arXiv:math-ph/0306070v3 9 Jul 2003

Mass Transportation, the Monge-Kantorovich problem, Hamilton-Jacobi equation and metrics on measure-valued functions

G. Wolansky[1]

April 11, 2008

1 Introduction

Consider a cloud of particles (say, a smoke cloud) drifted in the atmosphere. We can measure the density of the cloud at any instant of time we wish. What can we say about its velocity?

This question looks rather naive, since velocity is just the rate of change of position. Indeed, if the cloud is composed of a single (or few) particles then we can easily identify two close particles at times t and $t+\Delta T$, respectively, for ΔT small enough, as a single particle at two different positions $x(t)$ and $x(t+\Delta T)$ and approximate its velocity pretty well. However, if the cloud is composed of an astronomical number of particles, or if it is described as a continuous distribution $\rho = \rho(x,t)$, then we must make an assumption in order to be able to say something about its velocity field. Such an assumption can be, for example, that the cloud has the minimal possible kinetic energy, or, more generally, a minimum of some other moment of its velocity field.

Let us look, again, at our cloud at times t and $t + \Delta T$. How can we identify a pair of particles at these two instances as identical? This problem is known as "mass transportation", or the Monge problem.

The classical problem of mass transportation was suggested by Monge in the 18'th century [M]. Given a pair of Borel probability measures μ_0, μ_1 on a common probability space Ω, one looks at the problem of minimizing

$$\int c(|T(x) - x|)\mu_0(dx) \quad ; \quad T_\# \mu_0 = \mu_1 \tag{\textbf{M}}$$

along all mappings $T : \Omega \to \Omega$ which transport μ_0 into μ_1 ($T_\#\mu_0 = \mu_1$). Here $c(|\cdot|)$ is a pre-determined cost function.

A breakthrough in the Monge problem was introduced by Kantorovich relaxation in 1942 [K]: Instead of looking on a family of Borel maps satisfying $T_\#\mu_0 = \mu_1$ we look at a family of probability measures λ on $\Omega \times \Omega$ whose marginals are μ_0, μ_1 respectively. The Kantorovich problem is formulated as

$$\min_\lambda \int\int c(|x-y|)\lambda(dx, dy) \quad ; \ \pi^{(0)}\lambda = \mu_0 \ , \ \pi^{(1)}\lambda = \mu_1 \tag{\textbf{K}}$$

Here $\pi^{(i)}, i = 0, 1$ are the natural projections of $\Omega \times \Omega$ on its factors.

[1]Department of mathematics, Technion, Haifa 32000, Israel

The most striking advantage of the relaxed problem is that a minimizer always exists by the compactness of the set of probability measures (assuming Ω is compact). The "only" question left is the nature and uniqueness of this minimizer. In particular, is the minimizer supported on a graph of a measurable map T? If so, what is the nature of this map?

These questions (and related ones) were discussed in a countless number of publications in recent decades. Particular attention was given to the cost function $|x-y|^p$, leading to the Wasserstein metrics

$$d_p(\mu_1,\mu_2) = \left[\inf_\lambda \int\int |x-y|^p d\lambda\right]^{1/p}$$

with $p \geq 1$. There is also a dual formulation to the Wasserstein metric which, in the case $p=1$, takes the particular appealing form:

$$d_1(\mu_0,\mu_1) = \sup_{\phi\in C^1(\Omega);|\nabla\phi|=1} \int_\Omega \phi\left[\mu_1(dx) - \mu_0(dx)\right] \ . \tag{1.1}$$

Dual formulations exist also for $p>1$ (as well as for more general costs). However, their form is less elegant in the general case. For $p=2$, for example, the optimal solution is related to the dual problem

$$\max\left[\int_\Omega \phi\mu_0(dx) + \int_\Omega \psi\mu_1(dx) \quad ; \quad \phi(x)+\psi(y) \geq xy\right]$$

A remarkable result [B] is that the optimal solution of this dual problem (if exists) is given by a pair of convex-conjugate functions whose gradients yields the optimal Monge mapping $T=\nabla_x\phi$ and its inverse $T^{-1}=\nabla_x\psi$. In particular it was proved that, if μ_0 is continuous with respect to Lebesgue measure then such T exists and is unique. Moreover, any mapping which transport μ_0 to μ_1 is a composition $T\circ S$, where S is a mapping preserving μ_0. A generalization to this result also holds for more general cost functions, using special definitions of convexity [G.M].

At this point we should note that if T is a Borel map, then $T_\#\mu_0$ is not, in general, defined *unless* μ_0 is continuous with respect to Lebesgue measure. We may, of course, restrict the class of mappings (e.g. continuous maps), assuming that there exists a map T in this class for which $T_\#\mu_0=\mu_1$ and ask for the minimizer of (M) within this class. This is, of course, a legitimate problem which removes all limitation from μ_0. However, it pose some "topological" restrictions on μ_1.

In general, if μ_0 contains an atom, then there is no deterministic mapping T of any type which maps μ_0 into μ_1, so there is no sense to compare the Monge problem (M) with the Kantorovich problem (K) in that case.

One of the outcomes of this paper is to following paradigm, which offers a third interpretation to the deterministic (M) and probabilistic (K) mass transport: :

For given μ_0, μ_1 Borel probability measures in Ω, find a family of homomorphisms $T_{t_1}^{t_2}:\Omega\rightarrow\Omega$ for $0<t_1,t_2,<1$ and an orbit of probability measures μ_t on Ω for each $t\in[0,1]$ such that

- *$T^{t_2}_{t_1,\#}\mu_{t_1}=\mu_{t_2}$ for any $t_1,t_2\in(0,1)$. In addition, $T_{t_1}^{t_2}$ is an optimal Monge map from μ_{t_1} to μ_{t_2}.*

- *For each $t \in (0,1)$, both $T_t^1 := \lim_{\tau \to 1} T_t^\tau$ and $T_t^0 := \lim_{\tau \to 0} T_t^\tau$ exist and are continuous mappings. In addition, $\mu_1 = T^1_{t,\#}\mu_t$ and $\mu_0 = T^0_{t,\#}\mu_t$ for any $t \in (0,1)$.*

It seems that, in contrast to the Monge problem (M), there is no a-priori reason to pose a limitation on μ_0 and μ_1 in the above paradigm. The optimal Monge map from μ_0 to μ_1 is given by $T := T_t^1 \left(T_t^0\right)^{-1}$, for arbitrary $t \in (0,1)$, provided $\left(T_t^0\right)^{-1}$ exists.

Let us return now to our drifting cloud. If we describe the cloud as such a family of Borel probability measures μ_t for $t \in [0,1]$ it is natural to define the associated velocity field at $t_0 \in (0,1)$ as

$$\mathbf{v}(x,t_0) := \lim_{t \to t_0} \frac{T^t_{t_0}(x) - x}{t - t_0}$$

where $T^t_{t_0}$ is the optimal Monge map transporting μ_t to μ_{t_0}. There are, however, some serious problems with this definition. First, the optimal map $T^t_{t_0}$ may not exist, or it may not be unique, or, even if it exists and unique for any t, the limit may not exist.

Alternatively, one may use the Kantorovich approach and define a measure $\nu_{x,t}(dv)$ in a velocity space $v \in \mathbb{R}^n$ instead of a deterministic velocity field $\mathbf{v}$, via

$$\nu_{x_0,t_0}(dv) := \lim_{t \to t_0} \left(\frac{x - x_0}{t - t_0} \right)_{\#} \lambda^t_{t_0}$$

where $\lambda^t_{t_0}$ is the optimal Kantorovich measure (K) with respect to μ_t, μ_{t_0}. Still, it is not known (yet) how to classify the set of orbits $\mu := \mu_t \otimes dt$ for which the above limit exists $\mu-$a.e.

In this paper we take Ω to be the flat $n-$torus $\mathbb{R}^n/\mathbb{Z}^n$. The starting point is the following definition: Given an orbit μ_t of probability measures supported in Ω for $0 \leq t \leq 1$, define

$$||\mu||_p = \left[\inf_{\mathbf{v}} \int_0^1 \int_\Omega |\mathbf{v}|^p \mu_t(dx) dt \right]^{1/p}$$

where the infimum above is taken on all μ measurable vector fields $\mathbf{v}(x,t)$ compatible with μ (that is, satisfying a weak form of the continuity equation with μ). I denote the set for which $||\mu||_p < \infty$ as $\mathbf{H}_p$. This is a normed cone. In the next section I'll study some of its properties and prove a compactness embedding of $\mathbf{H}_p$ (for $p > 1$) in a set of orbits which satisfies Holder continuity in the weak topology. Then, I'll proved that the Wasserstein-p metric $d_p(\mu_0, \mu_1)$ is nothing but the infimum of $||\mu||_p$ over all orbits in $\mathbf{H}_p$ which satisfy the end conditions $\mu_{t=0} = \mu_0$, $\mu_{t=1} = \mu_1$.

Similar result (for $p = 2$) was obtained by Benamou et.al [BB, BBG].[2] However, the authors, motivated by an algorithmic approach, assumed that μ_0, μ_1 are "smooth enough" (see Proposition 1.1 in [BB]), and applied a time continuation of the optimal Monge map to obtain this result.

In section 3 I'll concentrate in the case $p = 2$. For this case there is a dual representation of $||\mu||_2$ as:

$$||\mu||_2 = \sqrt{\sup \left[\frac{\left(\int \phi_t \mu(dx)\right)^2}{\int |\nabla_x \phi|^2 \mu(dx)} \right]}$$

[2] I'm thankful to D.Kinderlehrer for turning my attention to these publications.

where the supremum is taken on the set of test functions $\phi(x,t) = \phi \in C_0^\infty(\Omega \times [0,1])$. The reader may observe that, with $\mu_t = \delta_{(x-x(t))}$, this norm is nothing but the $\mathbb{L}_2$ norm of $\dot{x}$. Alternatively:

$$\frac{1}{2}||\mu||_2^2 = \sup_{P,\phi}\left[\int_\Omega \phi(x,1)\mu_1(dx) - \int_\Omega \phi(x,0)\mu_0(dx) + \int_0^1\int_\Omega P\mu(dxdt)\right] \tag{1.2}$$

where the supremum is taken over all μ measurable functions P on $\Omega \times [0,1]$ and $\phi = \phi(x,t)$ satisfying

$$\phi_t + \frac{1}{2}|\nabla_x\phi|^2 + P = 0 \quad ; \quad \text{for almost any } (x,t) \in \Omega \times [0,1] \ .$$

If the above supremum is verified for some P and ϕ, then we can denote P as the *pressure field* associated with the orbit μ. At this stage, I do not know if the above supremum is verified for a given μ in $\mathbf{H}_2$. However, it is not difficult to show that, if $\phi = \phi(x,t)$ generates a unique flow $T_{t_0}^t$ via $\dot{x} = \nabla_x\phi(x,t)$ and $\mu = \mu_t \otimes dt$ is any orbit transported by this flow, then the supremum in (1.2) is verified for the above pair $\phi, P = -\phi_t - |\nabla_x\phi|^2/2$.

The first result of this section (see Theorem 1 at the end of Section 3) is the representation for the Wasserstein metric $d_2(\mu_0,\mu_1)$ as:

$$\frac{1}{2}[d_2(\mu_0,\mu_1)]^2 = \max_\phi\left[\int \phi(x,1)\mu_1(dx) - \int \phi(x,0)\mu_0(dx)\right] \tag{1.3}$$

where ϕ is in the set of Lipschitz function $\phi = \phi(x,t)$ which satisfies the homogeneous Hamilton-Jacobi equation:

$$\phi_t + \frac{1}{2}|\nabla_x\phi|^2/2 = 0 \quad ; \quad \text{for almost any } (x,t) \in \Omega \times [0,1] \ . \tag{1.4}$$

I shall further show that the maximizer ϕ of (1.3) satisfies $\phi \in C^1_{loc}(\Omega \times (0,1))$. Moreover, there exists a unique flow $T_{t_1}^t$ on Ω for every $t_1 \in (0,1)$, $t \in [0,1]$ generated by $\dot{x} = \nabla_x\phi(x,t)$ and an orbit $\mu_t \otimes dt$ verifying the Paradigm in the case of quadratic cost function (Theorem 2).

The Hamilton-Jacobi equation (1.4) (or *pressureless potential flow*) was also mentioned in [BB] in similar connection. However, to the best of my knowledge, there is no proof of an existence of a maximizer to problem (1.3) under general condition with no a-priori assumption of the existence of an optimal Monge map. In addition, the existence of optimal maps transporting μ_t ($t \in (0,1)$) to μ_1 and μ_0 does *not* follow from the theorem of existence of optimal transport introduced in [B]. In fact, μ_t is not necessarily continuous with respect to Lebesgue (even when $t \in (0,1)$) and, moreover, the optimal map obtained in [B] is not necessarily continuous.

The question of further regularity of the flow $T_{t_1}^{t_2}$ is also interesting. If we restrict (1.3) to *smooth* functions ϕ which satisfy the homogeneous Hamilton-Jacobi equation, we certainly obtain:

$$\sup_{\phi\in C^\infty}\left[\int \phi(x,1)\mu_1(dx) - \int \phi(x,0)\mu_0(dx)\right] \le \frac{1}{2}[d_2(\mu_0,\mu_1)]^2 \ . \tag{1.5}$$

The next result of Section 3 (see Theorem 3 at the end of that section) shows that the supremum on the left side of (1.5) is attained at a function $\phi \in C^{1,\alpha}_{loc}(\Omega \times (0,1)) \cap LIP(\Omega \times$

$[0,1])$ for any $\alpha < 1$, and the flow $\dot{x} = \nabla_x \phi(x,t)$ induces a family of "almost Lipschitz" homomorphisms $T_{t_1}^{t_2}$. In a future publication I hope to investigate conditions for an equality in (1.5).

The proofs of the main results are given in Section 4, together with a short review on forward, backward and reversible solutions of the Hamilton-Jacobi equation.

It is interesting to compare the dual representation of the d_1 Waserstain metric (1.1) with (1.3). If we restrict to the set of all steady propagating solutions of the Hamilton-Jacobi equation, of the form $\phi(x,t) = \psi(x) - Et$, then $|\nabla\psi|^2 = 2E$. If we optimize E in (1.3) we obtain precisely (1.1). Since, in general, $d_1(\mu_0,\mu_1) < d_2(\mu_0,\mu_1)$, we conclude that, in general, the maximizer of (1.3) is not a steady propagating solution of the Hamilton-Jacobi equation.

List of symbols and definitions

- $\Omega := \mathbb{R}^n/\mathbb{Z}^n$.
- $I = [0,1]$
- $\Omega_I = \Omega \times I$
- LIP_l is the set of all locally Lipschitz functions in $\Omega \times (0,1)$.
- $\mathcal{M}$ is the set of all probability Borel measures supported in Ω.
- $\mathcal{M}_I$ is the set of all Borel probability measures supported on Ω_I which are decomposable as $\mu \in \mathcal{M}_I \Longleftrightarrow \mu = \mu_t \otimes dt$ where $\mu_t \in \mathcal{M}$ a.e. $t \in I$.
- $\pi^{(0)}$ (res. $\pi^{(1)}$) is the natural projection of $\Omega \times \Omega$ on its first (res. second) factor Ω.
- For any pair $\mu_0, \mu_1 \in \mathcal{M}$, the Wasserstein-p metric is defined by

  $$d_p(\mu_0,\mu_1) := \inf_\lambda \int_\Omega \int_\Omega |x-y|^p \lambda(dxdy)$$

  where the infimum is on all probability measures on $\Omega \times \Omega$ such that $\pi^{(0)}_\# \lambda = \mu_0$, $\pi^{(1)}_\# \lambda = \mu_1$.
- $\mathbb{E}_\mu(\psi) := \int_0^1 \int_\Omega \psi(x,t)\mu_t(dx)dt$. Likewise, $\mathbb{E}_{\mu_t}(\psi) = \int_\Omega \psi(x,t)\mu_t(dx)$.
- A lifting ν of $\mu \in \mathcal{M}_I$ is a Borel measure on $\Omega_I \times \mathbb{R}^n$ such that

  $$\mathbb{E}_\nu(\psi) := \int_0^1 \int_\Omega \int_{\mathbb{R}^n} \psi(x,t)\nu(dxdtdv) = \mathbb{E}_\mu(\psi) \;\; ; \; \mathbb{E}_\nu(\psi_t + v \cdot \nabla_x \psi) = 0$$

  for all $\psi \in C_0^\infty(\Omega_I)$.

2 A metric space for measure's orbits

We start with the following
Definition: $\mu \in \mathcal{M}_I \in \mathbf{H}_p(I, \mathcal{M})$ if there exists a lifting ν of μ such that $\mathbb{E}_\nu(|v|^p) < \infty$. We shall also define the $\mathbf{H}_p$ norm of $\mu \in \mathbf{H}_p$ by:

$$||\mu||_p = \inf_\nu \left[\mathbb{E}_\nu(|v|^p)\right]^{1/p}$$

where the infimum is taken over all liftings of μ.

Lemma 2.1. *$\mathbf{H}_p$ is complete and locally compact under the weak C^* topology if $p > 1$. That is, from any bounded sequence μ_n in $\mathbf{H}_p$ we can extract a subsequence which converges in $C^*(\Omega_I)$ to some $\mu \in \mathbf{H}_p$. In addition:*

$$\lim_{n\to\infty} ||\mu_n||_p \geq ||\mu||_p \ .$$

Proof. By definition there exists a set of liftings ν_n corresponding to μ_n. Moreover, this sequence can be chosen so that $\mathbb{E}_{\nu_n}(|v|^p) < C$, so ν_n and $v\nu_n$ are tight on $\Omega_I \times \mathbb{R}^n$ (since $p > 1$). Hence the weak limit ν of ν_n is a lifting of the weak limit μ of μ_n, and $\mathbb{E}_\nu(|v|^p) < C$, hence $\mu \in \mathbf{H}_p$. The same argument also yields the lower-semi-continuity of $\mathbf{H}_p$. □

Lemma 2.2. *If $\mu = \mu_t \otimes dt \in \mathbf{H}_p$, $p > 1$ then the map $t \to \mu_t$ is a Holder $(p-1)/p$ continuous function from I into $\mathcal{M}$ with respect to the weak (C^*) topology.*

Remark: It is known that the Wasserstein-p metric on a compact domain is equivalent to the C^* topology for any $p \geq 1$ (see,e.g, [Am]).

Proof. The reader should notice the analogy of the definition of $\mathbf{H}_p$ with the Wasserstain p-metrics. In fact, it is evident, by definition, that an optimal lifting ν exists for $\mu \in \mathbf{H}_p$. Such a measure can be decomposed, by the Theorem of measure's decomposition [AFP], into $\nu = \mu_t(dx) \otimes dt \otimes \nu_{x,t}(dv)$, for μ a.a. (x,t). We may define now the velocity field

$$\mathbf{v}(x,t) = \mathbb{E}_{\nu_{x,t}}(v)$$

for μ a.a. (x,t). It follows that $\mathbf{v} \in \mathbb{L}_p(\mu(dx))$ and, moreover,

$$||\mu||_p = \left[\int_{\Omega_I} |\mathbf{v}|^p \mu(dxdt)\right]^{1/p} \ .$$

By definition:

$$\int_I \int_\Omega \frac{\partial \phi}{\partial t} \mu_t(dx)dt = -\int_I \int_\Omega \mathbf{v} \cdot \nabla_x \phi \mu_t(dx)dt \tag{2.1}$$

where $\phi = \phi(x,t)$ is in $C_0^\infty(\Omega_I)$. Let $\phi(x,t) = h_n(t)\phi(x)$ with $\phi \in C_0^\infty(\Omega)$ and $h_n \in C_0^\infty(I)$ such that $\lim_{n\to\infty} h_n(x) = \mathbf{1}_{t_1,t_1}(x)$ a.s and h_n is a monotone sequence. Here $\mathbf{1}_{t-1,t_2}(t) = 1$ if $t \in (t_1, t_2)$ and $\mathbf{1}_{t-1,t_2}(t) = 0$ if $t \notin [t_1, t_2]$. It follows that for almost any $t_1, t_2 \in I$ the LHS of (2.1) converges, as $n \to \infty$, to

$$\int_\Omega \phi \mu_{t_2}(dx) - \int_\Omega \phi \mu_{t_1}(dx)$$

while the RHS is estimated by

$$\left|\int_I\int_\Omega \mathbf{v}\cdot\nabla_x\phi\mu_t(dx)dt\right| \le |\nabla\phi|_\infty \left[\int_{\Omega_I} |\mathbf{v}|^p\mu(dxdt)\right]^{1/p} |t_2-t_1|^{(p-1)/p}$$

The result follows by the dual formulation for $p=1$ (1.1) and the remark above. □

Given μ_0 and $\mu_1 \in \mathcal{M}$, define the set

$$\Lambda_p(\mu_0,\mu_1) := \left\{\mu = \mu_t\otimes dt \in \mathbf{H}_p \ ; \ \mu_{(t=0)} = \mu_0 \ , \ \mu_{(t=1)} = \mu_1 ; \right\} \ .$$

Note that $\Lambda_p(\mu_0,\mu_1)$ can be defined equivalently in a weak way

$$\mathbb{E}_\nu(\phi_t + v\cdot\nabla_x\phi) = \mathbb{E}_{\mu_1}(\phi(\cdot,1)) - \mathbb{E}_{\mu_0}(\phi(\cdot,0)) \ \ \forall\ \phi\in C^\infty(\Omega_I) \tag{2.2}$$

where ν is a lifting of $\mu \in \mathbf{H}_p$.

Corollary 2.1. *The set $\Lambda_p(\mu_0,\mu_1)$ where $p>1$ is closed and locally compact in $C(I;C^*(\Omega))$.*

Following [Am], we may approximate any $\mu\in\mathbf{H}_p$ by a measure $\mu^\varepsilon = \rho_\varepsilon(x,t)dxdt$, where $\rho_\varepsilon \in C^\infty(\Omega_I)$. This is done by convoluting μ_t with a smooth, positive kernel $\varepsilon^{-n}\eta(x/\varepsilon)$. Such a regularization induces also a regularization of $\mathbf{v}$ into $\mathbf{v}_\varepsilon \in C^\infty(Supp(\mu))$ (note that $Supp(\mu)$ is the closure of an open set). Evidently,

$$||\mu^\varepsilon||_p \le \left[\int_{\Omega_I} |\mathbf{v}_\varepsilon|^p\mu^\varepsilon(dxdt)\right]^{1/p} \ ,$$

while $\lim_{\varepsilon\to 0}\mu^\varepsilon = \mu$ in $C^*(\Omega_I)$. An analogous argument is also valid for any pair of measures μ_{t_0},μ_{t_1} where $t_0,t_1\in I$, with respect to their optimal Kantorovich measure λ. From these, we obtain the following

Lemma 2.3. (Regularization Lemma): *If $\mu\in\mathbf{H}_p$ then there exists a sequence $\mu^\varepsilon\in\mathbf{H}_p$ with a smooth density so that $\mu = \lim_{\varepsilon\to 0}\mu_\varepsilon$ holds in $C^*(\Omega_I)$ and, moreover,*

$$\lim_{\varepsilon\to 0}||\mu^\varepsilon||_p = ||\mu||_p \ .$$

In addition, for any $t_0,t_1\in I$,

$$\lim_{\varepsilon\to 0} d_p(\mu^\varepsilon_{t_0},\mu^\varepsilon_{t_1}) = d_p(\mu_{t_0},\mu_{t_1}) \ .$$

We next consider the relation between $\mathbf{H}_p$ and the optimal solution of the Kantorovich problem.

Proposition 2.1. *Assume $p\ge 1$. Let $\mu_0,\mu_1\in\mathcal{M}$. Then $\Lambda_p(\mu_0,\mu_1)\neq\emptyset$. and*

$$\inf_{\mu\in\Lambda_p(\mu_0,\mu_1)} ||\mu||_p = d_p(\mu_0,\mu_1)$$

and the infimum above is attained at: $\mu_t = T^{(t)}_\#\lambda$ where $T^{(t)}:\Omega\times\Omega\to\Omega$ given by $T^{(t)}(x,y) = (1-t)x+ty$ and $\lambda(dx,dy)$ is an optimal solution of the Kantorovich problem (K).

The proof is similar to the proof of Theorem 4.2 of Ambrosio [Am] for the metric case $(p = 1)$. A sketch of it is given in the appendix.

We note that Corollary 2.1 is *not* valid in the case $p = 1$. To see it, consider the measure:

$$\mu = \sum_j \alpha_j(t)\delta_{(x-x_j(t)}\otimes dt$$

where $x_j = x_j(t) \in C^\infty(I;\Omega)$ and $\alpha_j \in C^\infty_+(I,\mathbb{R})$ such that $\sum_j \alpha_j(t) = 1\ \forall t \in I$. By Proposition 2.1 we can approximate μ by a sequence of measures $\mu_n \in \Lambda_1(\mu_0,\mu_1)$ as follows: For each $m \in \mathbb{N}$ consider the division $t_k^{(m)} = k/m$, $0 \le k \le m$ of I. Let $T_{m,k}$ be the optimal map of $d_1(\mu_{(t_k^{(m)})},\mu_{(t_{k+1}^{(m)})})$, and $T^{(t)}_{m,k} := \mathbf{Id} + (t-t_k^{(m)})\left[T_{m,k}-\mathbf{Id}\right]/(t_{k+1}^{(m)}-t_k^{(m)})$. Define μ_m as follows:

$$\mu_{m,(t_k)} = \mu_{(t_k)} \ \ ; \ \ \mu_{m,(t)} = T^{(t)}_{m,k,\#}\mu_{m,(t_k)} \ \ ; \ \ t_k^{(m)} \le t \le t_{k+1}^{(m)} \ .$$

Then, by Proposition 2.1, μ_m are bounded in $\mathbf{H}_1$ and $\mu_m \to \mu$. However, $\mu \notin \mathbf{H}_1$ unless α_j are constants in t. To see it, note that the continuity equation takes the form

$$0 = \sum_j \int_I (\alpha_j(t)\phi_t(x_j(t),t) + v_j(t)\cdot\nabla_x\phi(x_j(t),t)dt = \sum_j \int_I -\dot{\alpha}_j\phi(x_j(t),t) + [v_j(t)-\dot{x}_j]\cdot\nabla_x\phi(x_j(t),t)dt$$

where $v_j(t)$ are the velocities attributed to x_j. It is evident that, unless $\dot{\alpha}_j \equiv 0$, for any possible choice of v_j one can find $\phi = \phi(x,t)$ for which the integral on the right does not vanish.

3 Dual representation of $\mathbf{H}_2$

From now, we shall concentrate in the case $p = 2$. Let $J : \mathbb{R}^+ \to \mathbb{R}$ be a convex function which satisfies $J(0) = 0$. We make the convention $J(s) = \infty$ for $s < 0$. Assume that there is a smooth density $\rho(x,t) = \rho_\mu$ associated with $\mu \in \mathbf{H}_2$, and let $f(x,t,v) = f_\nu$ be a density of a lifting ν of μ. Define the functional

$$I_J^\varepsilon(f) = \int_{\Omega_I\times\mathbb{R}^n}\left[\varepsilon J(f) + \frac{1}{2}f|v|^2\right]dxdtdv \ . \tag{3.1}$$

Let

$$\overline{I}_J^\varepsilon(\rho) := \inf_f I_J^\varepsilon(f)$$

where the infimum is taken on the set of functions f satisfying

$$\int_{\mathbb{R}^n} f dv = \rho \ \ ; \ \ \int_{\mathbb{R}^n}\left[f_t + v\cdot\nabla_x f\right]dv = 0 \ \ \forall (x,t)\in\Omega_I \ . \tag{3.2}$$

The following Lemma is self-evident:

Lemma 3.1. *Assume $\mu \in \mathbf{H}_2$ has a density ρ_μ and there exists a lifting ν of μ which has a density $f = f(x,t,v)$, so that $J(f)$ is integrable over $\Omega_I\times\mathbb{R}^n$. Then*

$$\overline{I}_J^\varepsilon(\rho_\mu) \ge \frac{1}{2}||\mu||_2^2 \ \forall\varepsilon > 0 \quad \textit{and} \quad \lim_{\varepsilon\to 0}\overline{I}_J^\varepsilon(\rho_\mu) = \frac{1}{2}||\mu||_2^2 \ .$$

We now claim:

Lemma 3.2. *Let $\mu_0, \mu_1 \in \mathcal{M}$. Then there exists a connecting orbit $\mu \in \Lambda_2(\mu_0, \mu_1)$ with finite $\mathbf{H}_2$ norm and a lifting ν such that both μ and ν has densities in $\mathbb{L}_p(\Omega_I)$ (res. $\mathbb{L}_p(\Omega_I \times \mathbb{R}^n)$), where $1 \leq p < 1 + 1/n$.*

Proof. It is enough to show the lemma for Dirac measures $\mu_0 = \delta_{x_0}$ and $\mu_1 = \delta_{x_1}$. Moreover, we shall consider such an orbit which connect μ_0 at time t_0 to μ_1 at time t_1.

Let $\rho_1(r)$ be a smooth, positive function with compact support such that

$$|\mathbb{S}^{n-1}| \int_0^\infty r^{n-1} \rho_1(r) dr = 1 \ \ ; \ \ \int_0^\infty r^k \rho_1(r) dr := M_k \ \ ; \ \ \int_0^\infty r^{n-1} \rho_1^p dr := L(p) \ .$$

Set also

$$\rho_\alpha(r) = \alpha^n \rho_1(\alpha r) \ \ .$$

Define $\overline{x}(t) = x_0 + (t - t_0)(x_1 - x_0)$ and

$$\mathbf{v}(x,t) = \frac{|x - \overline{x}(t)|^{n-1}(x - \overline{x}(t))}{t - t_0} + \frac{x_1 - x_0}{t_1 - t_0} \ .$$

$$\rho(x,t) = \frac{1}{(t-t_0)^n} \rho_\alpha \left(\frac{|x - \overline{x}(t)|}{t - t_0} \right) \ .$$

A direct calculation shows that ρ satisfies the continuity equation:

$$\rho_t + \nabla_x \cdot \mathbf{v} \nabla_x \rho = 0 \ .$$

Finally, define the lifting of ρ as

$$f(x,t,v) = \sigma^{-n} \pi^{-2/n} \exp\left(-\frac{|v - \mathbf{v}|^2}{\sigma^2} \right) \rho(x,t) \ .$$

It follows immediately that

$$\int_{\mathbb{R}^n} v^2 f(x,t,v) dv = \frac{\sigma n}{2} \rho(x,t) + \mathbf{v}^2 \rho(x,t) \ \ ; \ \ \int_{\mathbb{R}^n} |f|^p dv = p^{-n/2} \pi^{2(1-p)/n} \sigma^{n(1-p)} \rho^p(x,t) \ .$$

Moreover:

$$\int_\Omega \rho(x,t) dx = 1 \ ; \ \int_\Omega \rho^p(x,t) dx = (t - t_0)^{n(1-p)} |\mathbb{S}^{n-1}| \int_0^\infty r^{n-1} \rho_\alpha^p(r) dr = \alpha^{n(p-1)} (t-t_0)^{n(1-p)} |\mathbb{S}^{n-1}| K(p)$$

$$\int_\Omega |\mathbf{v}(x,t)|^2 \rho(x,t) dx = \frac{|x_1 - x_0|^2}{(t_1 - t_0)^2} + |\mathbb{S}^{n-1}| \int_0^\infty ((t - t_0)^{2n-2} r^{3n-1} \rho_\alpha(r) dr =$$

$$\frac{|x_1 - x_0|^2}{(t_1 - t_0)^2} + |\mathbb{S}^{n-1}| t^{2n-2} \alpha^{-2n+1} M_{3n-1}$$

In particular:

$$\int_\Omega \int_{t_0}^{t_1} |v|^2 f = \frac{|x_1 - x_0|^2}{t_1 - t_0} + C_1 |t_1 - t_0|^{2n-1} \alpha^{-2n+1} + O(\sigma)$$

and

$$\int_\Omega \int_{t_0}^{t_1} \rho^p = C_3 |t_1 - t_0|^{n(1-p)+1} \alpha^{n(p-1)}$$

□

Let J^* be the Legender transform of J:

$$J^*(\lambda) = \sup_s \left[s\lambda - J(\lambda) \right] \ .$$

By our assumption on J we have that J^* is also convex and non-negative on $\mathbb{R}$. It satisfies $J^*(\lambda) = 0$ for $\lambda \leq 0$. Let

$$\langle J^* \rangle(s) = \int_{\mathbb{R}^n} J^*(s - |v|^2/2) dv$$

and $\langle J^* \rangle^*$ be the Legender transform of $\langle J^* \rangle$. The next lemma follows by direct calculations:

Lemma 3.3. *Suppose J is convex and $cp^\alpha < J(p) < C(1 + p^\alpha)$ for all $p > 0$, $J = \infty$ for $p < 0$. Then, for possibly other constants c, C, $cq^{\alpha/\alpha-1} < J^*(q) < (C+1)q^{\alpha/\alpha-1}$ and $cq^{\frac{\alpha}{\alpha-1}+n/2} < \langle J^* \rangle(q) < C(1 + q^{\frac{\alpha}{\alpha-1}+n/2})$ for $q > 0$. In particular,*

$$cp^{\frac{n(\alpha-1)+2\alpha}{n(\alpha-1)+2}} < \langle J^* \rangle^*(p) < C(1 + p^{\frac{n(\alpha-1)+2\alpha}{n(\alpha-1)+2}}) \ , so$$

$$\varepsilon^{1+n/2} \langle J^* \rangle^* \left(p\varepsilon^{-n/2} \right) < C \left(\varepsilon^{1+n/2} + \varepsilon^{\frac{2}{n(\alpha-1)+2}} p^{\frac{n(\alpha-1)+2\alpha}{n(\alpha-1)+2}} \right)$$

Define now:

$$I_J^\varepsilon(\phi, \rho_\mu) := \varepsilon^{1+n/2} \int_{\Omega_I} \langle J^* \rangle^* \left(\frac{\rho}{\varepsilon^{n/2}} \right) dxdt - \int_{\Omega_I} \rho(\phi_t + |\nabla_x \phi|^2/2) + \int_\Omega \phi(x,1)\mu_1(dx) - \int_\Omega \phi(x,0)\mu_0(dx) \ .$$

We extend the definition of $I_J^\varepsilon(\phi, \rho_\mu)$ to all functions ϕ which are locally Lipschitz in $\Omega \times (0,1)$ (note that $\phi_t + |\nabla_x \phi|^2/2$ is defined a.e.). In that case, we define

$$I_J^\varepsilon(\phi, \rho_\mu) := \varepsilon^{1+n/2} \int_{\Omega_I} \langle J^* \rangle^* \left(\frac{\rho}{\varepsilon^{n/2}} \right) dxdt - \int_{\Omega_I} \rho(\phi_t + |\nabla_x \phi|^2/2)$$

$$+ \limsup_{t \nearrow 1} \int_\Omega \phi(x,t)\mu_1(dx) - \liminf_{t \searrow 0} \int_\Omega \phi(x,t)\mu_0(dx) \ .$$

Lemma 3.4. *For every $\mu \in \Lambda_2(\mu_0, \mu_1)$ with a density $\rho = \rho_\mu$ for which $\langle J^* \rangle^*(\rho)$ is integrable,*

$$\overline{I}_J^\varepsilon(\rho_\mu) = \sup_{\phi \in C_0^\infty} I_J^\varepsilon(\phi, \rho_\mu) = \sup_{\phi \in C^\infty} I_J^\varepsilon(\phi, \rho_\mu) = \sup_{\phi \in LIP_l} I_J^\varepsilon(\phi, \rho_\mu).$$

Proof. A weak formulation of the constraint $\int f dx = \rho$ is

$$\int_{\Omega_I} \lambda \left[\int_{\mathbb{R}^n} f dv - \rho \right] dxdt = 0 \ , \quad \forall \lambda \in C^\infty(\Omega_I) \tag{3.3}$$

As for the second constraint in (3.2), we can write it in several versions: Either

$$\int_{\Omega_I \times \mathbb{R}^n} f\left[\phi_t + v \cdot \nabla_x \phi\right] dxdtdv = 0 \quad \forall \phi \in C_0^\infty(\Omega_I) \tag{3.4}$$

or

$$\int_{\Omega_I \times \mathbb{R}^n} f\left[\phi_t + v \cdot \nabla_x \phi\right] dxdtdv - \int_\Omega \phi(x,1)\mu_1(dx) + \int_\Omega \phi(x,0)\mu_0(dx) = 0 \quad \forall \phi \in C^\infty(\Omega_I)\ , \tag{3.5}$$

or

$$\int_{\Omega_I \times \mathbb{R}^n} f\left[\phi_t + v \cdot \nabla_x \phi\right] dxdtdv - \limsup_{t \nearrow 1} \int_\Omega \phi(x,t)\mu_1(dx) + \liminf_{t \searrow 0} \int_\Omega \phi(x,t)\mu_0(dx) \geq 0\ , \forall \phi \in LIP_l \tag{3.6}$$

Indeed, it is evident that, granted (3.3), both (3.4) and (3.5) are equivalent since ρ satisfies the end conditions μ_0, μ_1 as $t \to 0, 1$ by assumption. Now, (3.6) is equivalent to (3.5) for all $\phi \in C^\infty$ (just replace $\phi \to -\phi$). If $\phi \in LIP_l$ is the distributional limit of a sequence $\phi_n \in C^\infty$ then the inequality is preserved.

Define

$$F_\varepsilon(f, \rho, \phi, \lambda) := \int_{\Omega_I \times R^n} dxdtdv \left[\varepsilon J(f) - \left(\phi_t + v \cdot \nabla_x \phi - \frac{1}{2}|v|^2\right) f - \lambda(f - \rho)\right]$$

Then

$$\overline{I}_J^\varepsilon(\rho) = \sup_{\phi \in C_0^\infty} \sup_\lambda \inf_f F_\varepsilon(f, \rho, \phi, \lambda)$$

$$= \sup_{\phi \in C^\infty} \sup_\lambda \inf_f F_\varepsilon(f, \rho, \phi, \lambda) + \int_\Omega \phi(x,1)\mu_1(dx) - \int_\Omega \phi(x,0)\mu_0(dx)$$

$$= \sup_{\phi \in LIP_l} \sup_\lambda \inf_f F_\varepsilon(f, \rho, \phi, \lambda) + \limsup_{t \nearrow 1} \int_\Omega \phi(x,t)\mu_1(dx) - \liminf_{t \searrow 0} \int_\Omega \phi(x,t)\mu_0(dx)$$

We readily compute:

$$\inf_f F_\varepsilon(f, \rho, \phi, \lambda) = -\varepsilon \int_{\Omega_I \times R^n} dxdtdJ^* \left(\frac{\phi_t + v \cdot \nabla_x \phi - v^2/2 + \lambda}{\varepsilon}\right) + \int_{\Omega_I} \lambda\rho$$

Moreover,

$$\int_{R^n} dvJ^* \left(\frac{\phi_t + v \cdot \nabla_x \phi - v^2/2 - \lambda}{\varepsilon}\right) = \varepsilon^{n/2} \langle J^* \rangle \left(\frac{\phi_t + |\nabla_x \phi|^2/2 - \lambda}{\varepsilon}\right)$$

hence

$$F_\varepsilon(\rho, \phi, \lambda) := \inf_f F_\varepsilon(f, \rho, \phi, \lambda) = -\varepsilon^{1+n/2} \int_{\Omega_I} \langle J^* \rangle \left(\frac{\phi_t + |\nabla_x \phi|^2/2 + \lambda}{\varepsilon}\right) dxdt + \int_{\Omega_I} \lambda\rho$$

We now substitute $\lambda \to \hat{\lambda} = (\lambda + \phi_t + |\nabla_x \phi|^2/2)/\varepsilon$ to obtain

$$F_\varepsilon(\rho, \phi, \hat{\lambda}) = -\varepsilon^{1+n/2} \int_{\Omega_I} \langle J \rangle^* \left(\hat{\lambda}\right) dxdt + \varepsilon \int_{\Omega_I} \hat{\lambda}\rho - \int_{\Omega_I} \rho(\phi_t + |\nabla_x \phi|^2/2)\ .$$

Taking now the supremum above with respect to $\hat{\lambda} \in C^\infty$, we obtain the result. □

From Lemma 3.3 we can take $\varepsilon \to 0$ in the definition of $\overline{I}_J^\varepsilon$ and obtain, via Lemma 3.4, Lemma 3.1 and the regularization Lemma 2.3:

Corollary 3.1. *If $\mu \in \Lambda_2(\mu_0, \mu_1)$ then —*

$$\frac{1}{2}||\mu||_2^2 = -\inf_{\phi \in C_0^\infty} \left\{ \int_{\Omega_I} (\phi_t + |\nabla_x \phi|^2/2)\mu(dx) \right\} = \sup_{\phi \in C_0^\infty} \frac{\left(\int_{\Omega_I} \phi_t \mu(dx) \right)^2}{\int_{\Omega_I} |\nabla_x \phi|^2 \mu(dx)} . \tag{3.7}$$

as well as

$$\frac{1}{2}||\mu||_2^2 = -\inf_{\phi \in C^\infty} \left\{ \int_{\Omega_I} (\phi_t + |\nabla_x \phi|^2/2)\mu(dx) + \int_{\Omega} \phi(x,0)\mu_0(dx) - \int_{\Omega} \phi(x,1)\mu_1(dx) \right\} =$$

$$-\inf_{\phi \in LIP_l} \left\{ \int_{\Omega_I} (\phi_t + |\nabla_x \phi|^2/2)\mu(dx) + \liminf_{t \searrow 0} \int_{\Omega} \phi(x,t)\mu_0(dx) - \limsup_{t \nearrow 1} \int_{\Omega} \phi(x,t)\mu_1(dx) \right\} \tag{3.8}$$

An interesting application of the above results is:

Corollary 3.2. *If ϕ is a Lipschitz function, then, for any x_0, x_1 in Ω and any $t_1 > t_0$:*

$$\phi(x_1, t_1) - \phi(x_0, t_0) \le \frac{1}{2} \frac{|x_1 - x_0|^2}{t_1 - t_0} + C(s)||\phi_t + \frac{1}{2}|\nabla_x \phi|^2/2||_s (t_1 - t_0)^\gamma ,$$

where $s > n+1$ and $\gamma > \frac{s-1}{s} \frac{(2n-1)(n+1)}{2n(n+1)-1}$.

Proof. We use Corollary 3.1 with μ supported on $\Omega \times [t_1, t_0]$ and $\mu_{t_0} = \delta_{x_0}$, $\mu_{t_1} = \delta_{x_1}$, to obtain

$$\phi(x_1, t_1) - \phi(x_0, t_0) \le \frac{1}{2}||\mu||_2^2 + \left| \int_\Omega \int_{t_0}^{t_1} (\phi_t + |\nabla_x \phi|^2/2)\mu_t(dx)dt \right| .$$

By the proof of Lemma 3.2 we can find such a μ for which:

$$||\mu||_2^2 \le \frac{|x_1 - x_0|^2}{t_1 - t_0} + C_1 |t_1 - t_0|^{2n-1} \alpha^{-2n+1}$$

and, for the density $\rho = \rho_\mu$:

$$\int_\Omega \int_{t_0}^{t_1} \rho^p \le C_3 |t_1 - t_0|^{n(1-p)+1} \alpha^{n(p-1)}$$

where $p < 1 + 1/n$ and α any positive constant. Let $p* = p/(p-1)$, $\beta > 1$ and $s = \beta p*$. It follows that

$$\left| \int_\Omega \int_{t_0}^{t_1} (\phi_t + |\nabla_x \phi|^2/2)\rho dx dt \right| \le ||\rho||_p ||\phi_t + |\nabla_x \phi|^2/2||_{p*} \le ||\rho||_p ||\phi_t + |\nabla_x \phi|^2/2||_s |t_1 - t_0|^{1/\beta * p*}$$

Since $1/(\beta * p*) = 1/p * -1/s$ we obtain:

$$\phi(x_1, t_1) - \phi(x_0, t_0) \le \frac{1}{2} \frac{|x_1 - x_0|^2}{t_1 - t_0} + C_3^{1/p} |t_1 - t_0|^{n(1-p)/p + 1 - 1/s} \alpha^{n(p-1)/p} + C_1 |t_1 - t_0|^{2n-1} \alpha^{-2n+1} .$$

The choice $\alpha = (t_1 - t_0)^\gamma$ where $\gamma = 1 - \frac{s-1}{s(3n-1-n/p)}$ is the optimal choice and yields the desired result. □

Another application are the theorems below, to be proved in the next section:

Theorem 1. *For any μ_0, $\mu_1 \in \mathcal{M}$:*

$$\frac{d_2^2(\mu_0,\mu_1)}{2} = \max_{\phi\in LIP(\Omega_I)} \left\{ \int_\Omega \phi(x,1)\mu_1(dx) - \int_\Omega \phi(x,0)\mu_0(dx) \right\} \quad ; \phi \text{ satisfies } (3.11) \text{ a.e }. \tag{3.9}$$

Moreover, the maximizer is a reversible solution of (3.11) .

Theorem 2. *If $\phi \in LIP(\Omega_I)$ is a reversible solution of (3.11) then $\phi \in C^1(\Omega \times (0,1))$. Let $\mu_t = [(1-t)x + ty]_\#\lambda$ where λ is the optimal solution of the Kanterovich problem with respect to μ_0, μ_1. Then, for the maximizer of (3.9), the flow $T^t_{t_1}(x) := x + (t - t_1)\nabla_x\phi(x,t_1)$ is an homomorphism on Ω for any $t_1, t_2 \in (0,1)$, is continuous if $t \in [0,1]$, and is an optimal Monge transport from μ_{t_1} to μ_t for any $t_1 \in (0,1)$, $t \in [0,1]$.*

Theorem 3. *(i) For any μ_0, $\mu_1 \in \mathcal{M}$:*

$$\frac{d_2^2(\mu_0,\mu_1)}{2} \geq \sup_{\phi\in C^\infty(\Omega_I)} \left\{ \int_\Omega \phi(x,1)\mu_1(dx) - \int_\Omega \phi(x,0)\mu_0(dx) \right\} \tag{3.10}$$

where the supremum is taken on all **smooth** *functions $\phi = \phi(x,t)$ which satisfies the homogeneous Hamilton-Jacobi equation*

$$\phi_t + \frac{1}{2}|\nabla_x\phi|^2 = 0 \ . \tag{3.11}$$

(ii) There exists a function ϕ which realizes the supremum on the right side of (3.10). Moreover, $\phi \in C^{1,\alpha}_{loc}(\Omega_I) \cap LIP(\Omega_I)$ for any $0 \leq \alpha < 1$, (if $n = 1$ then $\phi \in C^{1,1}_{loc}(\Omega_I) \cap LIP(\Omega_I)$) and the induced flow $T^{t_1}_{t_0}$ is a family of homomorphisms on $\Omega \times (0,1)$ which satisfies:

$$\left|T^t_{t_0}(x) - T^t_{t_0}(y)\right| \leq C(t,t_0)|x-y| \left(1 + \ln^\beta \frac{1}{|x-y|}\right)$$

for some $\beta = \beta(t,t_0) \geq 0$ (if $n = 1$ then $\beta \equiv 0$).

4 Proof of main results

I'll review some facts about the equation

$$\phi_t + \frac{1}{2}|\nabla_x\phi|^2 = P \quad (x,t) \in \Omega_I \tag{4.1}$$

where $P \in LIP(\Omega_I)$. It is known that if ϕ (and hence P) are smooth, then $\forall x_0, x_1 \in \Omega$, $\forall t_1 > t_0$, the solution ϕ can be represented either in the forward version:

$$\phi(x_1,t_1) = \inf_{y=y(s)} \left[\int_{t_0}^{t_1} \left[\frac{|\dot{y}|^2}{2} + P(y(s),s) \right] ds + \phi(y(t_0),t_0) \right] \quad , \tag{\textbf{F}}$$

where the infimum above is taken on all orbits $y(s) : [t_0, t_1] \to \Omega$ such that $y(t_1) = x_1$, or in the backward version

$$\phi(x_0, t_0) = \sup_{y=y(s)} \left[\int_{t_0}^{t_1} \left[-\frac{|\dot{y}|^2}{2} - P(y(s), s) \right] ds + \phi(y(t_1), t_1) \right] . \qquad (\mathbf{B})$$

where the supremum above is taken on all orbits $y(s) : [t_0, t_1] \to \Omega$ such that $y(t_0) = x_0$. From the above it follows that one can solve the initial value problem (4.1) in Ω_I with given $\phi(x, 0)$, or the end value problem with given $\phi(x, 1)$, provided we *a-priori* know that the solution is smooth. Moreover, we know that, in this case, the solution is reversible in the following sense: If we solve the initial value problem with given $\phi(x, 0)$, using (**F**) with $(x, t) = (x_1, t_1)$ and $(x_0, t_0) = (x_0, 0)$, and then we use the obtained solution at $t = 1$ in the backward formulation (**B**) with $(x_0, t_0) = (x, t)$ and $(x_1, t_1) = (x_1, 1)$, we recover *the same* solution ϕ. Such a solution is called *reversible.*

The following claims are common knowledge (or, at least, should be):

Claim 1: *For any initial data $\phi(, 0) \in C(\Omega)$ and $P \in LIP(\Omega_I)$ one can construct a forward solution*

$$\phi(x, t) = \inf_{y=y(s)} \left[\int_0^t \left[\frac{|\dot{y}|^2}{2} + P(y(s), s) \right] ds + \phi(y(0), 0) \right] \quad , \qquad (\mathbf{F})$$

which satisfies **F** *for every $0 \le t_0 < t_1 \le 1$. Moreover, $\phi \in LIP(\Omega \times (0, 1])$ and satisfies* (4.1) *a.e. The analogous claim holds also for backward solution with prescribed $\phi(, 1)$. This time, $\phi \in LIP(\Omega \times [0, 1))$. If $\phi(, 0)$ (res. $\phi(, 1)$) are Lipschitz in Ω then $\phi \in LIP(\Omega_I)$.*

Claim 2 *If ϕ is a forward solution and ψ is a backward solution so that $\phi(, 1) = \psi(, 1)$ holds then $\psi(, t) \le \phi(, t)$* **for every** *$t \in I$. If, in addition, $\phi(, 0) = \psi(, 0)$ then $\phi \equiv \psi$ is a reversible solution. In particular, any backward solution ψ which coincide with some forward solution at $t = 1$ is a reversible solution.*

Claim 3 *If a sequence $\phi_n(, t_0)$ converges uniformly to $\phi(, t_0)$, then the corresponding forward solutions ϕ_n converges uniformly to a forward solution ϕ for any $t > t_0$. Same holds for backward solutions (where, this time, $t \le t_0$).*

We now turn to the proof of the main results of Section 3

Proof. (Theorem 1):
Use Proposition 2.1 and Lemma 3.3 to obtain

$$\lim_{\varepsilon \to 0} \inf_f I_J^\varepsilon(f) = \frac{1}{2} d_2^2(\mu_0, \mu_1)$$

where the infimum is taken over all f which satisfies (3.2) subjected to the end conditions $\int f(., 0, v) dv = \mu_0$ and $\int f(., 1, v) dv = \mu_1$. Proceeding as in the proof of Lemma 3.4 when we use now the constraint (3.5), we obtain

$$\inf_f I_J^\varepsilon(f) = \sup_{\phi \in C^\infty} \Psi_\varepsilon(\phi)$$

where

$$\Psi_\varepsilon(\phi) := \left[-\varepsilon^{1+n/2} \int_{\Omega_I} \langle J^* \rangle \left(\frac{\phi_t + |\nabla_x \phi|^2/2}{\varepsilon} \right) dxdt + \int_\Omega \phi(x, 1) \mu_1(dx) - \int_\Omega \phi(x, 0) \mu_0(dx) \right] .$$

By Lemma 3.3 we obtain that

$$\varepsilon^{1+n/2}\int_{\Omega_I}\langle J^*\rangle\left(\frac{\phi_t+|\nabla_x\phi|^2/2}{\varepsilon}\right)\geq c\varepsilon^{-1/(\alpha-1)}||\phi_t+|\nabla_x\phi|^2/2||_s^s$$

where $s>n+1$ provided we choose $J(s)\approx s^\alpha$ with $1<\alpha<1+(2/n)$. Now, take a sequence $\phi_k\in C^\infty$ and $\varepsilon_k\to 0$ such that $\lim_{k\to\infty}\Psi_{\varepsilon_k}(\phi_k)=d_2^2(\mu_0,\mu_1)/2$. By Corollary 3.2 we obtain that

$$\frac{d_2(\mu_0,\mu_1)}{2}\leftarrow\Psi_{\varepsilon_k}(\phi_k)<-c\varepsilon^{-1/(\alpha-1)}||\phi_{k,t}+|\nabla_x\phi_k|^2/2||_s^s+C||\phi_{k,t}+|\nabla_x\phi_k|^2/2||_s$$

hence $||\phi_{k,t}+|\nabla_x\phi_k|^2/2||_s\to 0$. Moreover,

$$\lim_{k\to\infty}\left[\int_\Omega\phi_k(x,1)\mu_1(dx)-\int_\Omega\phi_k(x,0)\mu_0(dx)\right]\geq\frac{d_2^2(\mu_0,\mu_1)}{2}\ , \tag{4.2}$$

while

$$\phi_{k,t}+\frac{1}{2}|\nabla_x\phi_k|^2:=P_k\to 0\ \text{ in }\mathbb{L}_s(\Omega_I)\ . \tag{4.3}$$

Since ϕ_k are defined up to a constant, we may set a convention by which $\min_\Omega\phi_k(x,0)=0$. Since $\langle J^*\rangle(s)=0$ for $s<0$ we may assume $P_k\geq 0$ for, otherwise, define $P_k^+=[P_k]_+$ and ϕ_k^+ to be the forward solution of the Hamilton-Jacobi equation with P_k^+ on the r.h.s and $\phi_k^+(,0)=\phi_k(,0)$. By Claim 1, $\phi_k^+\in Lip(\Omega_I)$ and $\phi_k^+\geq\phi_k$ for all $(x,t)\in\Omega_I$. Thus, $\phi_k^+\in LIP(\Omega_I)$ is a maximizing sequence as well, satisfying (4.2). Since, moreover, $\min_\Omega\phi_k^+(x,0)=0$ by assumption we obtain from Corollary 3.2 that $\phi_k^+(,1)$ are uniformly bounded in $\mathbb{L}_\infty(\Omega)$. In fact, we may assume that ϕ_k^+ are uniformly bounded in $\mathbb{L}_\infty(\Omega_I)$ as well since, otherwise, define ϕ_k^- to be the backward solution of $\phi_t+|\nabla_x\phi|^2/2=P_k^+$ such that $\phi_k^-(,1)=\phi_k^+(,1)$. By Claim 2 it follows that $\phi_k^-\leq\phi_k^+$ on Ω_I, so ϕ_k^- is, again, a maximizing sequence in the sense of (4.2). By definition of backward solution (and since $\phi_k^-(,1)$ are uniformly bounded by assumption and $P_k^+\geq 0$), we obtain that $\phi_k^-(,0)$ are uniformly bounded from above . Applying Corollary 3.2 again we see that $\phi_k^-(,0)$ are uniformly bounded from below as well. Finally, define ϕ_k^* to be the forward solution subjected to $\phi_k^*(,0)=\phi_k^-(,0)$. By the same argument as above this is a maximizing sequence which is uniformly bounded in $\mathbb{L}_\infty(\Omega_I)$.

As for now, we have a uniformly $\mathbb{L}_\infty$ bounded maximizing sequence $\phi_k^*\in Lip(\Omega_I)$ which are all forward solutions of the Hamilton-Jacobi equation with P_k^+ on the r.h.s, and $P_k^+\to 0$ in $\mathbb{L}_s$ where $s>n+1$.

Since both $\phi_n^*(,0)$ and $\phi_n^*(,1)$ are both *pointwise* bounded it follows that there exist weak limits, denoted by $\phi(,0)$ and $\phi(,1)$ respectively, such that

$$\int_\Omega\phi(x,i)\mu_i(dx)=\lim_{n\to\infty}\int_\Omega\phi_n^*(x,i)\mu_i(dx)\quad ;\quad i=0,1$$

hence

$$\int_\Omega\phi(x,1)\mu_1(dx)-\int_\Omega\phi(x,0)\mu_0(dx)\geq\frac{d_2^2(\mu_0,\mu_1)}{2}\ . \tag{4.4}$$

Let ψ_n be the forward solutions:

$$\psi_{n,t}+\frac{1}{2}|\nabla_x\psi_n|^2=0\quad ;\quad \psi(x,0)=\phi_n^*(x,0)\ .$$

Using Corollary 3.2 I claim that

$$\psi_n(x,1) \geq \phi_n^*(x,1) - O(||P_n^+||_s) \ .$$

Indeed, $\forall x_1 \in \Omega \exists x_0 \in \Omega$ such that $\psi_n(x_1,1) = \phi_n^*(x_0,0) + (x_1-x_0)^2/2$, and we apply Corollary 3.2 for $t_1 = 1$, $t_0 = 0$ to obtain the above result. In particular we obtain that

$$\lim_{n\to\infty} \int_\Omega \psi_n(,1)\mu_1(dx) \geq \int_\Omega \phi(,1)\mu_1(dx) \ ,$$

whenever the limit exists.

Being a sequence of forward solutions with uniformly bounded initial data, ψ_n are locally uniformly bounded in $Lip(\Omega\times(0,1])$. Therefore, by Claim 3, there is a limit $\psi \in Lip(\Omega\times(0,1])$ for some subsequence which is a forward solution. Moreover

$$\int_\Omega \psi(,1)\mu_1(dx) \geq \int_\Omega \phi(,1)\mu_1(dx) \ .$$

In addition we have by definition of forward solution (with $P \equiv 0$): $\psi_n(x,t) \leq \phi_n^*(x,0)$ for any $t > 0$, hence

$$\liminf_{t\searrow 0} \int_\Omega \psi(x,0)\mu_0(dx) \leq \int_\Omega \phi(x,0)\mu_0(dx) \ ,$$

and

$$\int_\Omega \psi(x,1)\mu_1(dx) - \liminf_{t\searrow 0} \int_\Omega \psi(x,0)\mu_0(dx) \geq \frac{d_2^2(\mu_0,\mu_1)}{2} \ . \tag{4.5}$$

Using Corollary 3.1 we obtain that, in fact, there is an equality in (4.5) so ψ is a forward maximizer in $LIP(\Omega\times(0,1])$. Now we build a backward solution η from $\psi(,1)$, using Claim 1. Since the later is Lipschitz, $\eta \in LIP(\Omega_I)$ and $\eta(x,t) \leq \psi(x,t)$ on Ω_I by Claim 2, so η is a backward maximizer. Using Claim 2 again, it is also a reversible solution. □

Proof. (Theorem 2):
Let $\tau \in (0,1)$ and $x \in \Omega$. Since ϕ is reversible, then

$$(i)\ \phi(x,\tau) - \phi(y_1,0) \leq \frac{|x-y_1|^2}{2\tau} \quad ; \quad (ii)\ \phi(y_2,1) - \phi(x,\tau) \leq \frac{|x-y_2|^2}{2(1-\tau)} \quad \forall y_1,y_2 \in \Omega \ ,$$

where an equality holds for some (possibly several) point. Adding (i) and (ii) we obtain:

$$\phi(y_2,1) - \phi(y_1,0) \leq \frac{|x-y_1|^2}{2\tau} + \frac{|x-y_2|^2}{2(1-\tau)} \leq \frac{|y_1-y_2|^2}{2} \ . \tag{4.6}$$

The last inequality is a strict one *unless* $x = (1-\tau)y_1 + \tau y_2$.

Let now $y_1^* \in \Omega$ for which an equality holds in (i). There exists some $y_2^* \in \Omega$ for which

$$\phi(y_2^*,1) - \phi(y_1^*,0) = \frac{|y_1^* - y_2^*|^2}{2} \ .$$

Therefore, if we substitute y_1^*, y_2^* for y_1, y_2 in (4.6), we must have equality in both places. This implies that there is an equality for $y_2 = y_2^*$ in (ii) as well as $x = (1-\tau)y_1^* + \tau y_2^*$.

Now, we substitute y_1^* and y_2^* in (i),(ii), and replace x by some $x^{'} \in \Omega$. We still have

$$(i)\ \phi(x^{'},\tau) - \phi(y_1^*,0) \leq \frac{|x^{'} - y_1^*|^2}{2\tau} \quad ; \quad (ii)\ \phi(y_2^*,1) - \phi(x^{'},\tau) \leq \frac{|x^{'} - y_2^*|^2}{2(1-\tau)} \quad \forall x^{'} \in \Omega$$

where equality holds if $x^{'} = x$. Therefore,

$$(i)\ \phi(x,\tau) - \phi(x^{'},\tau) \geq \frac{|x - y_1^*|^2}{2\tau} - \frac{|x^{'} - y_1^*|^2}{2\tau} = \frac{(x - x^{'})(x + x^{'} - 2y_1^*)}{2\tau}$$

$$(ii)\ \phi(x,\tau) - \phi(x^{'},\tau) \leq \frac{|x - y_2^*|^2}{2(1-\tau)} - \frac{|x^{'} - y_2^*|^2}{2(1-\tau)} = \frac{(x - x^{'})(x + x^{'} - 2y_2^*)}{2(1-\tau)}$$

It follows that

$$(i)\ \liminf_{x^{'}\to x} \frac{\phi(x,\tau) - \phi(x^{'},\tau)}{x - x^{'}} \geq \frac{(x - y_1^*)}{\tau} \qquad (ii)\ \limsup_{x^{'}\to x} \frac{\phi(x,\tau) - \phi(x^{'},\tau)}{x - x^{'}} \leq \frac{(x - y_2^*)}{(1-\tau)} \ .$$

However, $x = (1-\tau)y_1^* + \tau y_2^*$ so $\nabla_x \phi(x,\tau)$ exists for any $x \in \Omega$, $\tau \in (0,1)$ and

$$\nabla_x \phi(x,\tau) = \frac{(x - y_1^*)}{\tau} = \frac{(x - y_2^*)}{(1-\tau)} = y_2^* - y_1^* \ .$$

The above arguments also gurantee that y_1^* and y_2^* are uniquely determined by x. Hence

$$\phi(x,\tau) > \phi(y,0) + \frac{|x-y|^2}{2\tau} \quad \forall y \neq y_1^* \ .$$

Since $\phi(,\tau)$ and $\phi(,0)$ are both continuous functions (of x), it follows by standard arguments that y_1^* is a continuous function of x. Same holds for y_2^*, hence $\nabla_x \phi(,\tau)$ is continuous for all $\tau \in (0,1)$.

It follows that ϕ satisfies the Hamilton-Jacobi equation (3.11) *pointwise* for $(x,t) \in \Omega \times (0,1)$. The flow $\dot{x} = \nabla_x \phi(x,t)$ is defined. We show now that it is, in fact, unique and induces an homomorphism given by

$$T_\tau^t(y) = y + (t-\tau)\nabla_x \phi(y,\tau) \ . \tag{4.7}$$

For any $(y,\tau) \in \Omega \times (0,1)$ and any $t \in (0,1)$ we define $x = x(t)$ as the (unique) solution of

$$\phi(x,t) - \phi(y,\tau) = \frac{|x-y|^2}{2|t-\tau|} \ . \tag{4.8}$$

We proved that the equality

$$\nabla_x \phi(x,t) = \frac{x-y}{t-\tau} \tag{4.9}$$

is satisfied if x realizes (4.8). From (4.9) we obtain that $x(t)$ satisfies the differential equation $\dot{x} = \nabla_x \phi(x,t)$, subjected to $x(\tau) = y$, and is, actually, given implicitly by:

$$x(t) = y + (t-\tau)\nabla_x \phi(x,t) \tag{4.10}$$

However, by (4.9) again, $\nabla_x \phi(x(t),t)$ is differentialble as a function of t, and

$$\frac{d}{dt}\nabla_x \phi(x(t),t) = \frac{d}{dt}\frac{x(t)-t}{t-\tau} = (t-\tau)^{-1}\nabla_x\phi(x(t),t) - \frac{x-y}{(t-\tau)^2} = 0 \ ,$$

hence $\nabla_x\phi(x(t),t) = \nabla_x\phi(y,\tau)$. This, together with (4.10), yields (4.7).

Let λ be the optimal solution of the Kantorovich problem (K) for μ_0, μ_1. Set $\mu_t = [x+t(y-x)]_\#\lambda$. By Proposition 2.1, $||\mu||_2 = 2^{1/2}d_2(\mu_0,\mu_1)$. Insert μ in (3.8) of Corollary 3.1. By assumption, ϕ maximizes (3.8). Thus, if we substitute $\phi + \varepsilon\psi$ for $\psi \in C_0^\infty(\Omega_I)$ in (3.8) we obtain:

$$\int_{\Omega_I} [\psi_t + \nabla_x\psi\cdot\nabla_x\phi]\,\mu(dxdt) = 0$$

which is a weak form of the continuity equation for μ. Since the flow $T_{t_1}^{t_2}$ is generated by $\nabla_x\phi$ it follows that this flow transports μ_{t_1} to μ_{t_2}.

By (4.8) and (4.7) we have, for $t > \tau$:

$$\phi(T_\tau^t(x),t) - \phi(x,\tau) = \frac{1}{2}\frac{|T_\tau^t(x)-x|^2}{t-\tau}$$

hence

$$\int_\Omega |T_\tau^t(x)-x|^2\mu_\tau(dx) = 2(t-\tau)\int_\Omega \phi(x,t)\mu_t(dx) - \phi(x,\tau)\mu_\tau(dx) \le d_2^2(\mu_\tau,\mu_t) \le \int_\Omega |T_\tau^t(x)-x|^2\mu_\tau(dx)$$

where the first inequality above follows from (3.8) (adjusted to the interval $[\tau,t]$). In conclusion, an equality holds in both places hence T_τ^t is an optimal Monge map. □

Proof. (Theorem 3):
The first part follows from Corollary 3.1 and Proposition 2.1. Just consider (3.8) restricted to $\phi \in C^\infty$ which satisfies $\phi_t + |\nabla_x\phi|^2/2 = 0$, and use Proposition 2.1 to take the infimum of all $\mu \in \Lambda_2(\mu_0,\mu_1)$.

To prove part (ii), we first note that a maximizing sequence ϕ_n must be uniformly bounded in $LIP(\Omega_I)$. Indeed, since the functions ϕ_n are defined up to a constant, we may assume that $\min_\Omega \phi_n(x,0) = 0$. Using Claim 1 (with $P \equiv 0$) we obtain, since ϕ_n is a forward solution:

$$\phi_n(x,t) = \inf_{y\in\Omega}\left[\frac{(x-y)^2}{t} + \phi_n(y,0)\right]$$

Then $\phi_n(x,t) \in LIP_{loc}(\Omega\times(0,1])$ with $n-$independent estimate. However, ϕ_n are also backward solutions, hence

$$\phi_n(x,t) = \sup_{y\in\Omega}\left[-\frac{(x-y)^2}{1-t} + \phi_n(y,1)\right]$$

hence $\phi_n(x,t) \in LIP_{loc}(\Omega\times[0,1))$, with $n-$independent estimate as well.

Let us assume, first, that $n = 1$. Set $u := \phi_x^{'}$. Then u_n satisfies the Burgers equation:

$$u_{n,t} + u_n u_{n,x} = 0$$

If we further define $w = u_x^{'}$ then

$$w_{n,t} + u_n w_{n,x} + w_n^2 = 0 \ .$$

This implies that, along a characteristic line:

$$\pm \frac{d}{dt} w_n + w_n^2 = 0 \ . \tag{4.11}$$

where the $+$ $(-)$ sign corresponds to an integration up (down) the characteristic curve. Since $\phi_n \in C^\infty(\Omega_I)$ it follows that w_n are locally uniformly bounded (from above and below) in $\Omega \times (0,1)$ for, otherwise, the solution of (4.11) blows up at some $t \in (0,1)$. Hence $\phi_n \in C^2_{loc}(\Omega \times (0,1))$ with a uniform ($n-$independent) estimate. Hence, the maximizer ϕ, being a uniform limit of ϕ_n, is in $C^{1,1}_{loc}(\Omega \times (0,1))$.

In the case $n > 1$ we proceed similarly. Define $u^{(i)} = \partial_{x_i}\phi$ and $w^{(i,j)} = w^{(j,i)} = \partial^2_{x_i,x_j}\phi$. Using $\Delta_x \phi = \sum_i w^{(i,i)}$ we obtain the equality

$$\partial_t \Delta_x \phi_n + u_n \cdot \nabla_x \Delta_x \phi_n + \sum_i \sum_j \left[\partial^2_{x_i,x_j} \phi_n \right]^2 = 0 \ ,$$

which yields, again,

$$\pm \frac{d}{dt}(\Delta_x \phi) + \sum_i \sum_j \left[\partial^2_{x_i,x_j} \phi_n \right]^2 = 0$$

along the characteristic curves. Using $\sum_i \sum_j \left[\partial^2_{x_i,x_j} \phi_n \right]^2 \geq [\Delta_x \phi]^2$ we obtain, similarly to the case $n = 1$, a uniform bound $|\Delta_x \phi_n(,t)| < C(t)$ with $C \in \mathbb{L}_{\infty,loc}(0,1)$.

Now, we call upon [G,T] (Theorem 3.9), to obtain an estimate of the form

$$\frac{|\nabla_x \phi_n(x,t) - \nabla_y \phi_n(y,t)|}{|x-y|} \leq C \left[\sup_\Omega |\phi_n| + |\Delta_x \phi_n| \left| \log \frac{1}{|x-y|} \right| + 1 \right] \ .$$

The above estimate survives in the limit $n \to \infty$ for the maximizer ϕ. We obtain, therefore,

$$\frac{|\nabla_x \phi(x,t) - \nabla_y \phi(y,t)|}{|x-y|} \leq C \left[\sup_\Omega |\phi| + C(t) \left| \log \frac{1}{|x-y|} \right| + 1 \right] \ .$$

By Osgood criterion (see ODE book by Hartman), the above estimate is sufficient for the uniqueness of the flow $T^{t_1}_{t_0}$ generated by $\dot{x} = \nabla_x \phi(x,t)$, at least for $t \in (0,1)$. The modulus of continuity of this flow follows from standard estimates.

□

Appendix

Proof of Proposition 2.1: First, let $\mu_t = [x + t(y-x)]_{\#}\lambda$ where λ is the optimal solution of (K). We claim that $\nu = \nu_t \otimes dt$ where

$$\nu_t := [x + t(y-x), (y-x)]_{\#}\lambda \tag{.12}$$

is a lifting of $\mu = \mu_t \otimes dt$. Indeed, given $\phi \in C_0^\infty(\Omega_I)$, consider

$$\frac{d}{dt}\int_\Omega \phi\mu_t(dx) = \frac{d}{dt}\int_{\Omega\times\Omega} \phi(x+t(y-x),t)\lambda(dxdy)$$

$$= \int_{\Omega\times\Omega} \left[\phi_t(x+t(y-x),t) + (y-x)\cdot\nabla_x\phi(x+t(y-x),t)\right]\lambda(dxdy)$$

$$= \int_{\Omega\times\mathbb{R}^n} \left[\phi_t(x,t) + v\cdot\nabla_x\phi(x,t)\right]\nu_t(dxdv) \ .$$

Since $\phi \in C_0^\infty(\Omega_I)$ we obtain:

$$0 = \int_0^1 \frac{d}{dt}\int_\Omega \phi\mu_t(dx) = \int_{\Omega_I\times\mathbb{R}^n} \left[\phi_t(x,t) + v\cdot\nabla_x\phi(x,t)\right]\nu(dxdvdt) \ .$$

On the other hand, by definition (.12) :

$$\int_{\Omega_I\times\mathbb{R}^n} |v|^p\nu(dxdvdt) = \int_{\Omega\times\Omega} |x-y|^p\lambda(dxdy) = d_p^p(\mu_0,\mu_1) \ .$$

This proves the first part of the Proposition (and, in particular, $\Lambda_p(\mu_0,\mu_1) \neq \emptyset$).

Let now ν be a lifting of $\mu \in \Lambda_p(\mu_0,\mu_1)$. We shall prove that $\mathbb{E}_\nu(|v|^p) \geq d_p^p(\mu_0,\mu_1)$. For this, we use the regularization Lemma 2.3 to approximate μ by smooth densities $\mu_n = \rho_n(x,t)dxdt$. Let $\mathbf{v}_n$ be the regularized velocity field. Then

$$\lim_{n\to\infty}\int_{\Omega_I} |\mathbf{v}_n|^p(x,t)\rho_n(x,t)dxdt = ||\mu||_p^p \ . \tag{.13}$$

Define

$$m_n(x,t) = \rho_n(x,t)\mathbf{v}_n(x,t) \ .$$

Then $m_n \in C^\infty(\Omega_I)$. Define now

$$\mathbf{v}_n^\varepsilon(x,t) = \frac{m_n(x,t)}{\rho_n(x,t)+\varepsilon}$$

By assumption, $\mathbf{v}_n^\varepsilon$ is Lipschitz on Ω_I, $t \in I$. Define $\rho_n^{(\varepsilon)}(x,t)$ as the solution of

$$\frac{\partial\rho_n^{(\varepsilon)}}{\partial t} + \nabla_x\left[\mathbf{v}_n^\varepsilon\rho_n^{(\varepsilon)}\right] = 0 \ \ ; \ \ \rho_n^{(\varepsilon)}(x,0) = \rho_n(x,0) \ . \tag{.14}$$

Since $\mathbf{v}_n^{(\varepsilon)}$ is Lipschitz, we may define the flow associated with it as $\Gamma^t_{(\varepsilon)} : \Omega \to \Omega$ for $t \in I$, namely $\Gamma^t_{(\varepsilon)}(x) = y_{(x)}(t)$ where $\dot{y}_{(x)} = \mathbf{v}_n^{(\varepsilon)}(y_{(x)}(t), t)$ and $y_{(x)}(0) = x$. It follows that $\Gamma^t_{(\varepsilon),\#}\rho_n(\cdot, 0)dx = \rho_n^{(\varepsilon)}(\cdot, t)dx$ for all $t \in I$. In particular:

$$d_p^p\left(\rho_n(x,0)dx, \rho_n^{(\varepsilon)}(x,1)dx\right) \leq \int_\Omega \rho_n(x,0)|x - \Gamma^1_{(\varepsilon)}(x)|^p dx = \int_\Omega \rho_n(x,0)\left|\int_0^1 \mathbf{v}_n^\varepsilon(y_{(x)}(t), t)dt\right|^p dx$$

$$\leq \int_{\Omega_I} \rho_n(x,0)\left|\mathbf{v}_n^\varepsilon\right|^p (\Gamma^t_{(\varepsilon)}(x), t)dtdx = \int_{\Omega_I} \rho_n^{(\varepsilon)}\left|\mathbf{v}_n^\varepsilon\right|^p dxdt \leq \int_{\Omega_I} \rho_n^{(\varepsilon)}\left|\mathbf{v}_n\right|^p dxdt \ . \tag{.15}$$

We next show that

$$\lim_{\varepsilon \to 0} \int_\Omega \left|\rho_n^{(\varepsilon)}(x,t) - \rho_n(x,t)\right| dx = 0 \tag{.16}$$

for any $t \in I$. In fact, we note that $\rho_n + \varepsilon$ solves equation (.14), hence $w_n^{(\varepsilon)} := \rho_n - \rho_n^{(\varepsilon)} + \varepsilon$ solves this equation as well. Since $w_n^{(\varepsilon)}(x,0) = \varepsilon > 0$ we obtain that $w_n^{(\varepsilon)} \geq 0$ over Ω_I and, moreover,

$$\int_\Omega \left|\rho_n(x,t) - \rho_n^{(\varepsilon)}(x,t)\right| dx - |\Omega|\varepsilon \leq \int_\Omega \left|w_n^{(\varepsilon)}(x,t)\right| = \int_\Omega \left|w_n^{(\varepsilon)}(x,0)\right| = |\Omega|\varepsilon$$

for all $t \in I$. Now we take first the limit $\varepsilon \to 0$ then the limit $n \to \infty$ in (.15), use (.16), (.13) and Lemma 2.3 to obtain the desired result.

References

[Am] L. Ambrosio: *Lectures Notes on Optimal Transport Problems*, CVGMT Preprint: http://cvgmt.sns.it/papers/ambooa/

[AFP] L. Ambrosio, N.Fusco & D.Pallara: *Functions of Bounded Variations and Free Discontinuity Problems*, Oxford University Press, 2000.

[B] Y. Brenier: *Polar factorization and monotone rearrangement of vector valued functions*, Arch. Rational Mech &Anal., 122, (1993), 323-351.

[BB] J.D.Benamou, Y. Brenier: *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer.Math., 84 (2000), 375-393.

[BBG] J.D.Benamou, Y. Brenier and K.Guitter: *The Monge-Kantorovich mass transfer and its computational fluid mechanics formulation*, Inter. J. Numer.Meth.Fluids, 40 (2002), 21-30.

[G,M] W. Gangbo an& R.J. McCann: *The geometry of optimal transportation*, Acta Math., 177 (1996), 113-161

[G,T] D.Gilbert and N.S. Trudinger: *Elliptic Partial Differential Equations of Second Order*, Springer-Verlag, 1983.

[M] Monge,G: *Mémoire sur la théorie des déblais et de remblais*, Histoire de l'Académie Royale des Sciences de Paris, 1781, pp. 666-704

[K] L Kantorovich: *On the translocation of masses*, C.R (Doclady) Acad. Sci. URSS (N.S), 37, (1942), 199-201