In the critical case $p = d/(d-1)$ (where $q = 1$)

$$W_p(\mu+\nu, \mu) < \bar{C}\|\nu\|_{TV} \ln\left(\|\nu\|_{TV}^{-1}+1\right) \ .$$

For the proof of this proposition we need the following auxiliary lemma

Lemma 3.2. *Suppose $\mu, \nu_- \in \mathcal{M}_+$, μ is d-connected and $\nu_- \leq \mu$. Then there exists $\tilde{\nu} \in \mathcal{M}_+$ such that $\tilde{\nu}-\nu_- \in \mathcal{M}_0$, $\tilde{\nu} \leq \mu/2$, $\tilde{\nu}+\nu_- \leq \mu$ and a constant $\hat{C}(\mu)$ such that*

$$W_p(\mu-\nu_-, \mu-\tilde{\nu}) < \hat{C}(\mu)|\nu_-|^q$$

with $q = \min\{1, 1/p+1/d\}$.

Proof. Given $\epsilon_0 > 0$ it is enough to prove it for any $|\nu_-| < \epsilon_0$. So, let $|\nu_-| = \epsilon < \epsilon_0$. Let $\beta > 0$ large enough (independent of ϵ). For any such ϵ we divide the domain $\mathrm{supp}(\mu)$ into essentially disjoint, measurable cells $U_i \subset \Omega$ such that $\cup U_i \supset \mathrm{supp}(\mu)$, $U_i \cap U_j = \emptyset$ where $i \neq j$, and such that

i) Each cell contains a ball of radius $r_\epsilon := (4/K)^{1/d}\epsilon^{1/d}$ whose center is in $\mathrm{supp}(\mu)$. Here K is given by Definition 2.2.
ii) Each cell is contained in a concentric ball of radius βr_ϵ.

The existence of such a division can easily be demonstrated by tilling a neighborhood of supp(μ) by, say, identical boxes. The constant β depends *only* on the dimension of the embedding domain.

Let ν_i be the restriction of ν_- to U_i, $\alpha_i := |\nu_i|$, the mass of ν_- contained in U_i. By assumption, $\sum_i \alpha_i = \epsilon$.

By d-connectedness (see Definition 2.1) and (i), $\mu(U_i) \geq 4\epsilon$ for any i. Let

$$V_i := \{x \in U_i;\ d\nu_i/d\mu \leq 1/2\}$$

where $d\nu_i/d\mu$ stands for the Radon–Nikodym derivative. (Note that $d\nu_i/d\mu \leq 1$ since $\nu_i \leq \nu_- \leq \mu$). Then

$$\epsilon \geq \alpha_i \geq \int_{U_i - V_i} (d\nu_i/d\mu) d\mu \geq \frac{1}{2}\mu(U_i - V_i)$$

hence $\mu(U_i - V_i) \leq 2\epsilon$, so $\mu(V_i) \geq 4\epsilon - 2\epsilon \geq 2\alpha_i$.

Let $\tilde{V}_i \subset V_i$, a measurable set such that $\mu(\tilde{V}_i) = 2\alpha_i$. Define $\tilde{\nu}_i$ as the restriction of $\mu/2$ to $\tilde{V}_i$. In particular, $|\tilde{\nu}_i| = \alpha_i$, and $\tilde{\nu}_i \leq \mu/2$.

Let now $\tilde{\nu} := \sum_i \tilde{\nu}_i$. Since the sets $\tilde{V}_i$ are mutually disjoint, $\tilde{\nu} \leq \mu/2$, i.e. $d\tilde{\nu}/d\mu \leq 1/2$ μ-a.e. Moreover, $d\tilde{\nu}/d\mu + d\nu_-/d\mu \leq 1$ μ-a.e., since $d\tilde{\nu}/d\mu = 0$ if $d\nu_-/d\mu > 1/2$ by construction while $d\nu_-/d\mu \leq 1$ by the assumption $\nu_- \leq \mu$. So $\nu_- + \tilde{\nu} \leq \mu$. Finally, $|\tilde{\nu}| = |\nu_-| = \epsilon$, so $\tilde{\nu} - \nu_- \in \mathcal{M}_0$.

Since the diameter of the set U_i is not larger than $2\beta r_\epsilon$ (cf. (ii)), the W_p^p cost for shifting a mass α_i within U_i is not larger that $\alpha_i(2\beta r_\epsilon)^p$. Hence

$$W_p^p(\tilde{\nu}_i, \nu_i) \leq \alpha_i(2\beta r_\epsilon)^p = \alpha_i(2\beta)^p \left(\frac{4}{K}\right)^{p/d} \epsilon^{p/d} \tag{34}$$

Recalling $\nu_- = \sum \nu_i$, $\tilde{\nu} = \sum \tilde{\nu}_i$ we get $W_p^p(\tilde{\nu}, \nu_-) \leq \sum_i W_p^p(\tilde{\nu}_i, \nu_i) \leq$

$$\sum_i (2\beta)^p \left(\frac{4}{K}\right)^{p/d} \alpha_i \epsilon^{p/d} = (2\beta)^p \left(\frac{4}{K}\right)^{p/d} \epsilon^{p/d+1} = (2\beta)^p \left(\frac{4}{K}\right)^{p/d} |\nu_-|^{p/d+1}$$

were we used $\sum \alpha_i = \epsilon = |\nu_-|$.

Let now $\lambda := \mu - \nu_- - \tilde{\nu} \geq 0$. Then

$$W_p(\mu - \nu_-, \mu - \tilde{\nu}) = W_p(\lambda + \tilde{\nu}, \lambda + \nu_-) \leq W_p(\tilde{\nu}, \nu_-) \leq (2\beta)\left(\frac{4}{K}\right)^{1/d} |\nu_-|^q$$

by Proposition 3.2. □

Proof of Proposition 3.3. Let $\nu = \nu_+ - \nu_-$. We may assume by the principle of monotone additivity that $\nu_\pm$ are the positive/negative parts of ν, i.e. $\|\nu\|_{TV} = |\nu_+| + |\nu_-|$. Let $\bar{\mu} := \mu + \nu_+$. Then, by the triangle inequality,

$$W_p(\mu + \nu, \mu) = W_p(\bar{\mu} - \nu_-, \bar{\mu} - \nu_+) \leq W_p(\bar{\mu} - \nu_-, \bar{\mu} - \tilde{\nu}) + W_p(\bar{\mu} - \tilde{\nu}, \bar{\mu} - \nu_+) \tag{35}$$

where $\tilde{\nu}$ is as in Lemma 3.2 (in particular, $\bar{\mu}$ majorizes $\tilde{\nu}$, as well as ν_-, ν_+). Since $\bar{\mu} \geq \mu$ we get by monotone additivity and Lemma 3.2

$$W_p(\bar{\mu} - \nu_-, \bar{\mu} - \tilde{\nu}) \leq W_p(\mu - \nu_-, \mu - \tilde{\nu}) \leq \hat{C}(\mu)|\nu_-|^q \equiv 2^{-q}\hat{C}(\mu)\|\nu\|_{TV}^q\ . \tag{36}$$

Setting $\tilde{\mu} = \mu - \tilde{\nu} := \bar{\mu} - \nu_+ - \tilde{\nu}$ we get

$$W_p(\bar{\mu} - \tilde{\nu}, \bar{\mu} - \nu_+) = W_p(\tilde{\mu} + \nu_+, \tilde{\mu} + \tilde{\nu})\ . \tag{37}$$

Now, $\tilde{\mu} \geq \mu/2$ by Lemma 3.2, and since $\|\nu_+ - \tilde{\nu}\|_{TV} \leq |\nu_+| + |\tilde{\nu}| = |\nu_+| + |\nu_-| = \|\nu\|_{TV}$, we obtain from Theorem 1, (37) and Proposition 3.2

$$W_p(\bar{\mu} - \tilde{\nu}, \bar{\mu} - \nu_+) \leq C(\mu/2) \left\{ \begin{array}{c} \|\nu\|_{TV}^q \text{ if } p \neq \frac{d}{d-1} \\ \|\nu\|_{TV}\left(\ln(\|\nu\|_{TV}^{-1} + 1\right) \text{ if } p = \frac{d}{d-1} \end{array} \right\}. \tag{38}$$

The proposition now follows from (35), (36), (38) where $\bar{C}(\mu) = 2^{-q}\hat{C}(\mu) + C(\mu/2)$. □

Proof of Theorem 2.

i) Given $t > \tau \in I$, let $\nu = \mu_{(t)} - \mu_{(\tau)}$. Note that $\text{supp}(\nu) \subseteq \text{supp}(\mu_{(\tau)})$. Since $\vec{\mu} \in AC^r(I, TV)$

$$\|\nu\|_{TV} \leq \int_\tau^t m \leq \|m\|_r |t-\tau|^{1-1/r} . \tag{39}$$

The assumptions of Theorem 1 are satisfied so we obtain the result by Proposition 3.3, upon estimating $\|\nu\|_{TV}$ by (39). □

ii) Since $1 < p < d/(d-1)$ we get $q = 1$ so (33), where $\mu := \mu_{(\tau)}$, $\nu = \mu_{(t)} - \mu_{(\tau)}$, with the first inequality in (39) imply

$$W_p(\mu_{(t)}, \mu_{(\tau)}) \leq C \int_\tau^t m .$$

Since $\mu_{(\cdot)} \in AC^r(I, TV)$ by the assumption, then $m \in \mathbb{L}^r$ so $\vec{\mu} \in AC^r(I, W_p)$ as well. The existence of a vector field satisfying (19) follows from Theorem 8.3.1 in [1] (see also Theorem 5 in [7]).

4. Optimal teleportation and disconnected support

In the case of disconnected support of μ we obtain the following result:

Assumption 4.1.

1. $\mu \in \mathcal{M}_1$ and $\text{supp}(\mu)$ is composed of a finite number ($m \geq 2$) of disjoint components $\mu = \sum_{j=1}^m \mu_j$ where $\text{supp}(\mu_i) \cap \text{supp}(\mu_j) = \emptyset$ for any $i \neq j$.
2. Each μ_i satisfies the assumptions of Theorem 1.
3. $\nu = \nu_+ - \nu_- \in \mathcal{M}_0$, $\text{supp}(\nu_+) \cup \text{supp}(\nu_-) \subset \text{supp}(\mu)$.

Definition 4.1. $A_i := \text{supp}(\mu_i)$ are the connected components of $\text{supp}(\mu)$.

i) $\bar{\nu}_j := \langle \nu, 1_{A_j} \rangle$. By Assumption 4.1-(3), $\sum_{j=1}^m \bar{\nu}_j = 0$.

ii) $V := \{1..., m\}$, $V_+ := \{j \in V; \bar{\nu}_j > 0\}$, $V_- := \{j \in V; \bar{\nu}_j < 0\}$.

iii) For $i, j \in V$, $|E|_{i,j} := dist^p(A_i, A_j) \equiv \min_{x \in A_i, y \in A_j} |x-y|^p$.

iv) $G := (V, E)$ is a complete graph (i.e. any two vertices are connected by an edge) whose vertices V and the length of the edge $E_{i,j}$ connecting i to j is $|E|_{i,j}$.

v) Let $\mathcal{O}_{i,j}$ is the set of all orbits in V connecting i to j, that is, $o_{i,j} \in \mathcal{O}_{i,j}$ if

$$o_{i,j} = \{o_{i,j}^{(1)}, \dots o_{i,j}^{(n)}\} \subset V$$

such that $o_{i,j}^{(1)} = i$, $o_{i,j}^{(n)} = j$. The length of such an orbit is $|o_{i,j}| = n$ in that case.
Given $i, j \in V$, $d(i, j)$ is the geodesic distance corresponding to (V, E). That is:

$$d(i,j) := \min_{o_{i,j} \in \mathcal{O}_{i,j}} \sum_{l=1}^{|o_{i,j}|-1} |E|_{o_{i,j}^{(l)}, o_{i,j}^{(l+1)}} \tag{40}$$

See Fig. 2 for an illustration.

vi) Let now $\bar{\nu}_i > 0$ be the charge associated with the vertex $i \in V_+$, and $-\bar{\nu}_j > 0$ the charge associated with $j \in V_-$. Let $\|\nu\|_\mu$ be the optimal cost of transportation of $\sum_{i \in V_+} \bar{\nu}_i \delta_i$ to $\sum_{j \in V_-} (-\bar{\nu}_j) \delta_j$ subjected to the graph metric $d(i, j)$. That is:

$$\|\nu\|_\mu := \min_{\lambda \in \lambda(\nu)} \sum_{i \in V_+} \sum_{j \in V_-} \lambda_{i,j} d(i,j) := \sum_{i \in V_+} \sum_{j \in V_-} \lambda^*_{i,j} |d(i,j) \tag{41}$$

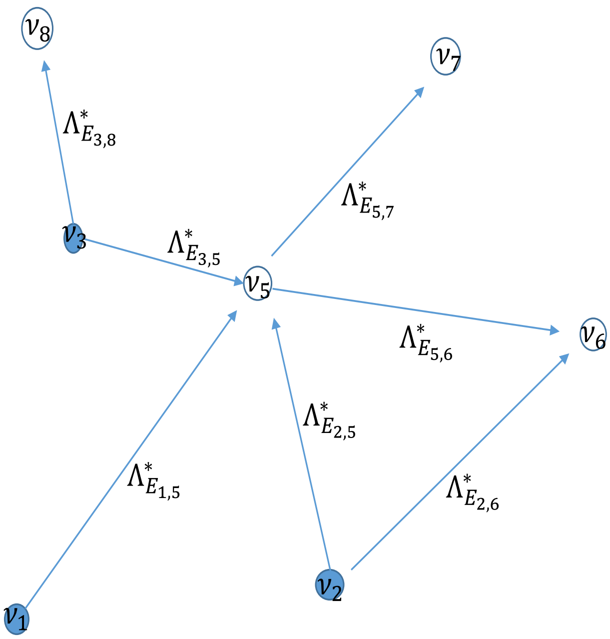

Fig. 2. Transfer plan via a directed graph. Sources ($\bar{\nu}_i > 0$) are filled circles while sinks ($\bar{\nu}_i < 0$) are empty circles. (cf. Definition 4.1-(i)). Geodesic arcs: $(1 \mapsto 5) = (1,5)$; $(1 \mapsto 6) = (1,5,6)$, $(1 \mapsto 7) = (1,5,7)$, $(1 \mapsto 8) = (1,5,3,8)$, $(2 \mapsto 5) = (2,5)$, $(2 \mapsto 6) = (2,6)$, $(2 \mapsto 7) = (2,5,7)$, $(2 \mapsto 8) = (2,5,3,8)$, $(3 \mapsto 5) = (3,5)$, $(3 \mapsto 6) = (3,5,6)$, $(3 \mapsto 7) = (3,5,7)$, $(3 \mapsto 8) = (3,8)$; Weighed arcs: $\Lambda^*_{E_{1,5}} = \lambda^*_{1,5} + \lambda^*_{1,6} + \lambda^*_{1,7}$, $\lambda^*_{E_{2,5}} = \lambda^*_{2,5} + \lambda^*_{2,7}$, $\Lambda^*_{E_{2,6}} = \lambda^*_{2,6}$, $\Lambda^*_{E_{5,6}} = \lambda^*_{3,6} + \lambda^*_{1,6}$, $\Lambda^*_{E_{5,7}} = \lambda^*_{1,7} + \lambda^*_{2,7} + \lambda^*_{3,7}$, $\Lambda^*_{E_{3,5}} = \lambda^*_{3,5} + \lambda^*_{3,6} + \lambda^*_{3,7}$. It is assumed that $\bar{\nu}_3$ is large enough to supply $\bar{\nu}_8$, so $\lambda^*_{1,8} = \lambda^*_{2,8} = 0$. Otherwise, the arrow $E_{3,5}$ should be reversed, and $\Lambda^*_{E_{1,5}} = \lambda^*_{1,5} + \lambda^*_{1,6} + \lambda^*_{1,7} + \lambda^*_{1,8}$, $\lambda^*_{E_{2,5}} = \lambda^*_{2,5} + \lambda^*_{2,7} + \lambda^*_{2,8}$, and $\Lambda^*_{E_{5,3}} = \lambda^*_{1,8} + \lambda^*_{2,8}$.

where $\lambda(\nu)$ is the set of non-negative $|V_+| \times |V_-|$ matrices $\{\lambda_{i,j}\}$ which satisfy:
$\sum_{j \in V_-} \lambda_{i,j} = \bar{\nu}_i$ if $i \in V_+$
$\sum_{i \in V_+} \lambda_{i,j} = -\bar{\nu}_j$ if $j \in V_-$.

Theorem 3. *If $\infty > p > 1$ and $\mu, \nu := \nu_+ - \nu_-$ satisfy Assumption 4.1 then*

$$\lim_{\epsilon \searrow 0} \epsilon^{-1/p} W_p(\mu, \mu + \epsilon\nu) = \|\nu\|_\mu^{1/p}$$

4.1. Proof of Theorem 3

We first state the inequality

$$\liminf_{\epsilon \searrow 0} \epsilon^{-1/p} W_p(\mu, \mu + \epsilon\nu) \geq \|\nu\|_\mu^{1/p} .$$

From the principle of monotone additivity it is enough to prove

$$\liminf_{\epsilon \searrow 0} \epsilon^{-1/p} W_p(\mu + \epsilon\nu_+, \mu + \epsilon\nu_-) \geq \|\nu\|_\mu^{1/p} . \tag{42}$$

Recall the dual formulation (22), (23). In fact, it is enough to restrict to $(\phi, \psi) \in \mathcal{C}_p(\text{supp}(\mu)) \equiv \mathcal{C}_p(\cup A_i)$. In the special case $\psi(x) = \phi(x) := z_i$ is a constant over A_i we get

$$W_p^p(\mu + \epsilon\nu_+, \mu + \epsilon\nu_-) \geq \epsilon \sum_{i \in \bar{V}} z_i \bar{\nu}_i \tag{43}$$

provided $z_i - z_j \le |x-y|^p$ for any $x \in A_i$, $y \in A_j$. In particular, if $z_i - z_j \le d(i,j)$ (see Definition 4.1-(iii, v)). From (43) and Definition 4.1-(ii) we get

$$W_p^p(\mu+\epsilon\nu_+,\mu+\epsilon\nu_-) \ge \epsilon \sup_{\{z\}} \sum_{i\in\bar{V}} z_i \bar{\nu}_i \tag{44}$$

where the supremum is on all possible values of $\{z_1 \dots, z_{\#\bar{V}}\}$ which satisfy $z_i - z_j \le d(i,j)$ for any $i, j \in \bar{V}$. Since $d(\cdot,\cdot)$ is a metric on the graph (V,E) via Definition 4.1-(v) we recall the dual formulation of the metric Monge problem, or the so called Kantorovich–Rubinstein Theorem (Theorem 1.14 in [13]) in discrete version:

$$\|\nu\|_\mu = \sup_{\{z\}} \sum_{i\in V} z_i \bar{\nu}_i \quad ; \quad z_i - z_j \le d(i,j) \tag{45}$$

(see also Definition 4.1-(vi)). Then (42) follows from (44)–(45).

To prove the opposite inequality we need some additional definitions:

Definition 4.2.

1. Denote $Z_i^j \in A_i$ to be the closest point in A_i to A_j (see Definition 4.1-(iii)).
2. For $i, j \in V$ let $\bar{o}_{i,j} = (o_{i,j}^{(1)} \dots o_{i,j}^{(n)})$, a choice of an optimal orbit realizing (40) in Definition 4.1-(v) (note that there can be more than one such orbit, but we choose only one). Let $|\bar{o}_{i,j}|$ be the cardinality of $\bar{o}_{i,j}$.
   For any $l \in V$, denote E_l^+ the set if all outgoing edges from l, that is, $E \in E_l^+$ iff, for some $i, j \in V$, $l \in \bar{o}_{i,j} = (o_{i,j}^{(1)} \dots o_{i,j}^{(n)})$, $l \neq o_{i,j}^{(n)}$.
   Likewise, denote E_l^- the set if all incoming edges to l, that is $E \in E_l^-$ iff $l \in \bar{o}_{i,j} = (o_{i,j}^{(1)} \dots o_{i,j}^{(n)})$, $l \neq o_{i,j}^{(1)}$.
3. For each $i, j \in V$, let

   $$E_{\bar{o}_{i,j}} := \{E;\, E = E_{o_{i,j}^{(k)}, o_{i,j}^{(k+1)}};\ 1 \le k \le |\bar{o}_{i,j}| - 1\}\ ,$$

   where $\bar{o}_{i,j}$ is the above choice of optimal orbit. Let

   $$\Lambda_E^* := \sum_{\{i,j\ ; E\in E_{\bar{o}_{i,j}}\}} \lambda_{i,j}^*\ , \tag{46}$$

   see (41) for $\lambda_{i,j}^*$. This is *the total flux* traversing E due to the optimal transport plan.
   Note that

   $$\sum_{E\in E_l^+} \Lambda_E^* - \sum_{E\in E_l^-} \Lambda_E^* = \bar{\nu}_l \tag{47}$$

   for any $l \in V$. Recall (Definition 4.1 (i, ii)) that $\bar{\nu}_l > 0$ if $l \in V_+$, $\bar{\nu}_l < 0$ if $l \in V_-$ and $\bar{\nu}_l = 0$ if $l \in V - \bar{V}$.
   Note also that the flux due to optimal plan is uni-directional, i.e. $\Lambda_E^* \cdot \Lambda_{-E}^* = 0$ for any edge E (here $-E$ represents the same edge in the opposite orientation).
4. For $k \in V$

   $$\hat{\nu}_k^+ := \sum_{E_{k,i}\in E_k^+} \Lambda_{E_{k,i}}^* \delta_{Z_k^i}\ . \tag{48}$$

   Here δ_x is the Dirac delta function at x. In particular, $\hat{\nu}_k^+$ is supported in A_k (see Definition 4.2 (1)), and

   $$|\hat{\nu}_k^+| = \sum_{E\in E_k^+} \Lambda_E^*\ . \tag{49}$$

5. For $i, j \in V$, let $B_r(Z_i^j)$ be the ball of radius r centered at $Z_i^j \in A_i$. Given $\epsilon > 0$ let $r_i^{j,\epsilon} > 0$ be the radius of the ball such that $\mu\left(A_j \cap B_{r_i^{j,\epsilon}}(Z_i^j)\right) = \epsilon$. See Fig. 3.
   Let $\hat{\mu}_{i,j}^\epsilon$ be the restriction of the measure μ to the set $A_j \cap B_{r_i^{j,\epsilon}}(Z_i^j)$ defined above.

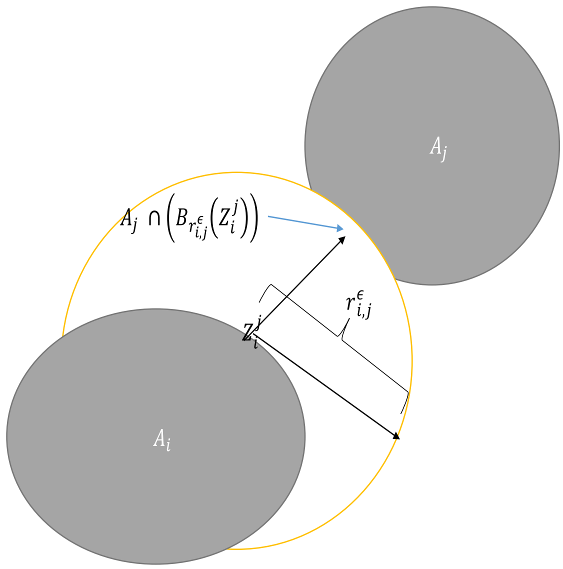

Fig. 3. Locating the support of $\hat{\mu}^\epsilon_{i,j}$.

6. Let

$$\hat{\nu}_k^-(\epsilon) := \sum_{E=E_{l,k}\in E_k^-} \hat{\mu}_{l,k}^{\epsilon\Lambda_E^*} . \tag{50}$$

In particular, $\hat{\nu}_k^-$ is supported in A_k and

$$|\hat{\nu}_k^-(\epsilon)| = \epsilon \sum_{E\in E_k^-} \Lambda_E^* . \tag{51}$$

7. $\hat{\nu}_+ := \sum_{k\in V} \hat{\nu}_k^+$; $\hat{\nu}_-(\epsilon) := \sum_{k\in V} \hat{\nu}_k^-(\epsilon)$; $\hat{\nu}(\epsilon) := \epsilon\hat{\nu}_+ - \hat{\nu}_-(\epsilon)$.

Note that $\hat{\nu}(\epsilon) \in \mathcal{M}_0$, i.e. $\epsilon|\hat{\nu}_+| = |\hat{\nu}_-(\epsilon)|$. In fact, we obtain from (47), (49, (51) that for each $k \in V$

$$\epsilon|\hat{\nu}_k^+| - |\hat{\nu}_k^-(\epsilon)| = \epsilon\bar{\nu}_k , \tag{52}$$

and $\sum_{k\in V} \bar{\nu}_k = 0$ (Definition 4.1-(i)).

Using the above we find form the metric property of W_p and the triangle inequality

$$W_p(\mu + \epsilon\nu, \mu) \le W_p(\mu + \epsilon\nu, \mu + \hat{\nu}(\epsilon)) + W_p(\mu, \mu + \hat{\nu}(\epsilon)) . \tag{53}$$

Let μ_k be the restriction of μ to A_k, ν_k the restriction of ν to A_k and $\hat{\nu}_k(\epsilon) = \epsilon\hat{\nu}_k^+ - \hat{\nu}_k^-(\epsilon)$. By (52) (recall $\bar{\nu}_k := |\nu_k|$), $W_p(\mu_k + \epsilon\nu_k, \mu_k + \hat{\nu}_k(\epsilon))$ is defined on each component. We can use the definition of Wasserstein metric to obtain

$$W_p^p(\mu + \epsilon\nu, \mu + \hat{\nu}(\epsilon)) \le \sum_{k\in V} W_p^p(\mu_k + \epsilon\nu_k, \mu_k + \hat{\nu}_k(\epsilon)) .$$

Now Theorem 1 applies to each of the components of this sum. By the assumption of the Theorem we obtain

$$W_p^p(\mu_k + \epsilon\nu_+^k, \mu_k + \epsilon\hat{\nu}_k^+) = O(\epsilon^{pq}) = o(\epsilon)$$

where $q > 1/p$ by its definition. Thus, the first term on the right of (53) is controlled by $o(\epsilon^{1/p})$.

To complete the proof we need to estimate the second term.

Proposition 4.1.

$$W_p^p(\mu, \mu + \hat{\nu}(\epsilon)) \le \epsilon\|\nu\|_\mu + o(\epsilon) .$$

For the proof we construct a transport plan π from μ to $\mu+\hat{\nu}(\epsilon)$. To illustrate this construction by a particular example see the directed tree in Fig. 2. A detailed description of the plan is given below.

For any positive measure $\sigma\in\mathcal{M}_+(\Omega)$ and $x\in\Omega$ define $\delta_x\otimes\sigma\in\mathcal{M}_+(\Omega\times\Omega)$ by its action on $\phi\in C(\Omega\times\Omega)$:

$$<\delta_x\otimes\sigma,\phi>:=\int_\Omega\phi(x,y)d\sigma(y)\ .$$

Let

$$\mu_-^\epsilon:=\mu-\hat{\nu}_-(\epsilon) \tag{54}$$

and $\pi_{\mu_-^\epsilon}$ be the diagonal lift of μ_-^ϵ to $\mathcal{M}_+(\Omega\times\Omega)$, that is,

$$<\pi_{\mu_-^\epsilon},\phi>:=\int_\Omega\phi(x,x)d\mu_-^\epsilon(x)\ .$$

Let now

$$\pi^\epsilon:=\pi_{\mu_-^\epsilon}+\sum_{l\in V}\sum_{k\in V}\delta_{Z_l^k}\otimes\hat{\mu}_{l,k}^{\epsilon\Lambda^*_{E_{l,k}}}. \tag{55}$$

Note that some terms in the double sum above my be zero. This is the case if the edge $E_{l,k}$ does not transverse an orbit of the optimal transport plan, i.e. $\Lambda^*_{E_{l,k}}=0$ (hence $\delta_{Z_l^k}\otimes\hat{\mu}_{l,k}^{\epsilon\Lambda^*_{E_{l,k}}}=0$).

Next, observe that $\pi^\epsilon\in\Pi(\mu+\hat{\nu}(\epsilon),\mu)$ (cf. (2)). In fact, from (54) and (55), for any $\phi=1(y)\psi(x)$

$$<\pi^\epsilon,\phi>=\int_\Omega\psi(x)d\mu_-^\epsilon(x)+\epsilon\sum_{l\in V}\sum_{k\in V}\psi(Z_k^l)\Lambda^*_{E(k,l)}=$$
$$\int_\Omega\psi(x)d\mu(x)-\int_\Omega\psi(x)d\hat{\nu}_-(\epsilon)(x)+\epsilon\int_\Omega\psi(x)d\hat{\nu}_+(x)=<\mu+\hat{\nu}(\epsilon),\psi> \tag{56}$$

where we used $\hat{\nu}_+:=\sum_{k\in V}\hat{\nu}_k^+$ and (48).

Setting now $\phi=1(x)\psi(y)$

$$<\ \pi^\epsilon,\phi\ >=\int_\Omega\psi(y)d\mu_-^\epsilon(y)\ +\ \sum_{l\in V}\sum_{k\in V}\int_\Omega\psi(y)d\hat{\mu}_{l,k}^{\epsilon\Lambda^*_{E_{l,k}}}(y)\ =<\ \mu,\psi\ > \tag{57}$$

where we used $\hat{\nu}_-(\epsilon):=\sum_{k\in V}\hat{\nu}_k^-(\epsilon)$ and (54), (50).

It then follows from (55) that

$$W_p^p(\mu,\mu+\hat{\nu}(\epsilon))\le\int_\Omega\int_\Omega|x-y|^pd\pi_\epsilon=\sum_{l\in V}\sum_{E=E_{l,k}}\int_\Omega\int_\Omega|Z_l^k-y|^p\hat{\mu}_{l,k}^{\epsilon\Lambda^*_E}(dy) \tag{58}$$

From Definition 4.2-1, 5 and Definition 4.1-(iii), we obtain $\int_\Omega|Z_l^k-y|^p\hat{\mu}_{l,k}^{\epsilon\Lambda^*_{E_{l,k}}}(dy)=\epsilon|E|_{l,k}\Lambda^*_E+o(\epsilon)$, so Definition 4.2-6, 7, together with (46) imply

$$\int_\Omega\int_\Omega|Z_l^k-y|^p\hat{\mu}_{l,k}^{\epsilon\lambda^*_E}(dy)=\epsilon|E|_{l,k}\sum_{(i,j);E_{l,k}\in\bar{e}_{i,j}}\lambda^*_{i,j}+o(\epsilon) \tag{59}$$

and (58), (59), (40) imply

$$W_p^p(\mu,\mu+\hat{\nu}(\epsilon))\le\epsilon\sum_{i,j\in V\times V}\lambda^*_{i,j}d(i,j)+o(\epsilon)=\epsilon\|\nu\|_\nu+o(\epsilon)\ .\quad\square$$

Conflict of interest statement

There is no conflict of interests.

Acknowledgement

This research was supported by the Israel Science Foundation (grant No. 988/15).

References

[1] L. Ambrosio, N. Gigli, G. Savaŕe, Gradient Flows in Metric Spaces and in the Space of Probability Measures, Lecture Notes in Mathematics, Birkhäuser, 2005.
[2] L. Ambrosio, N. Gigli, A user's guide to optimal transport, in: B. Piccoli, M. Rascle (Eds.), Modelling and Optimisation of Flows on Networks Cetraro, Italy, 2009, Springer, 2012.
[3] J.D. Benamou, Y. Brenier, A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem, Numer. Math. 84 (2000) 375–393.
[4] L.V. Kantorovich, On the translocation of masses, Dokl. Akad. Nauk. USSR 37 (1942) 199–201, English translation in J. Math. Sci. 133 (4) (2006) 1381–1382.
[5] L.V. Kantorovich, G.S. Rubinshtein, On a space of totally additive functions, Vestn. Leningr. Univ. 13 (7) (1958) 52, 59.
[6] E. Järvenpää, M. Järvenpää, A. Käenmäki, T. Rajala, S. Rogovin, V. Suomala, Packing dimension and Ahlfors regularity of porous sets in metric spaces, Math. Z. 266 (1) (September 2010) 83–105.
[7] S. Lisini, Characterization of absolutely continuous curves in Wasserstein spaces, Calc. Var. Partial Differ. Equ. 28 (2007) 85–120.
[8] R.J. McCann, A convexity principle for interacting gases, Adv. Math. 128 (1) (1997) 153–179.
[9] F. Otto, The geometry of dissipative evolution equations: the porous medium equation, Commun. Partial Differ. Equ. 26 (1–2) (2001) 101,174.
[10] S.T. Rachev, L.R. Rüschendorf, Mass Transportation Problems, vol. 1, Springer, 1998.
[11] H. Tanaka, An inequality for a functional of probability distributions and its application to Kac's one-dimensional model of a Maxwellian gas, Z. Wahrscheinlichkeitstheor. Verw. Geb. 27 (1973) 47, 52.
[12] L.N. Vasershtein, Markov processes over denumerable products of spaces describing large system of automata, Probl. Pereda. Inf. 5 (3) (1969) 64–72.
[13] C. Villani, Topics in Optimal Transportation, Graduate Studies in Mathematics, vol. 58, AMS, 2003.
[14] C. Villani, Optimal Transport: Old and New, Springer, 2009.
[15] A.M. Vershik, The Kantorovich metric: the initial history and little-known applications, Zap. Nauč. Semin. POMI 312 (2004) 69–85, Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 11 (2004) 311.
[16] G. Wolansky, Limit theorems for optimal mass transportation, Calc. Var. Partial Differ. Equ. 42 (3–4) (2011) 487–516
.