ESAIM: COCV 19 (2013) 888–905
DOI: 10.1051/cocv/2012037

**ESAIM: Control, Optimisation and Calculus of Variations**
www.esaim-cocv.org

# TRANSPORT PROBLEMS AND DISINTEGRATION MAPS

Luca Granieri[1] and Francesco Maddalena[1]

**Abstract.** By disintegration of transport plans it is introduced the notion of transport class. This allows to consider the Monge problem as a particular case of the Kantorovich transport problem, once a transport class is fixed. The transport problem constrained to a fixed transport class is equivalent to an abstract Monge problem over a Wasserstein space of probability measures. Concerning solvability of this kind of constrained problems, it turns out that in some sense the Monge problem corresponds to a *lucky* case.

**Mathematics Subject Classification.** 37J50, 49Q20, 49Q15.

Received July 6, 2012. Revised September 26, 2012.
Published online June 3, 2013.

## 1. Introduction

Optimal transport problems, also known as Monge−Kantorovich problems, have been very intensively studied in the last years giving rise to numerous and important applications to PDE, Shape Optimization and Calculus of Variations, so we witnessed a spectacular development of the field. The interested reader may look at the monographs and lecture notes [3, 5, 10, 13, 17–20] where the subject is fully developed.

Let us briefly recall the formulations of the Monge−Kantorovich problems.

Let $X, Y$ be two compact metric spaces and let $c : X \times Y \to \mathbb{R}^+$ be a Borel cost function. The Monge problem is formulated as follows: given two probability measures $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ find a measurable map $t : X \to Y$ such that $t_{\#}\mu = \nu$ ($\#$ denotes the push-forward of measures) and such that $t$ minimizes the total cost, *i.e.*

$$\min_{t:X\to Y} \left\{ \int_X c(x, t(x))\ \mathrm{d}\mu,\ | \, t_{\#}\mu = \nu \right\} \tag{1.1}$$

It may happens that the set of admissible maps is empty (*e.g.* $\mu = \delta_x$ and $\nu = \frac{1}{2}(\delta_y + \delta_z)$). Then the problem could be reformulated in its Kantorovich's relaxation: find $\gamma \in \mathcal{P}(X \times Y)$ such that $\pi^1_{\#}\gamma = \mu$ and $\pi^2_{\#}\gamma = \nu$ ($\pi^1$ and $\pi^2$ are the projections on the factors of $X \times Y$) and such that $\gamma$ minimizes the total cost, *i.e.*

$$\min_{\gamma} \left\{ \int_{X\times Y} c(x, y)\ \mathrm{d}\gamma(x, y) \mid \pi^1_{\#}\gamma = \mu,\ \pi^2_{\#}\gamma = \nu \right\}. \tag{1.2}$$

*Keywords and phrases.* Optimal mass transportation theory, Monge−Kantorovich problem, calculus of variations, shape analysis, geometric measure theory.

[1] Dipartimento di Matematica Politecnico di Bari, via Orabona 4, 70125 Bari, Italy.
l.granieri@poliba.it; f.maddalena@poliba.it

Article published by EDP Sciences

© EDP Sciences, SMAI 2013

https://doi.org/10.1051/cocv/2012037 Published online by Cambridge University Press

The admissible measures $\gamma$ for the Kantorovich problem are called transport plans. We denote by $\Pi(\mu,\nu)$ the set of transport plans with marginals $\mu$ and $\nu$. If $t$ is admissible for the Monge problem then the measure associated in the usual way to the graph of $t$, *i.e.* $\gamma = (I_X \times t)_{\#}\mu$, is admissible for the Kantorovich problem. However the class of admissible measures for the Kantorovich problem is never empty as it contains $\mu \otimes \nu$. Moreover, the Kantorovich problem is a linear one. Existence of minimizers for the Monge problem is difficult and may fails, while for the Kantorovich problem the semicontinuity of $c$ is enough to ensure existence of minimizers.

If $X = Y$, and $c = d$ is the distance function, for $p \geq 1$ the cost

$$W_p(\mu,\nu) = \left( \min \left\{ \int_{X\times X} \mathrm{d}^p(x,y)\ \mathrm{d}\gamma(x,y) \ :\ \gamma \in \Pi(\mu,\nu) \right\} \right)^{1/p}$$

defines a distance on $\mathcal{P}(M)$ called $p$-Wasserstein distance.

Let us recall that by Kantorovich duality (see [3, 13, 19, 20]) the 1-Wasserstein distance between $\mu$ and $\nu$, which we will simply denote by $W$, can be expressed as follows

$$W(\mu,\nu) = \sup \left\{ \int_X \varphi\ \mathrm{d}(\mu-\nu) \mid \varphi \in \mathrm{Lip}_1(X) \right\}, \tag{1.3}$$

where $\mathrm{Lip}_1(X)$ denotes the set of Lipschitz function having Lipschitz constant not greater than one.

## Description of the results

A relevant tool in mass transportation theory is constituted by the Disintegration theorem (Thm. 2.1) of measures which states that every transport plan $\gamma \in \mathcal{P}(X\times Y)$ can be written as $\gamma = f(x)\otimes\mu$ where $f(x) \in \mathcal{P}(Y)$. We shall call *disintegration map* every $f : X \to (\mathcal{P}(Y), W)$ such that

$$f(x) \otimes \mu \in \Pi(\mu,\nu).$$

In this paper we relate the structure of the set of transport plans $\Pi(\mu,\nu)$ with the push-forward of disintegration maps. Indeed, given the measure $\gamma = f(x)\otimes\mu$, obviously $\mu$ is the first marginal of $\gamma$, while the second marginal depends on the disintegration map $f$. An interesting feature of transport plans appears by looking to the measure $f_{\#}\mu$. Precisely, if $\eta = g(x)\otimes\mu$ is another transport plan, it results (see Lem. 2.5):

$$f_{\#}\mu = g_{\#}\mu \Rightarrow \pi^2_{\#}\gamma = \pi^2_{\#}\eta.$$

Therefore, the second marginals can be fixed by looking to the push-forward of disintegration maps. In this way the set of transport plans $\Pi(\mu,\nu)$ can be structured in transport classes (see Def. 2.7) by setting $\eta \in [\gamma] \Leftrightarrow f_{\#}\mu = g_{\#}\mu$. Roughly speaking (see Example 2.8), fixing a transport class leads to consider a constrained transport problem with respect to splitting masses or traveling ones. Lemma 2.3 shows that all transport plans induced by transport maps belong to the same transport class. Moreover, by density of transport maps in $\Pi(\mu,\nu)$, it follows that (see Prop. 2.9) such transport class characterizes the transport plans induced by transport maps.

Therefore, in this perspective the Monge problem can be seen as a constrained Kantorovich problem, namely

$$\min \left\{ \int_X c(x,t(x))\ \mathrm{d}\mu \ :\ t_{\#}\mu = \nu \right\} = \min \left\{ \int_{X\times Y} c(x,y)\ \mathrm{d}\gamma \ :\ \gamma \in [\delta_s \otimes \mu] \right\},$$

for a given transport map $s$. By density of transport maps, the Kantorovich transport problem also corresponds to

$$\min \left\{ \int_{X\times Y} c(x,y)\ \mathrm{d}\gamma \ :\ \gamma \in \Pi(\mu,\nu) \right\} = \min \left\{ \int_{X\times Y} c(x,y)\ \mathrm{d}\gamma \ :\ \gamma \in \overline{[\delta_s \otimes \mu]}^W \right\},$$

for a given transport map $s$.

https://doi.org/10.1051/cocv/2012037 Published online by Cambridge University Press

Hence the Monge problem represents a particular case of a more rich structure of problems, obtained by fixing a transport class in the Kantorovich formulation of transport problems. In this context, considering transport problems in a fixed transport class is quite natural. Since we are looking to the push-forward trough disintegration maps, fixing a transport class results equivalent (see Sect. 2) to consider measures $\Lambda \in \mathcal{P}(\mathcal{P}(Y))$ satisfying the *barycenter* constraint

$$\int_{\mathcal{P}(Y)} \lambda \, \mathrm{d}\Lambda(\lambda) = \nu.$$

The corresponding transport class is given by the transport plans $f(x)\otimes\mu$ such that $f_{\#}\mu = \Lambda$. In this formulation we see that fixing a transport class is equivalent to consider transport maps $f$ sending $\mu$ into $\Lambda$. So it is natural to consider the following Monge−Kantorovich problem in the class $\Lambda$:

$$MK_{\Lambda}(c,\mu,\nu) := \inf_{\gamma} \left\{ \int_{X\times Y} c(x,y) \, \mathrm{d}\gamma \mid \gamma = f \otimes \mu, \ f_{\#}\mu = \Lambda \right\}. \tag{1.4}$$

The above transport problem leads to consider an abstract Monge problem between the space $X$ and $\mathcal{P}(Y)$. Let us consider the following transport cost

$$\forall (x,\lambda) \in X \times \mathcal{P}(Y) \ : \ \tilde{c}(x,\lambda) = \int_Y c(x,y) \, \mathrm{d}\lambda.$$

We set

$$M(\tilde{c},\mu,\Lambda) := \inf_{f} \left\{ \int_X \tilde{c}(x,f(x)) \, \mathrm{d}\mu \mid f_{\#}\mu = \Lambda \right\}. \tag{1.5}$$

For every transport class $\Lambda$ (see Prop. 3.2) we have

$$M(\tilde{c},\mu,\Lambda) = MK_{\Lambda}(c,\mu,\nu).$$

Therefore, every existence result for the Monge problem $M(\tilde{c},\mu,\Lambda)$ in the abstract setting corresponds to an existence result for the Monge−Kantorovich problem in the transport class $\Lambda$. Of course, minimizing in a transport class could be as difficult as for the Monge problem since, of course, the transport classes are not in general closed. However, by this reformulation it comes out that the Monge case is peculiar. More precisely, since the Monge problem is a particular case of transportation in a transport class, one may asks what happens for others transport classes. In other words, the matter consists in establishing if the abstract Monge problem admits solutions. The existence results for the Monge problem are usually stated in the following form: under some assumption on the spaces, on the first marginal $\mu$ and on the cost $c(x,y)$, for every second marginal $\nu$ the Monge problem admits solutions. For the abstract Monge problem $M(\tilde{c},\mu,\Lambda)$ this is not the case. For discrete measures $\Lambda$, see Section 4, for the quadratic cost it results that $M(\tilde{c},\mu,\Lambda)$ may not admit solutions. Therefore, in the abstract setting, it could be also interesting to consider, under some assumption on the spaces, on the first marginal $\mu$ and on the cost $c(x,y)$, the question for what kind of second marginals the corresponding Monge problem admits solutions. From this point of view, in some sense the Monge problem corresponds to a *lucky* case.

## Disintegration maps and variational problems

Our main motivation in introducing the notion of *transport class* is inspired from variational problems arising in elasticity and shape analysis. Actually, a regular transport map can be viewed as a *deformation* of an elastic body $X$ into the deformed configuration $Y$. Since the mass is preserved, we may refer to these maps as *reformations* of $X$ into $Y$. Then, in the setting of nonlinear elasticity, one usually look for minimizers $t : X \to Y$ of the stored energy $\int_X W(\nabla t) \, \mathrm{d}x$. In this kind of approach of course $X$ and $Y$ should be also *regular*. For instance, if $X$ is connected, $Y$ has to be connected too. In order to deal with more general situations also allowing *fragmentations*, in [14] it is studied the notion of transport plan as a weak notion of reformation.

https://doi.org/10.1051/cocv/2012037 Published online by Cambridge University Press

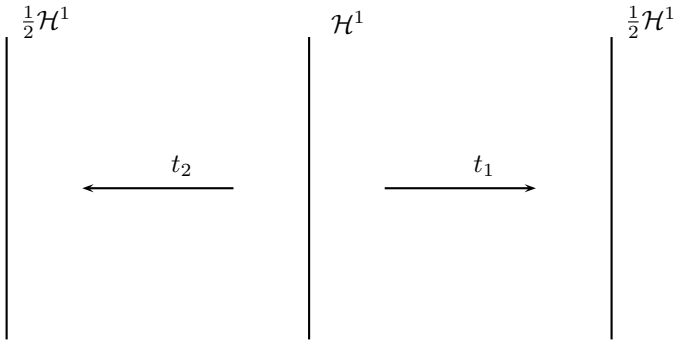



FIGURE 1. An isometric fractured reformation.

Precisely, the role of deformations in classical elasticity is played by the disintegration maps corresponding to transport plans. For instance, the configurations in Figure 1 are isometric, in the sense that the disintegration map corresponding to the transport plan that splits the mass one half on the right and one half to the left is an isometry with respect to the Wasserstein distance, although they cannot be compared by regular reformations.

To state a correspondent variational theory, some troubles appear. In this kind of problems, a key role is played by invertibility properties of maps. By change of variables formula, one is lead to deal with a push-forward condition like $f_{\#}\mu$. The matter is that by considering disintegration maps $f$, the corresponding push-forward is not constant with respect to $f$ and may change, adding an extra difficulty to handle with minimizing sequences. This phenomenon disappears in the classical approach since the push-forward of transport maps $t_{\#}\mu = \nu$ is fixed. Therefore, from this point of view it comes out quite natural to *classify* transport plans through the notion of transport class. We refer the reader to [14] for details.

## 2. Disintegration maps and transport classes

Let $X, Y \subset \mathbb{R}^N$ be two compact sets and let $\mathcal{M}(Y)$ be the space of Radon measures on $Y$. A map $\lambda : X \to \mathcal{M}(Y)$ is said to be Borel, or equivalently weakly*-measurable, if for any open set $B \subset Y$ the function $x \in X \mapsto \lambda_x(B)$ is a real valued Borel map. Equivalently, $x \mapsto \lambda_x$ is a Borel map if, for any Borel and bounded map $\varphi : X \times Y \to \mathbb{R}$, it results that the map

$$x \in X \mapsto \int_Y \varphi(x,y)\mathrm{d}\lambda_x$$

is Borel.

**Theorem 2.1** (Disintegration theorem)**.** *Let $\gamma \in \mathcal{P}(X \times Y)$ be given and let $\pi^1 : X \times Y \to X$ be the first projection map of $X \times Y$, we set $\mu = (\pi^1)_{\#}\gamma$. Then for $\mu - a.e.$ $x \in X$ there exists $\nu_x \in \mathcal{P}(Y)$ such that*

(i) *the map $x \mapsto \nu_x$ is Borel,*

(ii) $\forall \varphi \in \mathcal{C}_b(X \times Y) : \int_{X\times Y} \varphi(x,y)\mathrm{d}\gamma = \int_X \left(\int_Y \varphi(x,y)\mathrm{d}\nu_x(y)\right)\mathrm{d}\mu(x)$.

*Moreover the measures $\nu_x$ are uniquely determined up to a negligible set with respect to $\mu$.*

Let $\gamma \in \Pi(\mu,\nu)$, as usual we will write $\gamma = \nu_x \otimes \mu$, assuming that $\nu_x$ satisfy the condition (i) and (ii) of Theorem 2.1. Obviously, the transport plan $\mu \otimes \nu$ corresponds to the constant map $x \mapsto \nu_x = \nu$. Let $t : X \to Y$, be a transport map, observe that for the transport plan $\gamma_t := (I \times t)_{\#}\mu$, the Disintegration Theorem yields $\gamma_t = \delta_{t(x)} \otimes \mu$. Therefore, the disintegration procedure for a transport plan $\gamma = f(x) \otimes \mu$ produces a map

$$f : X \to (\mathcal{P}(Y), W), \text{ s.t. } x \mapsto f(x) \text{ is Borel.} \tag{2.1}$$

https://doi.org/10.1051/cocv/2012037 Published online by Cambridge University Press

We shall refer to such maps $f$ as *disintegration maps*. For a transport map $t$ the corresponding disintegration map is given by $x \mapsto \delta_{t(x)}$. Of course, it is possible to look at a disintegration map as a Borel map between $X$ and $(\mathcal{P}(Y), W)$. Indeed, we have the following result.

**Lemma 2.2.** *A map $f : X \to (\mathcal{P}(Y), W)$ is a disintegration map if and only if $f$ is Borel.*

*Proof.* Let $f : X \to (\mathcal{P}(Y), W)$ be Borel and let $A \subset Y$ be an open set. Observe that $f(x)(A) = \int_Y \chi_A(y) \mathrm{d}f(x)$. For a l.s.c. function $\varphi$ over $Y$, define

$$I_\varphi : (\mathcal{P}(Y), W) \to \mathbb{R}, \quad I_\varphi(\lambda) := \int_{\mathcal{P}(Y)} \varphi(y) \mathrm{d}\lambda. \tag{2.2}$$

Since $W$ metrizes the weak* topology of measures, we have that $I_\varphi$ is a l.s.c. map. For every $x \in X$ it results

$$\int_Y \varphi(y) \mathrm{d}f(x) = I_\varphi(f(x)).$$

If $f$ is a Borel map, it follows that the map $f(\cdot)(A) : X \to \mathbb{R}$ is also Borel as composition of a l.s.c map and a Borel one. Hence $f$ is a disintegration map. *Vice versa*, observe that by the Ascoli−Arzelá Theorem the space $\mathrm{Lip}_1(Y)$ is compact with respect to the uniform convergence. Fixed a countable dense subset $D \subset \mathrm{Lip}_1(Y)$, by Kantorovich duality we have

$$W(\nu_1, \nu_2) = \sup_{u \in \mathrm{Lip}_1(Y)} \int_Y u \ \mathrm{d}(\nu_1 - \nu_2) = \sup_{u \in D} \int_Y u \ \mathrm{d}(\nu_1 - \nu_2).$$

Since $x \mapsto f(x)$ is Borel, we have that, for every $u \in \mathrm{Lip}_1(Y)$, $g_u : X \to \mathbb{R}$ defined by $g_u(x) := \int_Y u \ \mathrm{d}(\nu - f(x))$ is a Borel map . To check that $f$ is a Borel map, it is sufficient to observe that

$$f^{-1}(B(\nu, r)) = \bigcap_{u \in D} g_u^{-1}(]-r, r[) := A$$

is Borel since $g_u$ is Borel and $D$ countable. It remains to check the above equality. If $x \in A$ we get $|g_u(x)| < r$, for every $u \in D$. Hence, by definition of $g_u$, it follows that $W(\nu, f(x)) < r$ and then $f(x) \in B(\nu, r)$. On the other hand, if $f(x) \in B(\nu, r)$, *i.e.* $W(\nu, f(x)) < r$, by Kantorovich duality $|g_u(x)| := |\int_Y u \ \mathrm{d}(\nu - f(x))| < r$ for every $u \in D$. This implies $x \in A$. □

Let $X \subset \mathbb{R}^N$, we recall that the barycenter of a measure $\mu \in \mathcal{P}(X)$ is given by

$$\beta(\mu) = \int_X x \ \mathrm{d}\mu.$$

Disintegration maps naturally produce measures of the form $f_\# \mu$ on the space $(\mathcal{P}(Y), W)$. By the following lemma, we see that this point of view is equivalent to fix the second marginal of transport plans induced by transport maps.

**Lemma 2.3.** *Let $t, s : X \to Y$ be two given Borel maps, $\mu \in \mathcal{P}(X)$ and let $f, g : X \to \mathcal{P}(Y)$ defined by $f(x) = \delta_{t(x)}$, $g(x) = \delta_{s(x)}$. Then*

$$t_\# \mu = s_\# \mu \ \Leftrightarrow \ f_\# \mu = g_\# \mu. \tag{2.3}$$

*Proof.* Assume $f_\# \mu = g_\# \mu$ and for any $\varphi \in C(Y)$ let us consider the function $I_\varphi$ defined in (2.2). Observe that $I_\varphi \in C((\mathcal{P}(Y), W))$. Hence we have

$$\int_{\mathcal{P}(Y)} I_\varphi(\lambda) \ \mathrm{d}(f_\# \mu) = \int_{\mathcal{P}(Y)} I_\varphi(\lambda) \ \mathrm{d}(g_\# \mu) \Leftrightarrow \int_X I_\varphi(f(x)) \ \mathrm{d}\mu = \int_X I_\varphi(g(x)) \ \mathrm{d}\mu$$

https://doi.org/10.1051/cocv/2012037 Published online by Cambridge University Press

$$\Leftrightarrow \int_X \left( \int_Y \varphi(y)\ \mathrm{d}f(x) \right) \mathrm{d}\mu = \int_X \left( \int_Y \varphi(y)\ \mathrm{d}g(x) \right) \mathrm{d}\mu. \tag{2.4}$$

Since

$$\int_Y \varphi(y)\ \mathrm{d}f(x) = \varphi(t(x)), \quad \int_Y \varphi(y)\ \mathrm{d}g(x) = \varphi(s(x)),$$

by (2.4) we get

$$\int_X \varphi(t(x))\ \mathrm{d}\mu = \int_X \varphi(s(x))\ \mathrm{d}\mu.$$

By the arbitrariness of $\varphi$ we infer $t_{\#}\mu = s_{\#}\mu$.

*Vice-versa*, for every $\psi \in C((\mathcal{P}(Y), W))$ let us consider the function $\varphi(y) = \psi(\delta_y)$. Observe that $\varphi \in C(Y)$. If $t_{\#}\mu = s_{\#}\mu$ we compute

$$\begin{aligned} \int_Y \psi\ \mathrm{d}(f_{\#}\mu) &= \int_X \psi(f(x))\ \mathrm{d}\mu = \int_X \psi(\delta_{t(x)})\ \mathrm{d}\mu \\ &= \int_X \varphi(t(x))\ \mathrm{d}\mu = \int_X \varphi(s(x))\ \mathrm{d}\mu = \int_Y \psi\ \mathrm{d}(g_{\#}\mu). \end{aligned}$$

By the arbitrariness of $\psi$ we obtain $f_{\#}\mu = g_{\#}\mu$. □

**Corollary 2.4.** *Let $t, s : X \to Y$ be two given Borel maps, $\mu \in \mathcal{P}(X)$, let $f, g : X \to \mathcal{P}(Y)$ defined by $f(x) = \delta_{t(x)}$, $g(x) = \delta_{s(x)}$ and let $\gamma = f(x) \otimes \mu$, $\eta = g(x) \otimes \mu$. Then*

$$\pi^2_{\#}\gamma = \pi^2_{\#}\eta \Leftrightarrow f_{\#}\mu = g_{\#}\mu. \tag{2.5}$$

Observe that the first part of the proof of the above Lemma works for general transport plans $\gamma = f(x) \otimes \mu$, $\eta = g(x) \otimes \mu$. Actually, by (2.4) we get the following

**Lemma 2.5.** *Let $\mu \in \mathcal{P}(X)$, $f, g : X \to \mathcal{P}(Y)$, $\gamma = f(x) \otimes \mu$, $\eta = g(x) \otimes \mu$ be given. Then the following implication holds true*

$$f_{\#}\mu = g_{\#}\mu \Rightarrow \pi^2_{\#}\gamma = \pi^2_{\#}\eta. \tag{2.6}$$

Therefore, also for transport plans, the second marginal can be fixed by fixing the push-forward of disintegration maps.

Notice that in general the converse of (2.6) is not true as we show in the next example.

**Example 2.6.** Let $f : X \to \mathcal{P}(Y)$ defined by $f(x) = \nu$ and let $\gamma = f(x) \otimes \mu$. Let $\eta = g(x) \otimes \mu$ where $g(x) = \delta_{t(x)}$ for a given transport map $t : X \to Y$ with $t_{\#}\mu = \nu$. For every $\psi \in \mathcal{C}((\mathcal{P}(Y), W))$ we have

$$\int_{\mathcal{P}(Y)} \psi\ \mathrm{d}(g_{\#}\mu) = \int_X \psi(\delta_{t(x)})\ \mathrm{d}\mu,$$

while

$$\int_{\mathcal{P}(Y)} \psi\ \mathrm{d}(f_{\#}\mu) = \psi(\nu).$$

For any $\varphi \in \mathcal{C}(Y)$ let us consider $\psi(\lambda) = \left| \int_Y \varphi(y) \mathrm{d}\lambda \right| = |I_\varphi(\lambda)|$. We compute

$$\int_X \psi(\delta_{t(x)})\ \mathrm{d}\mu = \int_X \left| \int_Y \varphi(y)\ \mathrm{d}\delta_{t(x)} \right|\ \mathrm{d}\mu = \int_X |\varphi(t(x))|\ \mathrm{d}\mu = \int_Y |\varphi(y)|\ \mathrm{d}\nu.$$

However, on the other hand $\psi(\nu) = \left| \int_Y \varphi(y)\ \mathrm{d}\nu \right|$.

https://doi.org/10.1051/cocv/2012037 Published online by Cambridge University Press

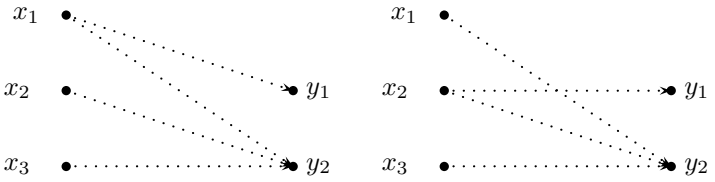



FIGURE 2. Transport plans in the same class.

The above arguments allow to characterize transport plans through the push forward of disintegration maps. We introduce the following notion of transport class.

**Definition 2.7** (Transport classes)**.** Let $\gamma, \eta \in \Pi(\mu, \nu)$ with $\gamma = f(x) \otimes \mu$, $\eta = g(x) \otimes \mu$ be given. We shall say that $\gamma$ and $\eta$ are equivalent (by disintegration), in symbols $\gamma \approx \eta$, if $f_{\#}\mu = g_{\#}\mu$.

For any given $\eta \in \Pi(\mu, \nu)$ with $\eta = g(x) \otimes \mu$, we shall call transport class of $\eta$ the equivalence class of transport plans given by

$$[\eta] = \{\gamma = f(x) \otimes \mu \mid f_{\#}\mu = g_{\#}\mu\}. \tag{2.7}$$

Notice that in the case of discrete first marginal $\mu = \sum_i \alpha_i \delta_{x_i}$, for any disintegration map it is easily seen that

$$f_{\#}\mu = \sum_i \alpha_i \delta_{f(x_i)}.$$

Therefore, transport classes are fixed by the range of $f$.

**Example 2.8.** Let

$$\mu = \frac{1}{3}\delta_{x_1} + \frac{1}{3}\delta_{x_2} + \frac{1}{3}\delta_{x_3}, \quad \nu = \frac{1}{6}\delta_{y_1} + \frac{5}{6}\delta_{y_2}.$$

Consider the transport plan which uniquely splits the mass at $x_1$. This transport plan corresponds to the disintegration map

$$f(x_1) = 3(a\delta_{y_1} + b\delta_{y_2}), \quad f(x_2) = \delta_{y_2}, \quad f(x_3) = \delta_{y_2}, \quad a = b = \frac{1}{6}. \tag{2.8}$$

By changing the point at which the mass is splitted, the range of the corresponding disintegration map does not change. For instance, for the second transport plan in Figure 2 we get the following disintegration map

$$g(x_1) = \delta_{y_2}, \quad g(x_2) = 3(a\delta_{y_1} + b\delta_{y_2}), \quad g(x_3) = \delta_{y_2}, \quad a = b = \frac{1}{6}.$$

It follows $f_{\#}\mu = g_{\#}\mu$. Analogously, all the transport plans with only one splitted mass belong to the same transport class.

On the other hand, by changing the number of splitted masses the corresponding disintegration range is changing.

Indeed, by looking at Figure 3, we may consider the disintegration map

$$h(x_1) = 3(a'\delta_{y_1} + b'\delta_{y_2}), \quad h(x_2) = 3(c'\delta_{y_1} + d'\delta_{y_2}), \quad h(x_3) = \delta_{y_2},$$
$$a' = \frac{3}{30}, \ b' = \frac{7}{30}, \ c' = \frac{2}{30}, \ d' = \frac{8}{30}.$$

In such a case we have that $f_{\#}\mu \neq h_{\#}\mu$. On the other hand, by keeping fixed the number of splitted masses, the transport class may be changed by modifying the amount of traveling masses. Consider for instance the

https://doi.org/10.1051/cocv/2012037 Published online by Cambridge University Press

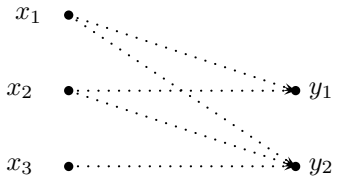



FIGURE 3. Two splitting masses.

disintegration

$$k(x_1) = 3(a''\delta_{y_1} + b''\delta_{y_2}), \quad k(x_2) = 3(c''\delta_{y_1} + d''\delta_{y_2}), \quad k(x_3) = \delta_{y_2},$$
$$a'' = \frac{1}{30},\ b'' = \frac{9}{30},\ c'' = \frac{4}{30},\ d'' = \frac{6}{30}.$$

We get $h_{\#}\mu \neq k_{\#}\mu$.

Therefore, to fix a transport class leads to consider a *constrained transport problem*, with respect to splitting masses or traveling ones.

By Lemma 2.3 it follows that all transport plans induced by transport maps belong to the same transport class. Since the transport maps are dense in $\Pi(\mu,\nu)$ we can prove the following result.

**Proposition 2.9.** *Let $s : X \to Y$ be a transport map,* i.e. *such that $s_{\#}\mu = \nu$, with $\mu$ non-atomic, and let $\eta = (I \times s)_{\#}\mu = \delta_{s(x)} \otimes \mu$. If $\gamma \in [\eta]$ then there exists a transport map $t : X \to Y$ such that $\gamma = \delta_{t(x)} \otimes \mu$,* i.e. *the transport plan $\gamma$ is induced by a transport map $t$. In particular, if $\gamma = f(x) \otimes \mu$, it results $t(x) = \beta(f(x))$ $\mu$-a.e..*

*Proof.* By applying Theorem 9.3 of [3], see also [10], and the same argument employed in Theorem 2.1 of [3], we find a sequence of Borel maps $t_n : X \to Y$ such that

$$\gamma = \lim_{n\to+\infty} \delta_{t_n(x)} \otimes \mu, \quad (t_n)_{\#}\mu = \nu \ \ \forall n \in \mathbb{N}$$

and therefore $\delta_{t_n(x)} \otimes \mu \in [\eta]$, $\forall n \in \mathbb{N}$. Consider $\varphi(y) = |y|^2$. By the push-forward constraint we get

$$\int_X |t_n(x)|^2 \ \mathrm{d}\mu = \int_X \varphi(t_n(x)) \ \mathrm{d}\mu = \int_Y \varphi(y) \ \mathrm{d}\nu = \int_Y |y|^2 \mathrm{d}\nu < +\infty.$$

Let us consider $\psi \in \mathcal{C}((\mathcal{P}(Y), W))$ defined by $\psi(\delta_y) = |y|^2$. In fact, setting $\Delta \subset \mathcal{P}(Y)$ the set of Dirac deltas, the function $\psi$ is Lipschitz, with respect to the Wasserstein distance, over $\Delta$. Hence it suffices to consider any Lipschitz extension of $\psi$ on the whole $\mathcal{P}(Y)$. For every $n \in \mathbb{N}$, since $(\delta_{t_n})_{\#}\mu = (\delta_s)_{\#}\mu$ we have

$$\int_X |t_n(x)|^2 \mathrm{d}\mu = \int_X \psi(\delta_{t_n(x)}) \ \mathrm{d}\mu = \int_X \psi(\delta_{s(x)}) \ \mathrm{d}\mu = \int_X |s|^2 \mathrm{d}\mu. \tag{2.9}$$

Therefore, by passing to a subsequence, we may suppose that $t_n$ is weakly convergent and let $t$ be the weak limit of $t_n$.

Let $\gamma = f(x) \otimes \mu$. By definition of weak convergence, by approximation with continuous functions, for any $g \in L^2(X, \mathbb{R}^N)$ we get

$$\int_X \langle g, t\rangle \mathrm{d}\mu = \lim_{n\to+\infty} \int_X \langle g, t_n\rangle \mathrm{d}\mu = \lim_{n\to+\infty} \int_X \left( \int_Y \langle g, y\rangle \mathrm{d}\delta_{t_n(x)}(y) \right) \mathrm{d}\mu$$
$$= \int_X \left( \int_Y \langle g, y\rangle \ df(x) \right) \mathrm{d}\mu = \int_X \left\langle g, \int_Y y \ \mathrm{d}f(x) \right\rangle \mathrm{d}\mu = \int_X \langle g, \beta(f(x))\rangle \mathrm{d}\mu.$$

https://doi.org/10.1051/cocv/2012037 Published online by Cambridge University Press

Therefore $t_n \rightharpoonup \beta(f(x))$. On the other hand, since $\gamma \in [\eta]$, *i.e.* $f_\#\mu = (\delta_{t_n})_\#\mu$, and by (2.9) we have

$$\int_X |\beta(f(x))|^2 \mathrm{d}\mu = \int_X |\beta(\delta_{s(x)})|^2 \mathrm{d}\mu = \int_X |s|^2 \mathrm{d}\mu = \int_X |t_n|^2 \mathrm{d}\mu.$$

Then, it follows that $t_n$ strongly converges to $\beta(f(x))$, hence by [3], Lemma 9.1, we deduce that $\gamma = \delta_{t(x)} \otimes \mu$ with $t(x) = \beta(f(x))$. □

Of course, the above arguments hold true as well by considering $L^p(X,\mu)$ with $p > 1$. This corresponds to consider the transport cost $d^p(x,y)$.

Proposition 2.9 allows to reformulate the Monge transport problem as follows:

$$\min\left\{\int_X c(x,t(x))\ \mathrm{d}\mu\ :\ t_\#\mu = \nu\right\} = \min\left\{\int_{X\times Y} c(x,y)\ \mathrm{d}\gamma\ :\ \gamma \in [\delta_s \otimes \mu]\right\}, \tag{2.10}$$

for a given transport map $s$.

By density of transport maps, the Kantorovich transport problem can be seen as

$$\min\left\{\int_{X\times Y} c(x,y))\ \mathrm{d}\gamma\ :\ \gamma \in \Pi(\mu,\nu)\right\} = \min\left\{\int_{X\times Y} c(x,y)\ \mathrm{d}\gamma\ :\ \gamma \in \overline{[\delta_t \otimes \mu]}^W\right\}, \tag{2.11}$$

for a given transport map $t$.

Therefore, the Monge problem corresponds to minimize the functional $\int_{X\times Y} c(x,y)\ \mathrm{d}\gamma$ in a fixed transport class of $\Pi(\mu,\nu)$, while the Kantorovich one corresponds to minimize the same functional on the whole $\Pi(\mu,\nu)$.

## 3. Monge–Kantorovich problems on transport classes

In the previous section we have seen that the Monge problem could be seen as a particular case of minimization on a transport class. Since the transport classes are determined through the push-forward of disintegration maps, they can be assigned by probability measures $\Lambda$ over $(\mathcal{P}(Y), W)$.

Actually, consider $f \otimes \mu \in \Pi(\mu,\nu)$ and $\Lambda = f_\#\mu$. Since $(\pi_2)_\#(f\otimes\mu) = \nu$, for every $\varphi \in \mathcal{C}(Y)$ we have

$$\begin{aligned}\int_Y \varphi(y)\ \mathrm{d}\nu &= \int_X \left(\int_Y \varphi(y)\ \mathrm{d}f(x)\right)\mathrm{d}\mu = \int_X I_\varphi(f(x))\ \mathrm{d}\mu \\ &= \int_{\mathcal{P}(Y)} I_\varphi(\lambda)\ \mathrm{d}\Lambda(\lambda) = \int_{\mathcal{P}(Y)}\left(\int_Y \varphi(y)\ \mathrm{d}\lambda\right)\mathrm{d}\Lambda.\end{aligned}$$

Therefore, in order to define a transport class, the measure $\Lambda$ has to satisfy the constraint

$$\int_{\mathcal{P}(Y)} \lambda\ \mathrm{d}\Lambda = \nu. \tag{3.1}$$

Hence, every probability measure $\Lambda$ over $(\mathcal{P}(Y), W)$ satisfying (3.1) defines a transport class $[\eta] = \{f\otimes\mu : f_\#\mu = \Lambda\}$.

For instance, the transport class $[\mu\otimes\nu]$ corresponds to the measure $\Lambda = \delta_\nu$. While for a transport map $t$, the transport class $[\delta_{t(x)}\otimes\mu]$ corresponds to $\Lambda = \int_X \delta_{\delta_{t(x)}}\mathrm{d}\mu$. On the other hand, the transport class in (2.8) corresponds to the discrete measure $\Lambda = \frac{1}{6}\delta_{\delta_{y_1}} + \frac{5}{6}\delta_{\delta_{y_2}}$. In this perspective, transport plans in the transport class $\Lambda$ can be seen as transport maps between $\mu$ and $\Lambda$. It is then natural to consider the Monge–Kantorovich problem in the class $\Lambda$ defined as follows

$$MK_\Lambda(c,\mu,\nu) := \inf_\gamma\left\{\int_{X\times Y} c(x,y)\ \mathrm{d}\gamma \mid \gamma = f\otimes\mu, f_\#\mu = \Lambda\right\} \tag{3.2}$$

By Proposition 2.9, the Monge problem corresponds to the transport class $\Lambda = \int_X \delta_{\delta_{t(x)}}\mathrm{d}\mu$.

https://doi.org/10.1051/cocv/2012037 Published online by Cambridge University Press

**Remark 3.1.** Observe that for a discrete transport class $\Lambda = \sum_i \alpha_i \delta_{\lambda_i}$, the Monge−Kantorovich problem in the class $\Lambda$ yields the *optimal allocation problem* consisting in the determination of a Borel subdivision $\{U_i\}$ of $X$ minimizing

$$\sum_i \int_{U_i} \left( \int_Y c(x,y) \mathrm{d}\lambda_i \right) \mathrm{d}\mu,$$

under the constraint $\sum_i \alpha_i \lambda_i = \nu$. For existence results in this framework see Section 4.

The notion of transport class leads naturally to consider an abstract Monge problem between the space $X$ and $\mathcal{P}(Y)$. Consider the following transport cost

$$\forall (x,\lambda) \in X \times \mathcal{P}(Y) \ : \ \tilde{c}(x,\lambda) = \int_Y c(x,y) \mathrm{d}\lambda. \tag{3.3}$$

We have the following

**Proposition 3.2.** *For every transport class $\Lambda \subset \Pi(\mu,\nu)$ we have*

$$M(\tilde{c},\mu,\Lambda) = MK_\Lambda(c,\mu,\nu).$$

*Proof.* It suffices to observe that for any disintegration map $f : X \to \mathcal{P}(Y)$ such that $f_\# \mu = \Lambda$, it results

$$\int_X \tilde{c}(x,f(x)) \ \mathrm{d}\mu = \int_X \left( \int_Y c(x,y) \mathrm{d}f(x) \right) \mathrm{d}\mu = \int_{X\times Y} c(x,y) \ \mathrm{d}(f \otimes \mu). \qquad \square$$

Observe that by the above proof it follows that $f$ is a solution of $M(\tilde{c},\mu,\Lambda)$ if and only if $f \otimes \mu$ is a solution of $MK_\Lambda(c,\mu,\nu)$. Therefore, every existence result for the Monge problem $M(\tilde{c},\mu,\Lambda)$ in the abstract setting corresponds to an existence result for the Monge−Kantorovich problem in the transport class $\Lambda$. The abstract setting has the advantage of considering a *nice* cost, since it is linear with respect to the second variable. Of course the disadvantage is the passage from the space $X \times Y \subset \mathbb{R}^N \times \mathbb{R}^N$ to the space $X \times \mathcal{P}(Y)$. Observe that the transport classes are always non-empty provided $\mu$ is non-atomic. Indeed, it is well-known, see for instance [17], Theorem 2.4, that whenever $\mu$ is non-atomic transport maps between $\mu$ and $\nu$ always exists. Therefore, the only obstacle in considering transport classes relies in the presence of atoms of $\mu$.

**Proposition 3.3.** *$c$ is continuous, l.s.c., Caratheodory, normal iff $\tilde{c}$ is.*

Where we say that a measurable map $c(x,y)$ is Caratheodory (resp. normal) if $c(x,\cdot)$ is continuous (resp. l.s.c.).

**Lemma 3.4.** *Let $c : X \times Y \to [0,+\infty]$ be a Borel cost function satisfying*

$$|c(x_1,y) - c(x_2,y)| \le \alpha(x_1 - x_2), \tag{3.4}$$

*for a given map $\alpha : X \to \mathbb{R}$ continuous at $x = 0$ and such that $\alpha(0) = 0$. We have the following*

1. *If $c(x,\cdot)$ is continuous then $c$ (and hence $\tilde{c}$) is continuous.*
2. *If $c(x,\cdot)$ is l.s.c. then $c$ (and hence $\tilde{c}$) is l.s.c.*

*Proof.* Let $c(x,\cdot)$ be continuous. If $(x_n,y_n) \to (x,y)$ on $X \times Y$ we compute

$$\begin{aligned} |c(x,y) - c(x_n,y_n)| &\le |c(x,y) - c(x,y_n)| + |c(x,y_n) - c(x_n,y_n)| \\ &\le |c(x,y) - c(x,y_n)| + \alpha(x - x_n) \to 0, \end{aligned}$$

as $n \to +\infty$. If $c(x,\cdot)$ is l.s.c. considering

$$c(x_n,y_n) = c(x_n,y_n) - c(x,y_n) + c(x,y_n),$$

by (3.4), passing to the liminf we obtain

$$\liminf_{n\to+\infty} c(x_n,y_n) = \liminf_{n\to+\infty} c(x,y_n) \ge c(x,y). \qquad \square$$

https://doi.org/10.1051/cocv/2012037 Published online by Cambridge University Press

In general, for the existence of optimal transport plans in the Kantorovich problem at least the lower semicontinuity property of the cost function is usually required. Actually, some regularity of the cost function is needed to obtain a useful duality theory or to ensure that the infimum of the Kantorovich problem is equal to the infimum of the Monge problem (see for instance [3,17]). However, it is not hard to verify that the Kantorovich problem admits solutions under more weak requirements on the cost function (see for instance [6]).

For reader's convenience here we provide some details based on disintegration maps.

Let $\pi_n = f_n \otimes \mu$ be a minimizing sequence for the Monge−Kantorovich problem (1.2). Passing to a subsequence we may suppose that $\pi_n \rightharpoonup f \otimes \mu = \pi \in \Pi(\mu,\nu)$. Now, for any $\psi(x) \in \mathcal{C}(X), \varphi(y) \in \mathcal{C}(Y)$, we get

$$\begin{aligned}\int_{X\times Y} \psi(x)\varphi(y)\ \mathrm{d}\pi &= \int_X \psi(x)\left(\int_Y \varphi(y)\mathrm{d}f(x)\right)\mathrm{d}\mu = \lim_{n\to+\infty}\int_X \psi(x)\left(\int_Y \varphi(y)\mathrm{d}f_n(x)\right)\mathrm{d}\mu \\ &= \lim_{n\to+\infty}\int_{X\times Y}\psi(x)\varphi(y)\ \mathrm{d}\pi_n. \end{aligned} \tag{3.5}$$

By density of continuous functions, the above limit holds for $\psi \in L^1(X,\mu)$ as well. Therefore, if the cost function has the form $c(x,y) = a(x)b(y)$, with $a \in L^1, b \in \mathcal{C}$, then by (3.5) it follows that $\pi$ is an optimal transport plan. Arguing component-wise, the same reasonings apply to linear costs $c(x,y) = \langle a(x), y\rangle$. For a Caratheodory cost function $c(x,y)$, *i.e.* a Borel map such that $c(x,\cdot)$ is continuous, observe that the disintegration maps $f_n : X \to \mathcal{P}(Y) \subset \mathcal{M}(Y,\mathbb{R})$ belongs to $L^\infty(X,\mathcal{M}(Y,\mathbb{R}))$, which is the dual of $L^1(X,\mathcal{C}(Y))$. Therefore, by passing to a subsequence we may suppose that $f_n \overset{*}{\rightharpoonup} f$, *i.e.*

$$\lim_{n\to+\infty}\int_X\left(\int_Y \psi(x,y)\mathrm{d}f_n(x)\right)\mathrm{d}\mu = \int_X\left(\int_Y \psi(x,y)\mathrm{d}f(x)\right)\mathrm{d}\mu \quad \forall\psi \in L^1(X,\mathcal{C}(Y)).$$

The above continuity property shows that $\pi = f\otimes\mu$ is an optimal transport plan, provided that $\int_Y \sup_y c(x,y)\ \mathrm{d}\mu < +\infty$.

If $c(x,y)$ is a normal cost, *i.e.* a Borel measurable map such that $c(x,\cdot)$ is lower semicontinuous, then it can be reduced to a Caratheodory cost by standard approximation procedures. For instance (see [9]), we may write

$$c(x,y) = \sup_h a_h(x)b_h(y), \quad b_h \in \mathcal{C}(Y).$$

Hence, the cost $c_j = \sup_{i\le j} a_i b_i$ is Caratheodory and $c_j \nearrow c$. Since

$$\int_{X\times Y} c_j(x,y)\mathrm{d}\pi_n \le \int_{X\times Y} c(x,y)\mathrm{d}\pi_n,$$

passing to the limit we obtain

$$\int_{X\times Y} c_j(x,y)\mathrm{d}\pi \le \liminf_{n\to+\infty}\int_{X\times Y} c(x,y)\mathrm{d}\pi_n.$$

Passing to the limit with respect to $j$ we get that $\pi$ is optimal. By representation of weakly* l.s.c. functionals (see [4]), the same reasonings apply to normal cost on $X\times\mathcal{P}(Y)$. For a related result see also [12].

In the following we compare the Kantorovich problem (1.2) with the abstract version formulated using the transport classes.

**Lemma 3.5.** *For every transport class $\Lambda \subset \Pi(\mu,\nu)$ it results*

$$MK(c,\mu,\nu) \le MK(\tilde{c},\mu,\Lambda).$$

https://doi.org/10.1051/cocv/2012037 Published online by Cambridge University Press

*Proof.* Let $\tilde{\pi} = \mathcal{N}(x) \otimes \mu \in \Pi(\mu, \Lambda)$. We compute

$$\int_{X\times\mathcal{P}(Y)} \tilde{c}\, \mathrm{d}\tilde{\pi} = \int_X \left( \int_{\mathcal{P}(Y)} \left( \int_Y c(x,y)\mathrm{d}\lambda \right) \mathrm{d}\mathcal{N}(x) \right) \mathrm{d}\mu$$
$$= \int_X \left( \int_Y c(x,y)\mathrm{d}f(x) \right) \mathrm{d}\mu = \int_{X\times Y} c(x,y)\ \mathrm{d}(f\otimes\mu), \tag{3.6}$$

where $f(x) = \int_{\mathcal{P}(Y)} \lambda\ \mathrm{d}\mathcal{N}(x)$. It remains to check that $(\pi_2)_\#(f\otimes\mu) = \nu$. By (3.1), for every $\varphi \in \mathcal{C}(Y)$ we have

$$\int_X \left( \int_Y \varphi(y)\mathrm{d}f(x) \right) \mathrm{d}\mu = \int_X \left( \int_{\mathcal{P}(Y)} \left( \int_Y \varphi(y)\mathrm{d}\lambda \right) \mathrm{d}\mathcal{N}(x) \right) \mathrm{d}\mu = \int_{X\times\mathcal{P}(Y)} I_\varphi(\lambda)\ \mathrm{d}\tilde{\pi}$$
$$= \int_{\mathcal{P}(Y)} I_\varphi(\lambda)\ \mathrm{d}\Lambda = \int_{\mathcal{P}(Y)} \left( \int_Y \varphi(y)\ \mathrm{d}\lambda \right) \mathrm{d}\lambda = \int_Y \varphi(y)\ \mathrm{d}\nu. \qquad \square$$

Observe that the above definition of $f(x) = \beta(\mathcal{N}(x))$ can be seen as a generalized barycenter map. Indeed, we have the following

**Lemma 3.6.** *The generalized barycenter map $\beta : \mathcal{P}(\mathcal{P}(Y)) \to \mathcal{P}(Y)$ defined by*

$$\beta(\mathcal{N}) = \int_{\mathcal{P}(Y)} \lambda\ \mathrm{d}\mathcal{N}$$

*is* 1*-Lipschitz with respect to the Wasserstein distance.*

*Proof.* First observe that if $\varphi \in \mathrm{Lip}_1(Y)$ then $I_\varphi \in \mathrm{Lip}_1(\mathcal{P}(Y))$. Fixed $\varphi \in \mathrm{Lip}_1(Y)$ we get

$$\int_Y \varphi\ \mathrm{d}(\beta(\mathcal{N}_1) - \beta(\mathcal{N}_2)) = \int_{\mathcal{P}(Y)} \left( \int_Y \varphi\ \mathrm{d}\lambda \right) \mathrm{d}(\mathcal{N}_1 - \mathcal{N}_2)$$
$$= \int_{\mathcal{P}(Y)} I_\varphi(\lambda)\ \mathrm{d}(\mathcal{N}_1 - \mathcal{N}_2) \le W(\mathcal{N}_1, \mathcal{N}_2).$$

Taking the supremum with respect to $\varphi \in \mathrm{Lip}_1(Y)$ it results

$$W(\beta(\mathcal{N}_1), \beta(\mathcal{N}_2)) \le W(\mathcal{N}_1, \mathcal{N}_2). \qquad \square$$

Observe that a probability measure $\Lambda$ on $\mathcal{P}(Y)$ defines a transport class iff its generalized barycenter is equal to $\nu$.

**Lemma 3.7.** *If $\tilde{c}$ is normal then there exists a transport class $\Lambda \subset \Pi(\mu,\nu)$ such that*

$$MK(c,\mu,\nu) = MK(\tilde{c},\mu,\Lambda).$$

*Proof.* Let $\pi_n = f_n \otimes \mu \in \Pi(\mu,\nu)$ be a minimizing sequence for $MK(c,\mu,\nu)$. Set $\Lambda_n = (f_n)_\#\mu$. By passing to a subsequence we have that $\Lambda_n \rightharpoonup \Lambda$ with respect to the weak convergence of measures. Observe that

$$\int_{\mathcal{P}(Y)} \left( \int_Y \varphi\ \mathrm{d}\lambda \right) \mathrm{d}\Lambda = \int_{\mathcal{P}(Y)} I_\varphi(\lambda)\ \mathrm{d}\Lambda = \lim_{n\to+\infty} \int_{\mathcal{P}(Y)} I_\varphi(\lambda)\ \mathrm{d}\Lambda_n = \lim_{n\to+\infty} \int_X I_\varphi(f_n(x))\ \mathrm{d}\mu$$
$$= \lim_{n\to+\infty} \int_X \left( \int_Y \varphi(y)\mathrm{d}f_n(x) \right) \mathrm{d}\mu = \int_Y \varphi\ \mathrm{d}\nu.$$

https://doi.org/10.1051/cocv/2012037 Published online by Cambridge University Press

Therefore, $\Lambda$ defines a transport class. Consider the transport plans $\tilde{\pi}_n = (I \times f_n)_{\#}\mu \in \Pi(\mu, \Lambda_n)$. Passing to a subsequence we may also suppose that weakly $\tilde{\pi}_n \rightharpoonup \tilde{\pi} \in \Pi(\mu, \Lambda)$. Since $\tilde{c}$ is normal we get

$$\begin{aligned} MK(\tilde{c}, \mu, \Lambda) \leq \int_{X\times\mathcal{P}(Y)} \tilde{c}\, \mathrm{d}\tilde{\pi} &\leq \liminf_{n\to+\infty} \int_{X\times\mathcal{P}(Y)} \tilde{c}\, \mathrm{d}\tilde{\pi}_n \\ &= \liminf_{n\to+\infty} \int_X \tilde{c}(x, f_n(x))\mathrm{d}\mu \\ &= \liminf_{n\to+\infty} \int_X \left( \int_Y c(x,y)\mathrm{d}f_n(x) \right) \mathrm{d}\mu = MK(c, \mu, \nu). \end{aligned}$$

The result follows by Lemma 3.5. □

By the above analysis the Kantorovich problem over $X \times \mathcal{P}(Y)$ is essentially equivalent to the usual Kantorovich's one. Indeed, if $\tilde{c}$ is normal (for instance if the cost $c$ satisfies the conditions of Lem. 3.4) by Lemma 3.7 we have $MK(c, \mu, \nu) = MK(\tilde{c}, \mu, \Lambda)$ for a transport class $\Lambda$. If $\tilde{\pi} = \mathcal{N}(x) \otimes \mu$ is an optimal transport plan for $MK(\tilde{c}, \mu, \Lambda)$, for such transport class $\Lambda$, setting $f(x) = \int_{\mathcal{P}(Y)} \lambda \, \mathrm{d}\mathcal{N}(x)$, by (3.6) it follows that $f(x) \otimes \mu$ is an optimal plan for $MK(c, \mu, \nu)$. For a related relaxation procedure see [11].

## 4. Existence and uniqueness for the Monge problem

In the previous section we have seen that the Monge−Kantorovich problem is essentially equivalent to the transport problem on transport classes. Of course, the Monge problem reveals hard to handle. For linear cost $c(x, y) = \langle x, y\rangle$, which is equivalent to the quadratic cost $c(x, y) = |x - y|^2$ because of the expansion $|x - y|^2 = |x|^2 + |y|^2 - 2\langle x, y\rangle$, it is relatively easy to find existence and uniqueness of optimal transport maps. We expect some advantage by considering the special form of the cost $\tilde{c}$ for the Monge problem on a transport class. In order to handle with the abstract Monge problem, we review here some usual tools for solving the Monge problem. The available approaches to existence and uniqueness rely more or less on two basic facts. The first one is based on the notion of $c$-cyclical monotonicity. A set $S \subset X \times Y$ is said $c$-cyclical monotone if for any finite set of pairs $(x_1, y_1), \ldots, (x_k, y_k)$ and any permutation $\sigma$ the following inequality holds true

$$\sum_{i=1}^{k} c(x_i, y_i) \leq \sum_{i=1}^{k} c(x_i, y_{\sigma(i)}).$$

A fundamental fact in mass transportation is that the support of every optimal transport plan is a $c$-cyclical monotone set and every $c$-cyclical monotone set is contained in the $c$-superdifferential $\partial^c\psi$ (or contact set) of a $c$-concave function $\psi$, where

$$\partial^c\psi = \{y \in Y \ : \ \psi(x') - \psi(x) \leq c(x', y) - c(x, y) \ \forall x' \in X\}. \tag{4.1}$$

A function $\psi$ is said to be $c$-concave if there exist $A \times B \subset Y \times \mathbb{R}$ such that

$$\psi(x) = \inf_{(y,t)\in A\times B} c(x, y) + t.$$

For details we refer the reader for instance to [17,20]. The $c$-transform of $\psi$ is defined by $\psi^c(y) = \inf_{x\in X}\{c(x, y) - \psi(x)\}$. It can be shown that $y \in \partial^c\psi \Leftrightarrow \psi(x) + \psi^c(y) = c(x, y)$. If one is able to show that for $\mu$ a.e. $x \in X$ the $c$-superdifferential is single valued, then every transport plan is supported on the graph of a transport map (see [3,17,20]). Namely, there exists a unique solution of the Monge problem. The same reasonings apply as well directly to the abstract problem $M(\tilde{c}, \mu, \Lambda)$. In other words, if the $\tilde{c}$-superdifferentials is single-valued, then the Monge problem $M(\tilde{c}, \mu, \Lambda)$ admits a unique solution. In this framework the two Monge problems are essentially

https://doi.org/10.1051/cocv/2012037 Published online by Cambridge University Press

related. Indeed, suppose that the $\tilde{c}$-superdifferentials contain just one Dirac delta and let $\psi$ be a $c$-concave function. We set

$$\tilde{A} = \{\delta_y : y \in A\} \times B \subset \mathcal{P}(Y) \times \mathbb{R}$$

and $\tilde{\psi}(x) = \inf_{(\lambda,t)\in\tilde{A}\times B} \tilde{c}(x,\lambda) + t$. It follows that $\tilde{\psi}$ is $\tilde{c}$-concave and $\tilde{\psi}(x) = \psi(x)$. Recalling that $\tilde{c}(x,\delta_y) = c(x,y)$, the following implications hold true

$$\begin{aligned} y \in \partial^c\psi(x) &\Leftrightarrow \psi(x') - \psi(x) \le c(x',y) - c(x,y) \\ &\Leftrightarrow \tilde{\psi}(x') - \tilde{\psi}(x) \le \tilde{c}(x',\delta_y) - \tilde{c}(x,\delta_y) \Leftrightarrow \delta_y \in \partial^{\tilde{c}}\tilde{\psi}(x). \end{aligned}$$

Therefore $\partial^c\psi(x) = \{y\}$.

Vice-versa, suppose that $c$-superdifferentials are single valued and let $\tilde{\psi}$ be a $\tilde{c}$-concave function. Consider the $c$-transform $(\tilde{\psi})^c(y) = \inf_x\{c(x,y) - \tilde{\psi}(x)\}$.

We have

$$\begin{aligned} \delta_y \in \partial^{\tilde{c}}\tilde{\psi}(x) &\Leftrightarrow \tilde{\psi}(x') - \tilde{\psi}(x) \le c(x',y) - c(x,y) \Leftrightarrow c(x,y) - \tilde{\psi}(x) \le c(x',y) - \tilde{\psi}(x') \\ &\Leftrightarrow c(x,y) - \tilde{\psi}(x) = (\tilde{\psi})^c(y) \Leftrightarrow y \in \partial^c\tilde{\psi}(x) \subset \partial^c u(x) \end{aligned}$$

for a $c$-concave function $u$ (see for instance [17], Rem. 3.12, Thm. 3.10). It follows that $\tilde{\psi}^c(y)$ contains just one delta. Of course this singleton condition of superdifferentials can be achieved under additional requirements on the cost function. For a differentiable cost a general condition relies in the so called *twist* (or *Spence–Mirrlees* in economic settings) condition, *i.e.*

$$x \mapsto c(x,y_1) - c(x,y_2) \text{ has no critical point } \forall\, y_1 \neq y_2 \tag{4.2}$$

(see [7, 10, 15]). For a generalization of such condition, in the case of suitable geometries see [2]. In the case $X = Y = M$ with $M$ a Riemannian manifold and for a Lagrangian cost it is enough for the cost $c$ to satisfy the Mather's shortening principle and the connectedness of the $c$-superdifferential, as it is shown in [20], Chapter 9. Observe that in this case the connectedness is a key property. This property is easily satisfied if $c(x,\cdot)$ is linear as happens just for the cost $c(x,y) = \langle x,y\rangle$. However, general forms of the cost $c$ which guarantee the connectedness of the $c$-superdifferential are not known. In this perspective, the consideration of the linear cost $\tilde{c}$ could be useful. Observe that for a cost $c$ linear with respect to the second variable, the twist condition is not in general satisfied. Consider for instance $c(x,y) = \langle a(x), y\rangle$ for possibly not invertible Jacobian matrix $\nabla a(x)$.

### 4.1. Monge–Mather's shortening principle

By considering the cost $\tilde{c}$, it may happen that the $\tilde{c}$-superdifferential contains many points, which actually are probability measures of $\mathcal{P}(Y)$, although the $c$-superdifferential is a singleton. However, we have the advantage that the $\tilde{c}$-superdifferential are convex sets. Consider $X = Y = M$ and a cost $c$ satisfying a shortening principle. We briefly sketch the reasonings of [20], Chapter 9. Suppose that the following conditions are satisfied

1. There exists $D \subset M$ with $\mu(D) = 0$ such that $D$ intersects every nontrivial continuous curve over $M$.
2. The cost $c$ satisfies a shortening principle.
3. The superdifferential $\partial^c\psi$ is connected.

By assumption (2) it is possible to define a function $F : \gamma^{x,y}(\frac{1}{2}) \mapsto x$ with $y \in \partial^c\psi(x)$ having as domain the mid point of geodesics over $M$. Indeed, by definition of $c$-superdifferential we get

$$\psi(x_1) - \psi(x) \le c(x_1,y) - c(x,y), \quad \psi(x) - \psi(x_1) \le c(x,y_1) - c(x_1,y_1)$$

for every $y \in \partial^c\psi(x), y_1 \in \partial^c\psi(x_1)$. It follows

$$c(x_1,y_1) - c(x,y_1) \le c(x_1,y) - c(x,y) \Rightarrow c(x_1,y_1) + c(x,y) \le c(x_1,y) + c(x,y_1).$$

https://doi.org/10.1051/cocv/2012037 Published online by Cambridge University Press

Hence, the shortening principle implies that

$$\mathrm{d}(x, x_1) \leq K\mathrm{d}\left(\gamma^{x,y}\left(\frac{1}{2}\right), \eta^{x_1,y_1}\left(\frac{1}{2}\right)\right). \tag{4.3}$$

By the inequality (4.3) it follows that $F$ is well defined and moreover it results a Lipschitz, actually also an Holder condition works as well, map. By condition (1) and (3) it is possible to show that the set of points on which $\partial^c\psi$ is not single valued is of null measure. Indeed, let $y_1, y_2 \in \partial^c\psi(x)$. By condition (3) consider a continuous curve $\rho_t$ lying in $\partial^c\psi(x)$ connecting $y_1, y_2$. Therefore it is defined the non-trivial continuous curve $m_t = \gamma^{x,\rho_t}(\frac{1}{2})$. Hence, $x = F(m_t)$. By point (1) it follows that $x \in F(D)$ which is a null measure set. Since $c(x, y) = \tilde{c}(x, \delta_y)$, we would like to prove that the $\tilde{c}$-superdifferential are a.e. single valued. The analogue condition for $\tilde{c}$ superdifferential is

$$\tilde{c}(x_1, \lambda_1) + \tilde{c}(x, \lambda) \leq \tilde{c}(x_1, \lambda) + \tilde{c}(x, \lambda_1).$$

By definition, the above inequality leads to

$$MK(c, \delta_{x_1}, \lambda_1) + MK(c, \delta_x, \lambda) \leq MK(c, \delta_{x_1}, \lambda) + MK(c, \delta_x, \lambda_1).$$

A shortening principle should allow to estimate the distance $\mathrm{d}(x, x_1)$ by the distance between mid points of geodesics in $M$ or in $\mathcal{P}(M)$. The main problem is to find a null set $D$, due to the fact that we are now dealing with the space $M \times \mathcal{P}(M)$.

### 4.2. Twist condition

Assume the cost function $c$ satisfies the twist condition (4.2). Under differentiability requirement, for instance on an open and bounded set $\Omega \subset \mathbb{R}^N$, with a uniform Lipschitz condition

$$|c(x, y) - c(x', y)| \leq K|x - x'|, \quad \forall y \in Y, \tag{4.4}$$

this means that the map $y \mapsto \nabla_x c(x, y)$ is injective. It turns out that $c$-concave functions $\psi$ are Lipschitz on $\Omega$. Indeed, it suffices to compute

$$\begin{aligned}\psi(x_2) &= \inf\{c(x_2, y) + t, (y, t) \in A \times B\} \\ &= \inf\{c(x_2, y) - c(x_1, y) + c(x_1, y) + t, (y, t) \in A \times B\} \leq K|x_1 - x_2| + \psi(x_1).\end{aligned}$$

Moreover, if $\psi$ is differentiable at $x \in \Omega$, for every $y \in \partial^c\psi(x)$, we have

$$\frac{\psi(x + tv) - \psi(x)}{t} \leq \frac{c(x + tv, y) - c(x, y)}{t}.$$

Passing to the limit as $t \to 0^+$ we get

$$\langle \nabla\psi - \nabla_x c(x, y), v\rangle \leq 0.$$

By the arbitrariness of $v$ it follows $\nabla\psi(x) = \nabla_x c(x, y)$ (see also [7]). Therefore, if $\mu$ is absolutely continuous with respect to the Lebesgue measure, then the twist condition implies that the superdifferential $\partial^c\psi(x)$ is a singleton for $\mu$-a.e. $x \in \Omega$. The same reasoning applies as well directly for the cost $\tilde{c}$. Indeed, since $\nabla_x\tilde{c}(x, \lambda) = \int_Y \nabla_x c(x, y)\ \mathrm{d}\lambda$ we have that $\tilde{c}$-superdifferentials contains at most one delta iff the cost $c$ satisfies the twist condition.

In the sequel we sketch a related approach to show existence and uniqueness for the Monge problem.

Assume that $c(x, y)$ is a Caratheodory function, hence by Lemma 3.4 $c$ is continuous. Approximate the target measure $\nu$ by a finite convex combination of Dirac deltas, say by $\nu_n \in \mathcal{P}(Y)$. Consider the problem $MK(c, \mu, \nu_n)$. The optimal transport plans of this approximation problem are of course extremal points of $\Pi(\mu, \nu_n)$. It can be shown that these extremal points, which are supported on the graph of a $c$-concave function $\psi_n$, are of the form $(Id \times t)_{\#}\mu$ (see Sect. 5), *i.e.* the Monge problem $M(c, \mu, \nu_n)$ admits a unique solution.

https://doi.org/10.1051/cocv/2012037 Published online by Cambridge University Press

By the uniform Lipschitz condition (4.4) it follows that the sequence of $c$-concave functions $\psi_n$ is equi-Lipschitz. Indeed, fixed $x, x' \in \Omega$, for $y_n \in \partial_c \psi_n(x)$ we have

$$\psi_n(x') - \psi_n(x) \leq c(x', y_n) - c(x, y_n) \leq K|x - x'|,$$

while for $y_n' \in \partial_c \psi_n(x')$ we have

$$\psi_n(x) - \psi_n(x') \leq c(x, y_n') - c(x', y_n') \leq K|x - x'|.$$

By the Ascoli−Arzelá Theorem we may suppose that $\psi_n \to \psi$, uniformly on compact subsets. Observe that $\psi$ is $c$-concave as well. Indeed, recall that a map $\psi$ is $c$-concave iff $\psi = \psi^{cc}$, where

$$\psi^{cc}(x) = \inf\{\psi(x) \leq f(x) : \ f \text{ c-concave}\}.$$

Fixed $\varepsilon > 0$, we find a large integer $n$ such that $|\psi_n(x) - \psi(x)| < \varepsilon$. Since $\psi_n$ is $c$-concave we get

$$\psi(x) < \varepsilon + \psi_n(x) \Rightarrow \psi^{cc}(x) \leq \psi_n(x) + \varepsilon \leq \psi(x) + 2\varepsilon.$$

By the arbitrariness of $\varepsilon$ we obtain $\psi^{cc} \leq \psi$.

Let $y_n = t_n(x) \in \partial_c \psi_n(x)$. By passing to a subsequence we may suppose that $y_n \to y$. Since

$$\psi_n(x') - \psi_n(x) \leq c(x', y_n) - c(x, y_n) \leq K|x - x'|,$$

passing to the limit as $n \to +\infty$ we get

$$\psi(x') - \psi(x) \leq c(x', y) - c(x, y).$$

Therefore $y \in \partial_c \psi(x)$. Since the $c$-superdifferential is a singleton, the whole sequence $y_n$ converges to $y$. We set $t(x) = y$. As limit of measurable maps, $t$ is measurable as well. Moreover, it results

$$\int_\Omega f(t(x))\mathrm{d}\mu = \lim_{n\to+\infty} \int_\Omega f(t_n(x))\mathrm{d}\mu = \lim_{n\to+\infty} \int_Y f(y)\mathrm{d}\nu_n = \int_Y f(y)\mathrm{d}\nu.$$

Hence $t_{\#}\mu = \nu$. Since the graph $(x, t(x))$ is supported on $\partial_c \psi(x)$ it follows, see [17], Theorem 3.22, that $t$ is an optimal transport map. For a related approximation procedure of the Kantorovich problem see [12].

Hence, under the twist condition for the cost $\tilde{c}$, we would have existence and uniqueness of optimal transport plans in every fixed transport class. Therefore, an interesting question is that of finding condition on the cost $c$ ensuring the twist condition for $\tilde{c}$.

## 5. Existence in discrete transport classes

To treat discrete measures we sketch existence for the Monge problem in this setting. For an extensive discussion of this case we refer to [1,8,16].

Let $\nu = \sum_i a_i \delta_{y_i}$ be a discrete probability measure over a metric space $Y$. Let $\gamma$ be an optimal transport plan between $\mu$ and $\nu$. Denote by $\Gamma = \mathrm{supp}(\gamma)$. By optimality, we have that $\Gamma$ is a $c$-cyclically monotone set of $X \times Y$. Define $A_{i,j} = \{x \in X \ : \ (x, y_i), (x, y_j) \in \Gamma\}$. Let $x, x' \in A_{i,j}$. By $c$-cyclically monotonicity we get

$$c(x, y_i) + c(x', y_j) \leq c(x, y_j) + c(x', y_i) \leq c(x, y_i) + c(x', y_j).$$

Therefore, we obtain

$$c(x, y_i) - c(x, y_j) = c(x', y_i) - c(x', y_j) = \lambda_{i,j}.$$

Under the assumption of $\mu$ $c$-continuous, *i.e.*

$$\mu\left(\{x \in X \ : \ c(x, y_i) - c(x, y_j) = \lambda_{i,j}\}\right) = 0,$$

for every $y_i, y_j \in Y, \lambda_{i,j} \in \mathbb{R}$, we have that $\mu(A_{i,j}) = 0$. Therefore, since the set of splitting masses is given by $A = \bigcup_{i,j} A_{i,j}$, if $\mu$ is $c$-continuous then $\mu(A) = 0$. Hence, the transport plan $\gamma$ is induced by a transport map.

https://doi.org/10.1051/cocv/2012037 Published online by Cambridge University Press

### 5.1. Existence in some transport classes

Let $\Lambda$ be an atomic transport class. Optimal transport plan for $MK(\tilde{c}, \mu, \Lambda)$ are of the form $\sum_i \alpha_i \delta_{f_i} \otimes \mu$ (see [11, 12]). For every index it results $f_i(x) \in \partial^{\tilde{c}}\psi(x)$. By linearity of $\tilde{c}$ it follows $f(x) = \sum \alpha_i f_i \in \partial^{\tilde{c}}\psi(x)$. Therefore, the transport map $f(x)$ is optimal for $M(\tilde{c}, \mu, (f)_{\#}\mu)$.

### 5.2. Non-existence in some transport classes

Consider a discrete transport class given by $\Lambda = \sum_i a_i \delta_{\lambda_i}$. Consider the cost $c(x, y) = \langle x, y \rangle$. For $i, j$ observe that

$$\tilde{c}(x, \lambda_i) - \tilde{c}(x, \lambda_j) = \left\langle x, \int_Y y \, \mathrm{d}(\lambda_i - \lambda_j) \right\rangle.$$

Therefore, $M(\tilde{c}, \mu, \Lambda)$ admits solution iff $\beta(\lambda_i) \neq \beta(\lambda_j)$. Analogously, for cost $c(x, y) = a(x)b(y)$, if $\mu(\{x \in X : a(x) = k\}) = 0$ for every $k \in \mathbb{R}$, it turns out that $M(\tilde{c}, \mu, \Lambda)$ admits solution iff $\int_Y b(y)\mathrm{d}(\lambda_i - \lambda_j) \neq 0$ for every $i, j$.

In this section we have seen that the usual approaches to solve the Monge problem give rise to some difficulties in the setting of the transport problem in a transport class. In some sense, these methods are specific for the transport class corresponding to transport maps. The question to establish existence in different transport classes remains open. We have shown that also for the case of a discrete transport class the answer could be negative. Therefore, from this point of view, the Monge problem corresponds to a *lucky* case for the fixed transport class. This feature of transport classes also naturally leads to the following question. The existence results for the Monge problem are usually stated in the following form: under some assumption on the spaces, on the first marginal $\mu$ and on the cost $c(x, y)$, for every second marginal $\nu$ the Monge problem admits solutions. Having in mind the abstract Monge problem $M(\tilde{c}, \mu, \Lambda)$, it could be also interesting to consider, under some assumption on the spaces, on the first marginal $\mu$ and on the cost $c(x, y)$, the question for what kind of second marginals the corresponding Monge problem admits solutions.

## References


[1] P.T. Abdellaoui and H. Heinich, Caracterisation d'une solution optimale au probleme de Monge−Kantorovich. *Bull. Soc. Math. France* **127** (1999) 429–443.

[2] N. Ahmad, H.K. Kim and R.J. McCann, Optimal transportation, topology and uniqueness. *Bull. Math. Sci.* **1** (2011) 13-32.

[3] L. Ambrosio, Lecture Notes on Transport Problems, in Mathematical Aspects of Evolving Interfaces. *Lect. Notes Math.* vol. 1812. Springer, Berlin (2003) 1–52.

[4] L. Ambrosio, N. Fusco and D. Pallara, Functions of Bounded Variation and Free Discontinuity Problems. Oxford University Press, New York (2000).

[5] L. Ambrosio, N. Gigli and G. Savaré, Gradient Flows in Metric Spaces and in the Space of Probability Measures. *Lect. Notes Math.* ETH Zürich, Birkhäuser (2005).

[6] P. Bernard, Young measures, superposition and transport. *Indiana Univ. Math. J.* **57** (2008) 247–276.

[7] G. Carlier and A. Lachapelle, A Planning Problem Combining Calculus of Variations and Optimal Transport. *Appl. Math. Optim.* **63** (2011) 1–9.

[8] J.A. Cuesta-Albertos and A. Tuero-Diaz, A characterization for the Solution of the Monge−Kantorovich Mass Transference Problem. *Statist. Probab. Lett.* **16** (1993) 147–152.

[9] I. Fonseca and G. Leoni, Modern Methods in the Calculus of Variations: $L^p$ spaces. Springer (2007).

[10] W. Gangbo, The Monge Transfer Problem and its Applications. *Contemp. Math.* **226** (1999) 79–104.

[11] J. Gonzalez-Hernandez and J. Gonzalez-Hernandez, Extreme Points of Sets of Randomized Strategies in Constrained Optimization and Control Problems. *SIAM J. Optim.* **15** (2005) 1085–1104.

[12] J. Gonzalez-Hernandez, J. Rigoberto Gabriel and J. Gonzalez-Hernandez, On Solutions to the Mass Transfer Problem. *SIAM J. Optim.* **17** (2006) 485–499.

[13] L. Granieri, *Optimal Transport and Minimizing Measures*. LAP Lambert Academic Publishing (2010).

[14] L. Granieri and F. Maddalena, A Metric Approach to Elastic reformations, preprint (2012), on `http://cvgmt.sns.it`.

[15] V. Levin, Abstract Cyclical Monotonicity and Monge Solutions for the General Monge−Kantorovich Problem. *Set-Valued Anal.* **7** (1999) 7–32.


https://doi.org/10.1051/cocv/2012037 Published online by Cambridge University Press

[16] M. McAsey and L. Mou, Optimal Locations and the Mass Transport Problem. *Contemp. Math.* **226** (1998) 131–148.
[17] A. Pratelli, *Existence of optimal transport maps and regularity of the transport density in mass transportation problems*, Ph.D. Thesis, Scuola Normale Superiore, Pisa (2003).
[18] S.T. Rachev and L. Ruschendorf, Mass Transportation Problems, *Probab. Appl.* Springer-Verlag, New York **I** (1998).
[19] C. Villani, Topics in Mass Transportation. *Grad. Stud. Math.*, vol. 58. AMS, Providence, RI (2004).
[20] C. Villani, *Optimal Transport, Old and New.* Springer (2009).

https://doi.org/10.1051/cocv/2012037 Published online by Cambridge University Press