THE MONGE MASS TRANSFER PROBLEM AND ITS APPLICATIONS

Wilfrid Gangbo
School of Mathematics, Georgia Institute of Technology
Atlanta, GA 30332, USA
gangbo@math.gatech.edu *

Dec 1997

Abstract

We unify Kantorovich and Young's theory by formulating the Monge mass transfer problem as a variational problem involving Young measures. This is done thanks to the disintegration theorem and a density result on the set of all Borel measures on $\mathbf{R}^d \times \mathbf{R}^d$ with fixed marginals. We mention applications, such as Bernoulli convolution.

Contents

*The author gratefully acknowledges the support provided by NSF grant DMS 96-22734. AMS code: 49J40, 28A50.

Introduction

We present two main results on existence of optimal maps for the Monge mass transfer problem. The first one was obtained by Sudakov [27] and the second one by Evans and Gangbo [14]. We write the Monge-Kantorovich problem as a relaxation of the Monge problem by using Young measures generated by measure-preserving maps. This observation is based on the fact that given two Borel probability measures μ^+ and μ^- on $\mathbf{R}^d$ that have no atoms the set $\mathcal{M}$ of all Borel measures on $\mathbf{R}^d \times \mathbf{R}^d$ that have $\mu^\pm$ as their marginals is the closure of the set of all p in $\mathcal{M}$ whose support lies in the graph of a one-to-one map of $\mathbf{R}^d$ into $\mathbf{R}^d$.

The original Monge problem consists of finding the optimal way for rearranging a Borel probability measure μ^+ on $\mathbf{R}^d$ onto a Borel probability measure μ^- on $\mathbf{R}^d$ against the cost function $c(\mathbf{z}) = ||\mathbf{z}||$. The physical interpretation given by Monge ([24]) is that we are dealing with a pile of soil with a given mass distribution which we want to transport to an excavation, with a given distribution. The work involved by a particle of mass $d\mu^+(\mathbf{x})$ moving from a point $\mathbf{x}$ to a point $\mathbf{r}(\mathbf{x})$ along a smooth path $t \to \mathbf{g}(t, \mathbf{x})$ between time 0 and 1 is

$$\int_0^1 ||\dot{\mathbf{g}}(t, \mathbf{x})|| dt d\mu^+(\mathbf{x}).$$

Hence, the total work involved by all particles is

$$L[\mathbf{g}] := \int_{\mathbf{R}^d} \int_0^1 ||\dot{\mathbf{g}}(t, \mathbf{x})|| dt d\mu^+(\mathbf{x}),$$

where $\mathbf{r}$ satisfies the mass conservation condition

$$\mu^+[\mathbf{r}^{-1}(A)] = \mu^-[A], \tag{1}$$

for all $A \subset \mathbf{R}^d$ Borel. As observed by Monge for $\mathbf{g}$ satisfying (1), if we define $\bar{\mathbf{g}}(t, \mathbf{x}) := (1-t)\mathbf{x} + t\mathbf{r}(\mathbf{x})$ then $L[\bar{\mathbf{g}}] \leq L[\mathbf{g}]$ (Jensen's inequality). The Monge problem reduces then to the variational problem

$$\min_{\mathbf{r} \in \mathcal{A}} I[\mathbf{r}], \tag{2}$$

where

$$I[\mathbf{r}] := \int_{\mathbf{R}^d} ||\mathbf{x} - \mathbf{r}(\mathbf{x})|| d\mu^+(\mathbf{x}),$$

and $\mathcal{A}$ is the set of all Borel maps $\mathbf{r} : \mathbf{R}^d \to \mathbf{R}^d$ satisfying (1).

The set $\mathcal{A}$ on which we minimize the convex function I is not convex. Using the canonical imbedding i defined in (4) we can consider $\mathcal{A}$ as a subset of $\mathcal{M}$, the set of all Borel measures q on $\mathbf{R}^d \times \mathbf{R}^d$ satisfying

$$q[A \times \mathbf{R}^d] = \mu^+[A], \quad q[\mathbf{R}^d \times A] = \mu^-[A],$$

for all $A \subset \mathbf{R}^d$ Borel. Define $J : \mathcal{M} \to [0, +\infty]$ by

$$J[q] := \int_{\mathbf{R}^d \times \mathbf{R}^d} ||\mathbf{x} - \mathbf{y}|| dq(\mathbf{x}, \mathbf{y}), \quad (q \in \mathcal{M}).$$

Then

$$I[\mathbf{r}] = J[i(\mathbf{r})], (\mathbf{r} \in \mathcal{A}), \tag{3}$$

where $i : \mathcal{A} \to \mathcal{M}$ is the canonical imbedding defined by

$$i(\mathbf{r})[E] := \mu^+\{\mathbf{x} \in \mathbf{R}^d : \ (\mathbf{x}, \mathbf{r}(\mathbf{x})) \in E\}, \tag{4}$$

for $E \subset \mathbf{R}^d \times \mathbf{R}^d$ Borel. Let us observe that $\mathcal{M}$ is a subset of $\mathcal{P}(\mathbf{R}^d \times \mathbf{R}^d)$, the set of all Borel probability measures on $\mathbf{R}^d \times \mathbf{R}^d$, which is itself a subset of the topological dual space to $C_o(\mathbf{R}^d \times \mathbf{R}^d)$. Here $C_o(\mathbf{R}^d \times \mathbf{R}^d)$ denotes the set of all continuous functions on $\mathbf{R}^d \times \mathbf{R}^d$ which vanish at infinity under the sup norm. We say that a sequence $(q_n)_n \subset \mathcal{M}$ converges weak $*$ to $q \in \mathcal{M}$ if

$$\lim_{n \to +\infty} \int_{\mathbf{R}^d \times \mathbf{R}^d} F(\mathbf{x}, \mathbf{y}) dq_n(\mathbf{x}, \mathbf{y}) = \int_{\mathbf{R}^d \times \mathbf{R}^d} F(\mathbf{x}, \mathbf{y}) dq(\mathbf{x}, \mathbf{y}),$$

for all $F \in C_o(\mathbf{R}^d \times \mathbf{R}^d)$. We prove that the set $i(\mathcal{A})$ is a dense subset of $\mathcal{M}$ endowed with the weak $*$ topology (Proposition A.3).

Monge conjectured that I admits a minimiser $\mathbf{s} \in \mathcal{A}$ and there exists a potential $u : \mathbf{R}^d \to \mathbf{R}$ with Lipschitz constant less than or equal to 1 such that

$$u(\mathbf{x}) - u(\mathbf{s}(\mathbf{x})) = ||\mathbf{x} - \mathbf{s}(\mathbf{x})||, \quad (\mathbf{x} \in \mathbf{R}^d), \tag{5}$$

provided that the measures $\mu^\pm$ have compact supports and are absolutely continuous with respect to the Lebesgue measure.

For the reader's convenience we recall two definitions needed in the sequel.

Definition 0.1 *Let X be a metric space and let μ be a positive, finite Borel measure on X. The support of μ is the smallest closed set $spt(\mu) \subset X$ such that $\mu(spt(\mu)) = \mu(X)$.*

Definition 0.2 *We say that $M \subset \mathbf{R}^d$ is a $(d-1)$-rectifiable set if M is a countable union of C^1 $(d-1)$-hypersurfaces and sets of zero $(d-1)$-dimensional Hausdorff measure.*

Young measures and Kantorovich approach. The first rigorous proof on the existence of a potential u associated to the Monge problem was given by Kantorovich when he introduced the following variational problem ([18], [19]) which is referred to as the Monge-Kantorovich problem: find $p \in \mathcal{M}$ such that

$$J[p] = \min_{q \in \mathcal{M}} J[q] := d_W(\mu^+, \mu^-). \tag{6}$$

Kantorovich's formulation is similar to Young's when he introduced the generalized functions (or parametrized Young measures [30]), a concept which is very useful in the Calculus of Variations, Partial Differential Equations ([28]) etc... For each $q \in \mathcal{M}$ we may find a family $(q_{\mathbf{x}})_{\mathbf{x} \in \mathbf{R}^d}$ of probability measures on $\mathbf{R}^d$ such that

$$\int_{\mathbf{R}^d \times \mathbf{R}^d} F dq = \int_{\mathbf{R}^d} [\int_{\mathbf{R}^d} F(\mathbf{x}, \mathbf{y}) dq_{\mathbf{x}}(\mathbf{y})] d\mu^+(\mathbf{x}), \tag{7}$$

for all Borel q-summable functions $F : \mathbf{R}^d \times \mathbf{R}^d \to [-\infty, +\infty]$. Using that $q \in \mathcal{M}$ we deduce that

$$\mu^-[A] = \int_{\mathbf{R}^d} q_{\mathbf{x}}[A] d\mu^+(\mathbf{x}), \tag{8}$$

for all $A \subset \mathbf{R}^d$ Borel. Conversely, one can readily check that given a family $(q_{\mathbf{x}})_{\mathbf{x} \in \mathbf{R}^d}$ of probability measures on $\mathbf{R}^d$ satisfying (8) the measure q defined by (7) is an element of $\mathcal{M}$.

The function d_W in (6), which is a metric on the set of probability measures (see [11]) and known as the Wasserstein-Rubinstein distance has been of great use in various fields such as Partial Differential Equations ([7]), Material Sciences ([6]), Probability ([25]), Functional Analysis ([1], ([20]), etc...

Writing (6) as an infinite dimensional linear programming minimization problem under the assumption that the measures $\mu^\pm$ have compact supports Kantorovich obtained a dual problem

$$\max_{w \in \mathcal{L}} K[w], \tag{9}$$

where,

$$K[w] := \int_{\mathbf{R}^d} w d\mu^+ - \int_{\mathbf{R}^d} w d\mu^-,$$

and,

$$\mathcal{L} := \{w : \mathbf{R}^d \to \mathbf{R} : \; Lip(w) := \sup_{\mathbf{x} \neq \mathbf{y}} \frac{|w(\mathbf{x}) - w(\mathbf{y})|}{||\mathbf{x} - \mathbf{y}||} \leq 1\}. \tag{10}$$

Since $\mu^+[\mathbf{R}^d] = \mu^-[\mathbf{R}^d]$ the supremum of K over $\mathcal{L}$ coincides with the supremum of K over the subset of all $w \in \mathcal{L}$ satisfying $w(0) = 0$. One can readily deduce that (9) admits a maximizer $u \in \mathcal{L}$. Observe that the duality relation between (6) and (9) implies

$$||\mathbf{x} - \mathbf{y}|| = u(\mathbf{x}) - u(\mathbf{y}), \quad \text{for } p \text{ a.e. } (\mathbf{x}, \mathbf{y}) \in \mathbf{R}^d \times \mathbf{R}^d \tag{11}$$

where p is any minimiser of (6). It is a known fact that existence of a minimiser of (6), a maximizer for (9), the duality between (6) and (9) still hold even if we don't impose that the supports of $\mu^\pm$ are compact but, assume that the first moments of the measures are bounded i.e.

$$\int_{\mathbf{R}^d} ||\mathbf{x}|| d\mu^+(\mathbf{x}), \ \int_{\mathbf{R}^d} ||\mathbf{x}|| d\mu^-(\mathbf{x}) < +\infty, \tag{12}$$

(see [21] and [25]). Note that (12) implies that J takes only finite values on $\mathcal{M}$.

If $\mu^\pm$ have no atoms then $i(\mathcal{A})$ is dense in $\mathcal{M}$ (see Proposition A.3), and using (3) and (12) we obtain

$$\min_{\mathbf{r} \in \mathcal{A}} I[\mathbf{r}] = \min_{q \in \mathcal{M}} J[q]. \tag{13}$$

The Monge-Kantorovich problem is then obtained as a relaxation of the Monge problem.

Extreme points of $\mathcal{M}$. Since J is a linear functional its minimum over $\mathcal{M}$ is achieved at an extreme point of the compact (with respect to the weak $*$ topology) convex set $\mathcal{M}$. If the supports of $\mu^\pm$ have both n elements then by the well-known Birkhoff–von Neumann theorem the set of extreme points of $\mathcal{M}$ is $i(\mathcal{A})$ and so, J achieves its minimum at some point $i(\mathbf{s})$. Clearly, $\mathbf{s}$ is a minimiser of the Monge problem. But, if $\mu^\pm$ have no atoms $i(\mathcal{A})$ is strictly included in the set of extreme points of $\mathcal{M}$ (see Remark C.2 and Corollary C.3).

Geometry of supports of optimal Young measures. Assume that the Borel probability measures $\mu^\pm$ have bounded, disjoint supports and denote by $X := \operatorname{spt} \mu^+$, $Y := \operatorname{spt} \mu^-$. Assume that ∂X and ∂Y are smooth. Then a maximizer u for (9) can be chosen to satisfy

$$u(\mathbf{x}) = \inf_{\mathbf{y} \in Y} \{||\mathbf{x} - \mathbf{y}|| + u(\mathbf{y})\}, \quad (\mathbf{x} \in O_X), \tag{14}$$

$$u(\mathbf{x}) = \sup_{\mathbf{x} \in X} \{u(\mathbf{x}) - ||\mathbf{x} - \mathbf{y}||\}, \quad (\mathbf{x} \in O_Y), \tag{15}$$

where O_X a neighborhood of X and O_Y a neighborhood of Y, and so, u is semiconcave on X, semiconvex on Y. Thus, u is differentiable at every point of $X \cup Y$ except may be on a set which is $(d-1)$-rectifiable and

$$Du(\mathbf{x}_o) = \frac{\mathbf{x}_o - \mathbf{y}_o}{||\mathbf{x}_o - \mathbf{y}_o||}, \tag{16}$$

whenever $Du(\mathbf{x}_o)$ exists and $u(\mathbf{x}_o) = ||\mathbf{x}_o - \mathbf{y}_o|| + u(\mathbf{y}_o)$. Let p be a minimiser of the Monge-Kantorovich problem (6) and let $(p_{\mathbf{x}})_{\mathbf{x}\in\mathbf{R}^d}$ the Young measures associated to p as in (7). If in addition we assume that $\mu^{\pm}$ vanish on $(d-1)$-rectifiable sets then (16) implies that $\operatorname{spt} p_{\mathbf{x}}$ is contained in a line segment through $\mathbf{x}$, parallel to $Du(\mathbf{x})$, i.e.,

$$\operatorname{spt} p_{\mathbf{x}} \subset R_{\mathbf{x}} := \{\mathbf{x} - \lambda Du(\mathbf{x}) : \quad \lambda \in [\lambda_o(\mathbf{x}), \lambda_1(\mathbf{x})]\}, \tag{17}$$

for μ^+-almost every $\mathbf{x} \in \mathbf{R}^d$. We call $R_{\mathbf{x}}$ the transport ray through $\mathbf{x}$, which is nothing but, the set $\{\mathbf{y} \in \mathbf{R}^d : \quad |u(\mathbf{x}) - u(\mathbf{y})| = ||\mathbf{x} - \mathbf{y}||\}$, and is a line segment for μ^+-almost every $\mathbf{x} \in \mathbf{R}^d$. Conversely, if a family $(p_{\mathbf{x}})_{\mathbf{x}\in\mathbf{R}^d}$ of probability measures on $\mathbf{R}^d$ satisfies (8) and (17) then p defined in (7) is a minimiser of the Monge-Kantorovich problem (6).

In [14] when the Borel measures are absolutely continuous with respect to the d-dimensional Lebesgue measure, $\mu^{\pm} = f^{\pm} d\mathbf{x}$, $f^{\pm}$ are Lipschitz and positive on the interior of their compact supports, an ODE

$$\begin{cases} \dot{g}_\delta(t, \mathbf{x}) & = & \mathbf{v}_\delta(t, g_\delta(t, \mathbf{x})) \\ \mathbf{g}_\delta(0, \mathbf{x}) & = & \mathbf{x}, \end{cases} \tag{18}$$

was identified such that $g_\delta(t, \mathbf{x})$ belongs to the transport ray $R_{\mathbf{x}}$, for μ^+-almost every $\mathbf{x} \in \mathbf{R}^d$,

$$g_\delta(1, \mathbf{x}) \;\text{ pushes }\; (\mu^+ + \delta) d\mathbf{x} \text{ forward to } (\mu^- + \delta) d\mathbf{x}, \tag{19}$$

and the limiting map

$$\mathbf{s}(\mathbf{x}) := \lim_{\delta \to 0^+} g_\delta(1, \mathbf{x})$$

exists. We have $\mathbf{s}(\mathbf{x}) \in R_{\mathbf{x}}$ for μ^+-almost every $\mathbf{x} \in \mathbf{R}^d$ and by (19) pushes $\mu^+ d\mathbf{x}$ forward to $\mu^- d\mathbf{x}$. Consequently, $\mathbf{s}$ is a minimiser of the Monge problem. To obtain (19) we introduce the following approximate variational problem:

$$\sup_{w \in W_o^{1,p}(B_R)} K_p[w], \tag{20}$$

where B_R is an open ball of center 0 and radius $R > 0$, large enough to contain the supports of $\mu^\pm$, $p > 1$ and

$$K_p[w] := K[w] - \frac{1}{p}||Dw||^p_{L^p(B_R)}.$$

Clearly, (20) admits a unique maximizer u_p, solution to the p-Laplacian equation

$$\begin{cases} -div(||Du_p||^{p-2}Du_p) &= f^+ - f^- \quad \text{in} \quad B_R \\ u_p &= 0 \quad \text{on} \quad \partial B_R, \end{cases} \tag{21}$$

in the weak sense. The sequence $(u_p)_{p\geq d+1}$ is bounded in $W^{1,d+1}_o(B_R)$ and so, we may extract a subsequence (u_{p_k}) converging uniformly to some $u \in W^{1,d+1}_o(B_R)$. It is straightforward that $u \in \mathcal{L}$ and is a maximizer for (9). Using that $f^\pm$ are Lipschitz functions, bounded below by a positive constant on their compact supports a careful analysis yields the sequence $(||Du_p||^{p-2})$ is bounded in $L^\infty(B_R)$, converges to some $a \in L^\infty(B_R)$ weak $*$, and letting p go to infinity in (21) yields

$$\begin{cases} -div(aDu) &= f^+ - f^- \quad \text{in} \quad B_R \\ u &= 0 \quad \text{on} \quad \partial B_R, \end{cases} \tag{22}$$

We may interpret the density function a as a Lagrange multiplier for (9). Note that (22) is a continuity equation of the form

$$\frac{\partial \rho}{\partial t} + div(\rho \mathbf{v}) = 0, \tag{23}$$

where

$$\rho(t, \mathbf{z}) = (1-t)f^+(\mathbf{z}) + tf^-(\mathbf{z})$$

and

$$\mathbf{v}(t, \mathbf{z}) = \frac{-a(\mathbf{z})Du(\mathbf{z})}{(1-t)f^+(\mathbf{z}) + tf^-(\mathbf{z})} \tag{24}$$

is defined for γ-almost every $(t, \mathbf{z}) \in [0,1] \times \mathbf{R}^d$, γ being the measure on $[0,1] \times \mathbf{R}^d$ defined by

$$\gamma[B] = \int_B \rho(t, \mathbf{z}) dt d\mathbf{z},$$

for all Borel sets $B \subset [0,1] \times \mathbf{R}^d$. We approximate $\mathbf{v}$ by velocities $\mathbf{v}_\delta$, $(\delta > 0)$

$$\mathbf{v}_\delta(t, \mathbf{z}) = \frac{-a(\mathbf{z})Du(\mathbf{z})}{(1-t)f^+(\mathbf{z}) + tf^-(\mathbf{z}) + \delta},$$

and ρ by densities

$$\rho_\delta(t, \mathbf{z}) = (1-t)f^+(\mathbf{z}) + tf^-(\mathbf{z}) + \delta.$$

Note that the following continuity equation holds:

$$\frac{\partial \rho_\delta}{\partial t} + div(\rho_\delta \mathbf{v}_\delta) = 0.$$

If aDu is smooth then the flow $\mathbf{g}_\delta$ defined in (18) is such that $\mathbf{s}_\delta := \mathbf{g}(1, \cdot)$ pushes $(\mu^+ + \delta)d\mathbf{x}$ forward to $(\mu^- + \delta)d\mathbf{x}$. Also, for any point $\mathbf{x}$ where u is differentiable Du is constant along the ray $R_\mathbf{x}$ and so, the solution in (18) is unique once we impose that $\mathbf{g}(t, x) \in R_\mathbf{x}$. Observe that $a \geq 0$ implies $(\mathbf{s}_\delta(\mathbf{x}))_{0<\delta<1}$ is monotonically rearranged along $R_\mathbf{x}$ and $\mathbf{s}(\mathbf{x}) := \lim_{\delta \to 0^+} \mathbf{s}_\delta(\mathbf{x})$ exists. It is straightforward to check that $\mathbf{s}$ pushes $\mu^+ d\mathbf{x}$ forward to $\mu^- d\mathbf{x}$, and $\mathbf{s}(\mathbf{x}) \in R_\mathbf{x}$ i.e.

$$u(\mathbf{x}) - u(\mathbf{s}(\mathbf{x})) = ||\mathbf{x} - s(\mathbf{x})|| \quad \text{for } \mu^+ - \text{a.e.} \quad \mathbf{x} \in \mathbf{R}^d. \tag{25}$$

Using (25) and the fact that (2) and (6) are dual we obtain that $\mathbf{s}$ is a minimiser of the Monge problem (2). Unfortunately, aDu is not known to be smooth and we must approximate aDu, f^+, and f^- by smooth functions $(aDu)_\epsilon$, f^+_ϵ, and f^-_ϵ such that

$$-div(aDu)_\epsilon = f^+_\epsilon - f^-_\epsilon.$$

Accordingly, we introduce the velocities

$$\mathbf{v}_{\delta,\epsilon}(t, \mathbf{z}) = \frac{(aDu)_\epsilon(\mathbf{z})}{(1-t)f^+_\epsilon(\mathbf{z}) + tf^-_\epsilon(\mathbf{z}) + \delta},$$

and the flow

$$\dot{g}_{\delta,\epsilon}(t, \mathbf{x}) = \mathbf{v}_{\delta,\epsilon}(t, g_{\delta,\epsilon}(t, \mathbf{x})), \quad g_{\delta,\epsilon}(0, \mathbf{x}) = \mathbf{x}$$

which satisfies $\mathbf{s}_{\delta,\epsilon} := \mathbf{g}_{\delta,\epsilon}(1, \cdot)$ pushes $(f^+_\epsilon + \delta)d\mathbf{x}$ forward to $(f^-_\epsilon + \delta)d\mathbf{x}$. Proving that $\mathbf{s}_{\delta,\epsilon}(\mathbf{x})$ converges for μ^+-almost every $\mathbf{x}$ to $\mathbf{s}_\delta(\mathbf{x})$ as ϵ goes to 0 requires that we know fine properties of the restriction of the functions a and u to the neighborhood of transport rays. By (14) and (15) u is semiconcave in a neighborhood of X, semiconvex in a neighborhood of Y and so D^2u is a Radon measure on $X \cup Y$. Using (22) we obtain that for almost every $\mathbf{x} \in X$ the restriction of a to the transport ray $R_\mathbf{x}$ is locally Lipstchitz and satisfies

$$-(a_\mathbf{n} + a[\triangle u]_{ac}) = f^+ - f^- \quad H^1 a.e. \text{ on } X \tag{26}$$

$$-(a_{\mathbf{n}} + a[\triangle u]_{ac}) = f^+ - f^- \quad H^1 a.e. \text{ on } Y \tag{27}$$

where $\mathbf{n} = Du(\mathbf{x})$, H^1 stands for the one-dimensional Hausdorff measure, and $[\triangle u]_{ac}$ for the trace of the absolutely continuous part of D^2u (see [14]). We may interpret (26) and (27) as $-(Da \cdot Du + a[\triangle u]_{ac}) = f^+ - f^-$ which is a formal way of writing (22). Both (26) and (27) are used to prove that a vanishes at the endpoints of transport rays $R_{\mathbf{x}}$ for μ^+-almost every $\mathbf{x} \in \mathbf{R}^d$.

The remainder of the paper is organized as follows. In section 1 we describe how the dual problem (9) is obtained via the p-Laplacian and recall properties of the density function a occuring as a Lagrange multiplier for (9) and the potential function u maximizer in (9). In section 2 we state Sudakov's result ([27]) and construct a minimiser of (2) as done in [14]. In section 3 we use the Wasserstein-Rubinstein distance to study Bernoulli's convolution. In this paper we include an appendix consisting of three parts. In Appendix A we prove that if X and Y are two uncountable complete metric spaces, μ^+, is a finite Borel measure on X, μ^- is a finite Borel measure on Y, $\mu^\pm$ have no atoms and $\mu^+[X] = \mu^-[Y] > 0$ then every Borel measure γ on $X \times Y$ having $\mu^\pm$ as their marginals can be obtained as the weak $*$ limit of a sequence of the form $\{i(\mathbf{r}_n)\}$ where $\mathbf{r}_n : X \to Y$ are Borel maps that push μ^+ forward to μ^- and are one-to-one. Here $i(\mathbf{r})[E] := \mu^+\{\mathbf{x} \in \mathbf{R}^d : \ (\mathbf{x}, \mathbf{r}(\mathbf{x})) \in E\}$, for $E \subset X \times Y$ Borel set. In Appendix B we obtain a generalization of Fubini's theorem as a straightforward application of the disintegration of measures, a very useful tool in ergodic theory which goes back to von Neumann. In the last Appendix C we prove that in general the extreme points of the set of all Borel measures that have $\mu^\pm$ as their marginal is wider than $i(\mathcal{A})$ where $\mathcal{A}$ is the set of all Borel maps that push μ^+ forward to μ^-.

The author would like to thank M. Milman and the organisers of the CBM-NSF conference The Monge-Ampère Equation: Application to Geometry and Optimisation for their hospitality and financial support. He thanks the referee for stylistic improvement of the manuscript. He would like also to thank C. Heil, C. Houdré for references, and A. Swiech for comments on the paper.

1 Duality

In this section we study a linear maximization problem dual to the Monge problem. A detailed proof of the results stated in this section can be found in [14]. Given $f \in L^{\infty}(\mathbf{R}^d)$

$$f \text{ compactly supported} \tag{28}$$

find $u \in \mathcal{L}$ solving

$$K[u] = \max_{w \in \mathcal{L}} K[w], \tag{29}$$

where,

$$K[w] := \int_{\mathbf{R}^d} f w d\mathbf{x},$$

and $\mathcal{L}$ is the subset of Lipschitz functions defined in (10). To ensure that the supremum in (29) is finite we impose that

$$\int_{\mathbf{R}^d} f d\mathbf{x} = 0. \tag{30}$$

Since (30) implies

$$K[w + c] = K[w], \tag{31}$$

for all $w \in \mathcal{L}$ and all $c \in \mathbf{R}$, we deduce that if we choose $S > 0$, such that

$$B(0, S) \text{ contains the support of } f, \tag{32}$$

then

$$\max_{w \in \mathcal{L}} K[w] = \max_{w \in W_o^{1,\infty}(B(0,R))} K[w], \tag{33}$$

for all $R > 2S$. We next consider a family of variational problems related to (29): find $u_p \in W_o^{1,p}(B(0,R))$ such that

$$K_p[u_p] = \max_{w \in W_o^{1,p}(B(0,R))} K_p[w], \tag{34}$$

where

$$K_p[w] := K[w] - \frac{||Dw||_p^p}{p}.$$

Note that

$$\lim_{p \to +\infty} K_p[w] = \begin{cases} K[w] & \text{if} \quad w \in \mathcal{L} \\ -\infty & \text{if} \quad w \notin \mathcal{L}, \end{cases} \tag{35}$$

u_p is the unique solution to the PDE

$$\begin{cases} -div(||Du_p||^{p-2} Du_p) & = & f & \text{in} & B(0,R) \\ u_p & = & 0 & \text{on} & \partial B(0,R), \end{cases} \tag{36}$$

in the weak sense. In addition $u_p \in C^{1,\alpha}(B(0,R))$ for some $\alpha \equiv \alpha(p)$, according to [23] and [29]. Since

$$-div(||Du_p||^{p-2} Du_p) = 0 \quad \text{in } B(0,R) \setminus B(0,S) \tag{37}$$

we deduce that there exists $\mathbf{x}_o \in \partial B(0,S)$ such that

$$u_p(\mathbf{x}_o) = 0. \tag{38}$$

Multiplying (36) by $u_p - \frac{1}{|B(0,S)|}\int_{B(0,S)} u_p d\mathbf{x}$, integrating by parts, using (30) and Poincaré-Wirtinger's inequality on $B(0,S)$ we deduce that $(||Du_p||^p)$ is bounded in $L^1(B(0,R))$ by a constant which depends only on $||f||_\infty$ and S. This, together with (38) and Sobolev's Imbedding Theorem implies

$$\max_{B(0,S)} |u_p| \leq C_1, \tag{39}$$

where C_1 is a constant depending only on $||f||_\infty$ and S. Thanks to the maximum principle (37) and (39) imply C_1 is a bound for u_p over $B(0,R)$. To have a better estimate on $||Du_p||$ useful for the Monge problem we assume in addition that

$$f \text{ is a Lipschitz function} \tag{40}$$

to obtain the following conclusions.

Lemma 1.1 *Assume (28), (30) and (40) hold. Then there exist a constant $C > 0$ and a large radius $R > 0$ such that*

$$\max_{B(0,R)} |u_p| \leq C$$

and

$$\max_{B(0,R)} ||Du_p||^p \leq C$$

for $d+1 \leq p < +\infty$.

In light of Lemma 1.1 we can extract a subsequence $p_k \to +\infty$ so that

$$u_{p_k} \to u \text{ uniformly in } B(0,R), \tag{41}$$

$$Du_{p_k} \to Du \text{ weak } * \text{ in } B(0,R), \tag{42}$$

$$||Du_{p_k}||^{p_k-2} \to a \text{ weak } * \text{ in } B(0,R), \tag{43}$$

where $u \in W_o^{1,\infty}(B(0,R)) \cap \mathcal{L}$, $a \in L^\infty(B(0,R))$ and $a \geq 0$. The first main result of this section is obtained using (41)–(43).

Theorem 1.2 *Assume (28), (30) and (40) hold. Then there exists a radius $R > 0$ large enough such that*

(i)

$$-div(aDu) = f \quad in \;\; B(0, R)$$

in the weak sense,

$$for\, a.e. \mathbf{x} \in B(0, R) \quad a(\mathbf{x}) > 0 \;\; implies \;\; ||Du(\mathbf{x})|| = 1.$$

(ii)

$$K[u] = \max_{w \in \mathcal{L}} K[w].$$

For technical reasons we assume henceforth that $R > 0$ is large enough and

$$\begin{cases} X \cap Y & = & \emptyset \\ \partial X, \partial Y & & \text{are smooth} \\ f \neq 0 & & \text{in the interior of its support,} \end{cases} \tag{44}$$

where X is the support of $f^+ := \max\{0, f\}$ and Y is the support of $f^- := \max\{0, -f\}$. We introduce the transport set which is a compact set containing $X \cup Y$.

$$T := \{\mathbf{z} \in B(0, R) \;:\; u_*(\mathbf{z}) = u(\mathbf{z}) = u^*(\mathbf{z})\},$$

where

$$u_*(\mathbf{z}) := \max_{\mathbf{x} \in X} \{u(\mathbf{x}) - ||\mathbf{z} - \mathbf{x}||\} \quad (\mathbf{z} \in B(0, R))$$

and

$$u^*(\mathbf{z}) := \min_{\mathbf{y} \in Y} \{u(\mathbf{y}) + ||\mathbf{z} - \mathbf{y}||\} \quad (\mathbf{z} \in B(0, R)).$$

Since $||Du||_\infty \leq 1$ we have $u_* \leq u \leq u^*$ on $B(0, R)$. Using that u is a maximizer for (29), $f^+ > 0$ on X^o, $f^- > 0$ on Y^o, we deduce that $u_* = u = u^*$ on $X \cup Y$. The transport set T is made of transport rays

$$R_{\mathbf{z}_o} := \{\mathbf{z} \in B(0, R) \;:\; |u(\mathbf{z}_o) - u(\mathbf{z})| = ||\mathbf{z}_o - \mathbf{z}||\}. \tag{45}$$

These transport rays $R_{\mathbf{z}_o}$ are line segments with endpoints $a_o(\mathbf{z}_o) \in X$, $b_o(\mathbf{z}_o) \in Y$ (see (14)–(17)) whenever $Du(\mathbf{z}_o)$ exists and $\mathbf{z}_o \in T$.

Proposition 1.3 *Assume (28), (30), (40) and (44) hold. Then, $a \equiv 0$ on $B(0, R) \setminus T$.*

For $\mathbf{z}_o \in T$ such that $R_{\mathbf{z}_o}$ is a line segment with endpoints $a_o(\mathbf{z}_o), b_o(\mathbf{z}_o)$ we define

$$R^\sigma_{\mathbf{z}_o} := R_{\mathbf{z}_o} \setminus [B(a_o(\mathbf{z}_o),\sigma) \cup B(b_o(\mathbf{z}_o),\sigma)], \quad (\sigma > 0).$$

The restriction of the function u to such a transport ray grows with rate one and so Du is constant along the ray. The following theorem improves the smoothness property of Du is a neighborhood of $R^\sigma_{\mathbf{z}_o}$.

Proposition 1.4 *Assume (28), (30), (40) and (44) hold. Then, for each transport ray $R_{\mathbf{z}_o}$ such that u is differentiable at $\mathbf{z}_o \in T$ and $\sigma > 0$ there exists a constant $C_\sigma > 0$ and a tubular neighborhood N of $R^\sigma_{\mathbf{z}_o}$ such that*

$$||Du(\mathbf{z}) - Du(\hat{\mathbf{z}})|| \leq C_\sigma ||\mathbf{z} - \hat{\mathbf{z}}||$$

for each $\mathbf{z} \in N \cap T$ at which $Du(\mathbf{z})$ exists. Here, $\hat{\mathbf{z}}$ denotes the projection of $\mathbf{z}$ onto $R_{\mathbf{z}_o}$.

Proposition 1.4 is also used to obtain the following result.

Proposition 1.5 *Assume (28), (30), (40) and (44) hold. Then, for a.e.$\mathbf{z}_o \in T$*

(i) $a_{|R_{\mathbf{z}_o}}$ is locally Lipschitz along $R_{\mathbf{z}_o}$.

(ii) $a_{|R_{\mathbf{z}_o} \cap X^o}$ and $a_{|R_{\mathbf{z}_o} \cap Y^o}$ are both strictly positive and vanish at the endpoints of $R_{\mathbf{z}_o}$.

Let E be the set of all $\mathbf{z} \in X \cup Y$ such that $\mathbf{z}$ is an endpoint for some transport ray.

Proposition 1.6 *Assume (28), and (40) hold. Then, $|E| = 0$ i.e. the d-dimensional Lebesgue measure of E is zero.*

2 Existence of Optimal Maps

Throughout this section we assume that $\mu^\pm$ are Borel probability measures on $\mathbf{R}^d$

$$\text{spt}\,(\mu^\pm) \subset B(0,S), \tag{46}$$

$\mu^\pm$ are absolutely continuous with respect to the d-dimensional Lebesgue measure

$$\mu^\pm = f^\pm d\mathbf{x}, \tag{47}$$

and we define

$$f := f^+ - f^-.$$

We recall that X is the support of μ^+ and Y is the support of μ^-. The first main result of this section is the existence of an optimal map for the Monge problem under (46) and (47). The second main result is the identification of an ODE to construct an optimal solution when (46), (47) hold and $f^\pm$ are smooth. Using the same notations as in section 1 we recall that $S > 0$ is chosen so that (32) holds and $R > 2S$ is large enough.

Theorem 2.1 *[Sudakov] Take $\mu^\pm$ so that (46) and (47) hold. Then there exists an optimal solution to the Monge problem (2).*

Sketch of proof We refer the reader to [27] for details.
Clearly, the dual problem (9) admits a maximizer $u \in \mathcal{L}$. For each $\mathbf{x} \in \mathbf{R}^d$ recall that the transport ray through $\mathbf{x}$ is

$$R_\mathbf{x} = \{\mathbf{y} \in B(0,R) \; : \; |u(\mathbf{x}) - u(\mathbf{y})| = ||\mathbf{x} - \mathbf{y}||\}.$$

If u is differentiable at $\mathbf{x}$, then $R_\mathbf{x}$ is either a single point or a line segment. Except for $\mathbf{x} \in M$ where M is a set of d-dimensional Lebesgue measure zero, $R_\mathbf{x}$ is a convex set, and is contained in a level set of Du. Thus $(R_\mathbf{x})_{\mathbf{x}\in B(0,R)}$ is an affine decomposition of $B(0,R)$, and so, the conditional measures on $R_\mathbf{x}$ of $\mu^\pm$ are absolutely continuous with respect to the 1-dimensional Lebesgue measure on $R_\mathbf{x}$. This reduces the Monge problem to a transport problem on a straight line where the measures involved are absolutely continuous with respect to the 1-dimensional Lebesgue measure. The *one*-dimensional problem is known to admit a solution. QED.

An alternative method to Sudakov's is next presented. Let us recall that the transport set T introduced in section 1 is a compact set and is the union of the transport rays $R_\mathbf{x}$. A PDE is identified to first reduce the Monge problem to *one*-dimensional transport problem analogously to Sudakov's decomposition of measures theory. Then an ODE is identified to solve the *one*-dimensional transport problems. To avoid technical difficulties we assume that $f^\pm$ are Lipschitz, ∂X, ∂Y are smooth, X and Y don't intersect. We also assume f does not vanish in the interior of its compact support $X \cup Y$ (see (44)). The first proposition asserts that the conditional measures on $R_{\mathbf{x}}$ of $\mu^{\pm}$ are absolutely continuous with respect to the one-dimensional Hausdorff measure H^1 for almost every $\mathbf{x} \in B(0, R)$.

Proposition 2.2 *Take $\mu^{\pm}$ so that $f^{\pm}$ are Lipschitz functions and assume that (44) holds. Let $N \subset B(0, R)$ be a set of d-dimensional Lebesgue measure zero. Then $H^1(R_{\mathbf{x}} \cap N) = 0$ for H^d- almost every $\mathbf{x} \in B(0, R)$.*

Proof: see [14] QED.

Theorem 2.3 *[Evans & Gangbo] Take $\mu^{\pm}$ so that (46), (47) hold and $f^{\pm}$ are Lipschitz functions. Assume furthermore that (44) holds. Let a and u be two functions as obtained in Theorem 1.2. Define the flow $(\mathbf{g}_\delta)$ solution of the ODE*

$$\begin{cases} \dot{\mathbf{g}}_\delta(t, \mathbf{x}) & = & \mathbf{b}_\delta(t, \mathbf{g}_\delta(t, \mathbf{x})) \\ \mathbf{g}_\delta(0, \mathbf{x}) & = & \mathbf{x} \end{cases}$$

where $\mathbf{b}_\delta(t, \mathbf{z}) := \frac{a(\mathbf{z})\mathbf{n}}{(1-t)f^+(\mathbf{z})+tf^-(\mathbf{z})+\delta}$ and $\mathbf{n} := -Du(\mathbf{x})$. Define $\mathbf{s}_\delta(\mathbf{x}) := \mathbf{g}_\delta(1, \mathbf{x})$. Then $\lim_{\delta \to 0^+} \mathbf{s}_\delta(\mathbf{x}) := \mathbf{s}(\mathbf{x})$ exists for H^d-almost every $\mathbf{x} \in X$. Futhermore $\mathbf{s}$ is an optimal solution to the Monge problem (2).

Sketch of proof We refer the reader to [14] for details.
1. If $\mathbf{x} \in T$ is such that $Du(\mathbf{x})$ then u is differentiable at every point in the relative interior of the ray $R_{\mathbf{x}}$ and Du is constant along $R_{\mathbf{x}}$ (see (14), (16)). By Proposition 1.6 we may as well assume that $\mathbf{x}$ is not an endpoint and so, thanks to Proposition 1.5 a and $\mathbf{b}_\delta(t, \dot{)}$ restricted to $R_{\mathbf{x}}$ are Lipschitz functions in a neighborhood of $\mathbf{x}$. Hence, the ODE (22) is well-defined and $\mathbf{g}_\delta(t, \mathbf{x})$ is uniquely determined. Using Proposition 1.5 again we obtain that a vanishes at the endpoint of $R_{\mathbf{x}}$ and so, $\mathbf{g}_\delta(t, \mathbf{x})$ remains in T.

2. We approximate aDu, f^+, and f^- by smooth functions $(aDu)_\epsilon$, f^+_ϵ, and f^-_ϵ such that

$$-div(aDu)_\epsilon = f^+_\epsilon - f^-_\epsilon.$$

Define

$$\mathbf{v}_{\delta,\epsilon}(t, \mathbf{z}) := \frac{-(aDu)_\epsilon(\mathbf{z})}{(1-t)f^+_\epsilon(\mathbf{z}) + tf^-_\epsilon(\mathbf{z}) + \delta}$$

and let $\mathbf{g}_{\delta,\epsilon}$ solve the ODE

$$\begin{cases} \dot{\mathbf{g}}_{\delta,\epsilon}(t,\mathbf{x}) & = & \mathbf{v}_{\delta,\epsilon}(t,\mathbf{g}_{\delta,\epsilon}(t,\mathbf{x})) \\ \mathbf{g}_{\delta,\epsilon}(0,\mathbf{x}) & = & \mathbf{x} \end{cases}$$

Clearly, $\mathbf{s}_{\delta,\epsilon} := \mathbf{g}_{\delta,\epsilon}(1,\cdot)$ pushes $(f_\epsilon^+ + \delta)d\mathbf{x}$ forward to $(f_\epsilon^- + \delta)d\mathbf{x}$. Using Proposition 1.3, 1.4 and the fact that $(aDu)_\epsilon$, f_ϵ^+, and f_ϵ^- converge to aDu, f^+, and f^- almost everywhere as ϵ tends to 0 we deduce that $\lim_{\epsilon\to 0^+} \mathbf{s}_{\delta,\epsilon}(\mathbf{x})$ exists and coincides with $\mathbf{s}_\delta(\mathbf{x})$ for almost every $\mathbf{x} \in T$ and so, $\mathbf{s}_\delta$ pushes $(f^+ + \delta)d\mathbf{x}$ forward to $(f^- + \delta)d\mathbf{x}$. Since $a, f^+, f^- \geq 0$ we obtain that $(\mathbf{s}_\delta(\mathbf{x}))_{0\leq\delta\leq 1}$ is monotonically arranged along $R_\mathbf{x}$ and $\mathbf{s}(\mathbf{x}) := \lim_{\delta\to 0^+} \mathbf{s}_\delta(\mathbf{x})$ exists and belongs to $R_\mathbf{x}$. Hence $\mathbf{s}$ pushes $f^+ d\mathbf{x}$ forward to $f^- d\mathbf{x}$. Since in addition

$$u(\mathbf{x}) - u(\mathbf{s}(\mathbf{x})) = ||\mathbf{x} - \mathbf{s}(\mathbf{x})||$$

and (2) and (9) are dual we deduce that $\mathbf{s}$ is a minimiser of the Monge problem. QED.

3 Bernoulli's convolution

Given a probability measure ν on $[0,1]$, a Borel map $\mathbf{m} : [0,1] \times \mathbf{R}^d \to \mathbf{R}^d$ satisfying

$$\sup_{t\in \text{spt }(\nu)} ||\mathbf{m}(t,0)|| := K < +\infty \tag{48}$$

and the Lipschitz condition

$$||\mathbf{m}(t,\mathbf{x}) - \mathbf{m}(t,\mathbf{y})|| \leq \beta ||\mathbf{x} - \mathbf{y}||, \quad (\mathbf{x},\mathbf{y} \in \mathbf{R}^d, \quad \nu \; a.e. \; t \in [0,1]) \tag{49}$$

for some $\beta \in (0,1)$, we prove existence and uniqueness of a Borel probability measure μ on $\mathbf{R}^d$ of compact support such that

$$\mu[A] = \int_0^1 \mu[\mathbf{m}_t^{-1}(A)] d\nu(t) \tag{50}$$

for all bounded Borel sets $A \subset \mathbf{R}^d$. Here $\mathbf{m}_t := \mathbf{m}(t,\cdot)$. A similar statement can be found in [4] (page 356) and [17], where the measure ν is a linear combination of dirac masses.

Let $\mathcal{P}_1$ be the set of all Borel probability measures on $\mathbf{R}^d$ having bounded first moments. Define k from $\mathcal{P}_1$ into $\mathcal{P}_1$ by

$$k(\mu)[A] := \int_0^1 \mu[\mathbf{m}_t^{-1}(A)] d\nu(t),$$

for all Borel sets $A \subset \mathbf{R}^d$. Note that μ satisfies (50) if and only if μ is a fixed point of k. Our plan is to prove that k is a contraction map with respect to the Wasserstein-Rubinstein distance d_W introduced in (6).

Theorem 3.1 *Under (48) and (49), k is a contraction map on $\mathcal{P}_1$ with respect to the metric d_W, and the equation $k(\mu) = \mu$ admits a unique solution $\mu_o \in \mathcal{P}_1$. Furthermore, the support of μ_o is contained in the closed ball of center 0 and radius $R_o := \frac{K}{1-\beta}$.*

Proof: 1. *We prove that k is a contraction.* Using (49) and that (2) and (9) are dual we obtain that if $\mu_1, \mu_2 \in \mathcal{P}_1$ and p is a measure on $\mathbf{R}^d \times \mathbf{R}^d$ having μ_1 and μ_2 as its marginals then

$$\begin{aligned} & d_W(k(\mu_1), k(\mu_2)) \\ = & \sup_{||Du||_\infty \leq 1} \int_0^1 [\int_{\mathbf{R}^d \times \mathbf{R}^d} [u(\mathbf{m}(t, \mathbf{x})) - [u(\mathbf{m}(t, \mathbf{y}))] dp(\mathbf{x}, \mathbf{y})] d\nu(t) \\ \leq & \beta \int_{\mathbf{R}^d \times \mathbf{R}^d} ||\mathbf{x} - \mathbf{y}|| dp(\mathbf{x}, \mathbf{y}). \end{aligned}$$

and so,

$$d_W(k(\mu_1), k(\mu_2)) \leq \beta d_W(\mu_1, \mu_2). \tag{51}$$

Thus, k is a contraction map.

2. *Existence of an invariant measure.* Assume $\mu_1 \in \mathcal{P}_1$ is of compact support, say $\text{spt}\,(\mu_1) \subset B(0, R)$, and define recursively

$$\mu_n = k(\mu_{n-1}), \quad (n = 2, 3, \cdots).$$

Combining (48) and (49) we obtain inductively that the support of μ_n is contained in the ball of center 0 and radius $R_n := \beta^{n-1} R + \frac{1-\beta^{n-1}}{1-\beta} K$. Hence,

$$\text{spt}\,(\mu_n) \subset B(0, R + \frac{K}{1-\beta}), \quad (n = 2, 3, \cdots). \tag{52}$$

Let (μ_{n_j}) be a subsequence of (μ_n) converging weak $*$ to $\mu_o \in \mathcal{P}_1$. The sequence $(d_W(\mu_{n_j}, \mu_o))$ converges to 0 (see [11]) and by (51) and (52) we deduce that in fact

$$\lim_{n \to +\infty} d_W(\mu_n, \mu_o) = 0. \tag{53}$$

Combining (51) and (53) we have

$$d_W(k(\mu_o), \mu_o) \leq \liminf_{n \to +\infty} [\beta d_W(\mu_o, \mu_{n-1}) + d_W(\mu_n, \mu_o)] = 0.$$

Hence,

$$k(\mu_o) = \mu_o.$$

3. Since by (51) k is a contraction map, μ_o is the unique solution of the equation $k(\mu) = \mu$. Using that in (52) $R > 0$ is any arbitrary positive number we deduce that the support of μ_o is contained in the closed ball of center 0 and radius $R_o := \frac{K}{1-\beta}$. QED.

Remark 3.2 *If μ_o is the invariant measure of Theorem 3.1 then $\lambda\mu_o$ is also an invariant measure in the sense that it satisfies (50) for all $\lambda > 0$.*

Example[Bernoulli's convolution] Bernoulli's convolution arises in spline theory, in constructing wavelets of compact support, in constructing fractals (see [4], [9], [10]). Assume we are given $2N + 1$ real numbers $c_1, \cdots, c_N > 0$, $\beta_1, \cdots, \beta_N$ and $\alpha > 1$ satisfying the compatibility condition $\sum_{i=1}^N c_i = \alpha$. The problem is to determine whether or not there exists a nonnegative function $f \in L^1(\mathbf{R})$ with compact support such that

$$f(x) = \sum_{i=1}^N c_i f(\alpha x - \beta_i) \quad (x \in \mathbf{R}). \tag{54}$$

Note that solving (54) is equivalent to proving that there exists a measure $\mu_o \in \mathcal{P}_1$, which absolutely continuous with respect to the Lebesgue measure and which is invariant in the sense that

$$\mu_o[A] = \sum_{i=1}^N \frac{c_i}{\alpha}\mu_o[m_i^{-1}(A)] \tag{55}$$

for all Borel sets $A \subset \mathbf{R}$. Here, $m_i : \mathbf{R} \to \mathbf{R}$ are defined by

$$m_i(x) = \frac{x + \beta_i}{\alpha}.$$

Problem (54) is unsolved in general even in cases which appear simple at a first glance. For instance, it is not known whether or not the equation

$$f(x) = \frac{3}{4}f(\frac{3x}{2}) + \frac{3}{4}f(\frac{3x}{2} - \frac{1}{2}) \quad (x \in \mathbf{R}), \tag{56}$$

admits a solution $f \in L^1(\mathbf{R})$ which is a nonnegative function with compact support (see [12], [13]). Although we cannot answer that question, using the Wasserstein-Rubinstein distance d_W we can recover many results obtained by various authors (e.g. [12], [13]) which we next describe. Existence of a solution to (55) is given by Theorem 3.1. We next argue that the unique probability measure μ_o solution to (55) is either absolutely continuous with respect to the Lebesgue measure or is singular to the Lebesgue measure. Indeed, by Lebesgue decomposition theorem

$$\mu_o = f dx + \mu_s,$$

where μ_s is singular with respect to the Lebesgue measure. Since m_i are affine maps fdx satisfies (55) i.e., f is a solution to (54). Consequently, $\mu_s = \mu_o - fdx$ satisfies (55) too and so, fdx and μ_s must be colinear. Thus, one of them must be the null measure.

A Density of the set of measure-preserving mappings

The main result in this appendix is Proposition A.3. Its proof might already exist in the literature but we could not find it. If X is a topological space we denote by $\mathcal{B}(X)$ the set of all Borel subsets of X.

Definition A.1 *Assume that μ^+ is a Borel measure on a topological space X and μ^- is a Borel measure on a topological space Y. We say that $(X, \mathcal{B}(X), \mu^+)$ is isomorphic to $(Y, \mathcal{B}(Y), \mu^-)$ if there exists a one-to-one map T of X onto Y such that for all $A \in \mathcal{B}(X)$ we have $T(A) \in \mathcal{B}(Y)$ and $\mu^+[A] = \mu^-[T(A)]$, and for all $B \in \mathcal{B}(Y)$ we have $T^{-1}(B) \in \mathcal{B}(X)$ and $\mu^+[T^{-1}(B)] = \mu^-[B]$. For brevity we say that μ^+ is isomorphic to μ^-.*

Theorem A.2 *Let μ^+ be a finite Borel measure on a complete separable metric space X. Assume that μ^+ has no atoms and $\mu^+[X] = 1$. Then $(X, \mathcal{B}(X), \mu^+)$ is isomorphic to $([0,1], \mathcal{B}([0,1]), \lambda_1)$, where λ_1 stands for the one-dimensional Lebesgue measure on $[0,1]$.*

Proof: We refer the reader to [26], Theorem 16. QED.

Define $\mathcal{A}(X,Y)$ to be the set of all Borel maps $\mathbf{r} : X \to Y$ that push μ^+ forward to μ^- and define $\mathcal{M}(X,Y)$ to be the set of all Borel measures on $X \times Y$ that have μ^+ and μ^- as marginals. Let i be the canonical imbedding $i : \mathcal{A}(X,Y) \to \mathcal{M}(X,Y)$ defined by

$$i(\mathbf{r})[E] := \mu^+\{\mathbf{x} \in X : \ (\mathbf{x}, \mathbf{r}(\mathbf{x})) \in E\}, \tag{57}$$

for $E \subset X \times Y$.

Proposition A.3 *Assume that X and Y are two complete, separable, metric spaces. Assume that μ^+ is a Borel measure on X with no atoms, μ^- is a Borel measure on Y with no atoms, and $\mu^+[X] = \mu^-[Y] = 1$. Then every $\gamma \in \mathcal{A}(X, Y)$ can be approximated in the weak $*$ topology by a sequence $\{i(\mathbf{r}_n)\}$ where $\mathbf{r}_n \in \mathcal{A}$ is one-to-one, i.e.*

$$\int_{X\times Y} F d\gamma = \lim_{n\to+\infty} \int_X F(\mathbf{x}, \mathbf{r}_n(\mathbf{x}))d\mu^+(\mathbf{x}),$$

for all bounded $F \in C(X \times Y)$.

Proof: Denote by λ_2 the 2-dimensional Lebesgue measure and set

$$Z := [0, 1]^4.$$

Since by Theorem A.2 μ^+ and μ^- are isomorphic to λ_2 we may assume without loss of generality that

$$\mu^+ = \mu^- = \lambda_2,$$

and

$$X = Y = [0, 1]^2.$$

Note that γ has no atoms and so, using Theorem A.2 again we find an isomorphism $T = (T_1, T_2)$ of $([0, 1]^2, \mathcal{B}([0, 1]^2), \lambda_2)$ onto $(Z, \mathcal{B}(Z), \gamma)$. Note that $T_1, T_2 : [0, 1]^2 \to [0, 1]^2$ push λ_2 forward to λ_2, i.e.

$$\lambda_2[T_i^{-1}(A)] = \lambda_2[A], \quad (i = 1, 2) \tag{58}$$

for all Borel $A \subset [0, 1]^2$. Thanks to (58) we may find (see [5]) sequences

$$T_1^n, T_2^n : [0, 1]^2 \to [0, 1]^2$$

of smooth, one-to-one measure measure-preserving diffeomorphisms such that

$$\lim_{n\to+\infty} T_i^n(\mathbf{x}) = T_i(\mathbf{x}) \quad (i = 1, 2) \tag{59}$$

for λ_2-almost every $\mathbf{x} \in [0, 1]^2$. The maps

$$\mathbf{r}_n := T_2^n \circ (T_1^n)^{-1}$$

are one-to-one, push λ_2 forward to λ_2 and satisfy

$$\int_{X\times Y} F d\gamma = \lim_{n\to+\infty} \int_X F(\mathbf{x}, \mathbf{r}_n(\mathbf{x}))d\mu^+(\mathbf{x}),$$

for all bounded $F \in C(X \times Y)$. QED.

Remark A.4 *Note that there are only two smooth one-to-one maps $T: [0,1] \to [0,1]$ that push λ_1 the 1-dimensional Lebesgue measure forward to λ_1. Consequently, we may not find smooth, one-to-one sequences $T_1^n, T_2^n : [0,1] \to [0,1]$ satisfying (59) if we substitute $[0,1]$ to $[0,1]^2$ in the proof of Theorem A.3.*

B Disintegration theorem

Proposition B.1 *[Fubini's theorem for doubly stochastic measures] Assume that μ^+ is a finite Borel measure on $\mathbf{R}^d$, μ^- is a finite Borel measure on $\mathbf{R}^N$ and γ is a Borel measure on $\mathbf{R}^d \times \mathbf{R}^N$ having μ^+ and μ^- as its marginals. Then there exists a family $(\mu_{\mathbf{x}})_{\mathbf{x}\in\mathbf{R}^d}$ of probability measures on $\mathbf{R}^N$ such that for each $S \subset \mathbf{R}^d \times \mathbf{R}^N$ Borel set $\mathbf{x} \to \mu_{\mathbf{x}}[S_{\mathbf{x}}]$ is μ^+-measurable and*

$$\gamma[S] = \int_{\mathbf{R}^d} \mu_{\mathbf{x}}[S_{\mathbf{x}}] d\mu^+(\mathbf{x}),$$

where $S_{\mathbf{x}} := \{\mathbf{y} \in \mathbf{R}^N : \ (\mathbf{x},\mathbf{y}) \in S\}$.

Proof: 1. Define the projection $\mathbf{p} : \mathbf{R}^d \times \mathbf{R}^N \to \mathbf{R}^d$ by

$$\mathbf{p}(\mathbf{x},\mathbf{y}) = \mathbf{x} \quad ((\mathbf{x},\mathbf{y}) \in \mathbf{R}^d \times \mathbf{R}^N).$$

Then, γ is a σ-finite Borel measure on $\mathbf{R}^d \times \mathbf{R}^N$ and

$$\gamma[\mathbf{p}^{-1}(B)] = 0$$

for all $B \subset \mathbf{R}^d$ such that $\mu^+[B] = 0$. In light of the disintegration of measures theorem ([16]) we deduce that there exists a collection $(\gamma_{\mathbf{x}})_{\mathbf{x}\in\mathbf{R}^d}$ of Borel measures on $\mathbf{R}^d \times \mathbf{R}^N$ such that for every Borel set $S \subset \mathbf{R}^d \times \mathbf{R}^N$

$$(i) \quad \mathbf{x} \to \gamma_{\mathbf{x}}[S] \text{ is a Borel map}$$

$$(ii) \quad \gamma[S] = \int_{\mathbf{R}^d \times \mathbf{R}^N} \gamma_{\mathbf{x}}[S] d\mu^+(\mathbf{x}),$$

and

$$(iii) \quad \gamma_{\mathbf{x}}[\mathbf{R}^d \times \mathbf{R}^N \setminus \{\mathbf{x}\} \times \mathbf{R}^N] = 0.$$

2. Define the measures

$$\mu_{\mathbf{x}}[B] := \gamma_{\mathbf{x}}[\{\mathbf{x}\} \times B]$$

for $B \subset \mathbf{R}^N$ Borel. Let $S \subset \mathbf{R}^d \times \mathbf{R}^N$ be a Borel set. Since

$$S = [(\{\mathbf{x}\} \times S_{\mathbf{x}}) \cap S] \cup [(\{\mathbf{x}\}^c \times \mathbf{R}^N) \cap S]$$

using (iii) we obtain that

$$\gamma_{\mathbf{x}}[S] \leq \gamma_{\mathbf{x}}[\{\mathbf{x}\} \times S_{\mathbf{x}}] \leq \gamma_{\mathbf{x}}[S].$$

Thus, $\gamma_{\mathbf{x}}[S] = \mu_{\mathbf{x}}[S_{\mathbf{x}}]$, which, together with (i) and (ii) yields $\mathbf{x} \to \mu_{\mathbf{x}}[S_{\mathbf{x}}]$ is a Borel map and

$$\gamma[S] = \int_{\mathbf{R}^d} \mu_{\mathbf{x}}[S_{\mathbf{x}}] d\mu^+(\mathbf{x}). \tag{60}$$

3. Writing (60) for $S = A \times \mathbf{R}^N$ where $A \subset \mathbf{R}^d$ is an arbitrary Borel set we deduce that $\mu_{\mathbf{x}}[\mathbf{R}^d] = 1$ for μ^+-almost every $\mathbf{x} \in \mathbf{R}^d$. QED.

Corollary B.2 *Take $\mu^\pm$ and γ as in Proposition B.1. Then there exists a family $(\mu_{\mathbf{x}})_{\mathbf{x} \in \mathbf{R}^d}$ of probability measures on $\mathbf{R}^N$ such that for all Borel maps $F : \mathbf{R}^d \times \mathbf{R}^N \to [-\infty, +\infty]$ that are γ-summable we have $\mathbf{x} \to \int_{\mathbf{R}^N} F(\mathbf{x}, \mathbf{y}) d\mu_{\mathbf{x}}(\mathbf{y})$ is μ^+-integrable and*

$$\int_{\mathbf{R}^d \times \mathbf{R}^N} F d\gamma = \int_{\mathbf{R}^d} [\int_{\mathbf{R}^N} F(\mathbf{x}, \mathbf{y}) d\mu_{\mathbf{x}}(\mathbf{y})] d\mu^+(\mathbf{x}).$$

Remark B.3 *If $\mu^+ = \chi_\Omega d\mathbf{x}$ where $\Omega \subset \mathbf{R}^d$ is an open bounded set then $(\mu_{\mathbf{x}})_{\mathbf{x} \in \mathbf{R}^d}$ coincides with the usual Young measures introduced in [30] and used in [28], etc...*

C Extremal measures

Recall that λ_1 stands for the one-dimensional Lebesgue measure on $[0,1]$. Assume that

$$\mu^+ \quad \text{is a finite Borel measure on } \mathbf{R}^d, \tag{61}$$

$$\mu^- \quad \text{is a finite Borel measure on } \mathbf{R}^N. \tag{62}$$

Let $\mathcal{M}_{dN}$ be the set of all Borel measures on $\mathbf{R}^d \times \mathbf{R}^N$ that have μ^+ and μ^- as marginals. Denote by $\mathcal{M}_{11}$ the set of all Borel measures on $[0,1]^2$ that have λ_1 and λ_1 as marginals. For a Borel map $L : \mathbf{R}^d \times \mathbf{R}^N \to [0,1]^2$ and a Borel measure γ on $\mathbf{R}^d \times \mathbf{R}^N$ we denote the push forward of γ through L by $L\sharp\gamma$ defined by

$$L\sharp\gamma[B] = \gamma[L^{-1}(B)].$$

Proposition C.1 *Take $\mu^{\pm}$ so that (61) and (62) hold. Let T be an isomorphism of μ^+ onto λ_1 and let $\bar{T}$ be an isomorphism of μ^- onto λ_1. Define $L := (T, \bar{T})$, and let j be the map from the set of all finite Borel measures on $\mathbf{R}^d \times \mathbf{R}^N$ to the set of all finite Borel measures on $[0,1]^2$ defined by $\gamma \to L\sharp\gamma$. Then the following hold:*

(i) the restriction $j : \mathcal{M}_{dN} \to \mathcal{M}_{11}$ is one-to-one.

(ii) If γ is an extreme point of the convex set $\mathcal{M}_{dN}$ then $j(\gamma)$ is an extreme point of the convex set $\mathcal{M}_{11}$.

(iii) The support of $\gamma \in \mathcal{M}_{dN}$ lies in the graph of a Borel map $\mathbf{r} : \mathbf{R}^d \to \mathbf{R}^N$ if and only if the support of $j(\gamma)$ lies in the graph of the Borel map $\bar{T} \circ \mathbf{r} \circ T^{-1} : [0,1] \to [0,1]$.

Proof: The first assertion is trivial and the inverse of j is $\bar{\gamma} \to L^{-1}\sharp\bar{\gamma}$. Note that

$$j(t\gamma_1 + (1-t)\gamma_2) = tj(\gamma_1) + (1-t)j(\gamma_2), \tag{63}$$

for all $t \in (0,1)$ and all $\gamma_1, \gamma_2 \in \mathcal{M}_{dN}$. Combining (63) and (i) we obtain (ii). Assume next that $\gamma \in \mathcal{M}_{dN}$ has its support contained in the graph of a Borel map $\mathbf{r} : \mathbf{R}^d \to \mathbf{R}^N$. By Proposition B.1

$$\gamma[S] = \mu^+\{\mathbf{x} \in \mathbf{R}^d : \ (\mathbf{x}, \mathbf{r}(\mathbf{x})) \in S\},$$

and so,

$$j(\gamma)[E] = \lambda_1\{s \in [0,1] : \ (s, \bar{T} \circ \mathbf{r} \circ T^{-1}(s)) \in E\}, \tag{64}$$

for all $E \subset [0,1]^2$ Borel sets. Thus, the support of $j(\gamma)$ lies in the graph of the Borel map $\bar{T} \circ \mathbf{r} \circ T^{-1}$. Similarly, if the support $j(\gamma)$ lies in the graph of the Borel map $\mathbf{r'} : [0,1] \to [0,1]$ then the support of γ lies in the graph of $\mathbf{r} := \bar{T}^{-1} \circ \mathbf{r'} \circ T$. QED.

Remark C.2 *Take $\mu^{\pm}$ so that (61) and (62) hold. If the support of $\gamma \in \mathcal{M}_{dN}$ lies in the graph of a Borel map $\mathbf{r} : \mathbf{R}^d \to \mathbf{R}^N$ then γ is an extreme point of $\mathcal{M}_{dN}$.*

Proof: Assume that $\gamma = t\gamma_1 + (1-t)\gamma_2$ where $\gamma_1, \gamma_2 \in \mathcal{M}_{dN}$. Then the supports of γ_1 and γ_2 must be contained in the support of γ which is a subset of the graph of $\mathbf{r}$. Thus, By Proposition B.1

$$\gamma_1[S] = \mu^+\{\mathbf{x} \in \mathbf{R}^d : \ (\mathbf{x}, \mathbf{r}(\mathbf{x})) \in S\} = \gamma_2[S],$$

for all Borel $S \subset \mathbf{R}^d \times \mathbf{R}^N$. Consequently, $\gamma_1 = \gamma_2$. QED.

Corollary C.3 *Take $\mu^{\pm}$ and γ as in Proposition B.1. Then there exists an extreme point of $\mathcal{M}_{dN}$ whose support does not lie in the graph of a Borel map.*

Proof: In light of Proposition C.1 we may assume without loss of generality that $X := [-1,1] \times \{0\}$, μ^+ is the restriction of λ_1 to X, $Y := ([-1,1] \times \{-1\}) \cup ([-1,1] \times \{1\})$, and μ^- is the restriction of $\frac{1}{2}\lambda_1$ to Y. As observed in [27] one can readily check that the Monge problem admits a unique minimiser $p \in \mathcal{M}_{dN}$ whose support is concentrated on the segments $\{(t,0,t,1) :\ t \in [-1,1]\}$ and $\{(t,0,t,-1) :\ t \in [-1,1]\}$. Clearly, p is not in $i(\mathcal{A})$. More examples are provided in [15]

QED.

References

[1] G.P. Akilov and L. Kantorovich. *Functional Analysis*, 2nd Ed., Pergamon, New-York, 1982.

[2] G. Alberti. On the structure of singular sets of convex functions. *Calc. Var. Partial Differential Equations* **2**, 17–27 (1994).

[3] A.D. Aleksandrov. Existence and uniqueness of a convex surface with a given integral curvature. *C.R. (Doklady) Acad. Sci. URSS (N.S.)* **35**, 131–134 (1942).

[4] M. Barnsley. *Fractals everywhere*, Academic Press, INC, 1988.

[5] Y. Brenier. Lecture notes (preprint).

[6] G. Bouchitté, G. Buttazzo and P. Seppecher. Shape optimization solutions via Monge-Kantorovich equation. *C.R. Acad. Sci. Paris.* (In preprint).

[7] J.E. Colliander and R.L. Jerrard. Vortex dynamics for the Ginzburg-Landau-Schrodinger equation. (In preprint).

[8] J.A. Cuesta and C. Matrán. Notes on the Wasserstein metric in Hilbert Spaces. *Ann. Probab.* **17**, 1264–1276 (1989).

[9] I. Daubechies and J.C. Lagarias. Two-scale difference equations I. Existence and global regularity of solutions *SIAM J. Math. Anal.* vol 22, **5**, 1388–1410 (1991).

[10] I. Daubechies and J.C. Lagarias. Two-scale difference equations II. Local regularity, infinite products of matrices and fractals. *SIAM J. Math. Anal.* vol 23, **4**, 1031–1079 (1992).

[11] R.M. Dudley. Probability and metrics *Aarhus University: Aarhus* (1976).

[12] P. Erdös. On a family of symmetric Bernoulli convolutions. *Amer. J. Math.* **61**, 974–976 (1939).

[13] P. Erdös. On the smoothness properties of a family of Bernoulli convolutions. *Amer. J. Math.* **62**, 180–186 (1940).

[14] L.C. Evans and W. Gangbo. Differential equations methods for the Monge-Kantorovich mass transfer problem. (In preprint).

[15] W. Gangbo and R. McCann. Shape recognition via Wasserstein distance. (In preparation).

[16] S. Graf and R.D. Mauldin. A classification of disintegrations of measures. *Contemporary Mathematics* **94**, 147–158 (1989).

[17] J. Hutchinson. Fractals and self similarity. *Indiana Uni. Math. J.* **30** , no 5, 713–747 (1981).

[18] L. Kantorovich. On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)* **37**, 199–201 (1942).

[19] L. Kantorovich. On a problem of Monge (In Russian). *Uspekhi Math. Nauk.* **3**, 225–226 (1948).

[20] L. Kantorovich and G.Sh. Rubinstein. On a functional space and certain extremal problems, (In Russian). *Dokl. Akad. Nauk. SSSR* **115**, 1058–1061 (1957).

[21] H.G. Kellerer. Duality theorems for marginal problems. *Z. Wahrsch. Verw. Gebiete* **67**, 399–432 (1984).

[22] M. Knott and C.S. Smith. On the optimal mapping of distributions. *J. Optim. Theory Appl.* **43**, 39–49 (1984).

[23] G. Lieberman. Boundary regularity for solutions of degenerate elliptic equations. *Nonlinear Analysis* **12** , 1203–1220 (1988).

[24] G. Monge. Mémoire sur la théorie des déblais et de remblais. *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, pages 666–704 (1781).

[25] S.T. Rachev. The Monge-Kantorovich mass transference problem and its stochastic applications. *Theory Probab. Appl.* **29**, 647–676 (1984).

[26] Royden. *Real Analysis*, 3rd Ed., Macmillan, 1988.

[27] V.N. Sudakov. Geometric problems in the theory of infinite-dimensional probability distributions. *Proc. Steklov Inst. Math.* **141**, 1–178 (1979).

[28] L. Tartar. Compensated compactness and applications to partial differential equations. *Nonlinear Anal. Mech.: Heriott Watt Symposium*, vol. IV (Knops, R., ed.) Pitman Res. Notes in Math. **39**, 136–212 (1979).

[29] K. Uhlenbeck. Regularity for a class of nonlinear elliptic systems. *Acta. Math.*, **138**, 219–240 (1977).

[30] L.C. Young. *Lectures on calculus of variations and optimal control theory,* Saunders, 1969.