On Monge's problem for Bregman-like cost functions

G. Carlier, C. Jimenez *

May 5, 2006

Abstract

We consider Monge's optimal transportation in the case where the transportation cost is the symmetric part of the Bregman distance associated to some smooth and strictly convex function. We prove existence, uniqueness and characterize the solutions.

* Université Paris Dauphine, CEREMADE, Pl. de Lattre de Tassigny, 75775 Paris Cedex 16, FRANCE carlier@ceremade.dauphine.fr, jimenez@ceremade.dauphine.fr.

1 Introduction

Given two probability measures μ and ν on $\mathbb{R}^n$, and a transportation cost function c on $\mathbb{R}^n \times \mathbb{R}^n$, the Monge-Kantorovich optimal transportation problem (introduced in [13]) consists in finding a probability measure γ on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals μ and ν (throughout, such a γ will be refered to as a transport plan) minimizing the cost functional:

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y) d\gamma(x, y).$$

Before we go further, let us define some notation. Given a probability space $(\Omega_1, \mathcal{A}_1, \mu_1)$, a measurable space $(\Omega_2, \mathcal{A}_2)$ and a measurable map f: $\Omega_1 \to \Omega_2$, the push-forward of μ_1 through f, denoted $f\sharp\mu_1$ is the probability measure on $(\Omega_2, \mathcal{A}_2)$ defined by:

$$f\sharp\mu_1(F) := \mu_1(f^{-1}(F))$$

for every $F \in \mathcal{A}_2$. Denoting by

$$(\pi_1(x, y), \pi_2(x, y)) := (x, y), \ \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n$$

the canonical projections, the set of transport plans denoted $\Pi(\mu, \nu)$ is then by definition the set of probability measures γ on $\mathbb{R}^n \times \mathbb{R}^n$ such that $\pi_1\sharp\gamma = \mu$ and $\pi_2\sharp\gamma = \nu$.

The Monge-Kantorovich problem appears as a natural (linear) relaxation of the Monge's problem (see [14]) which consists in finding a transportation map, i.e., a measurable map s that pushes μ forward to ν (meaning that $s\sharp\mu = \nu$) minimizing:

$$\int_{\mathbb{R}^n} c(x, s(x)) d\mu(x).$$

The problem of finding an optimal s has been solved first by Y. Brenier in [4] for a cost which is the square of the Euclidean distance. Other costs have also been studied later in the literature (see [10], [6], [12] and [7].) Optimal transportation problems have been the subject of an intensive stream of research since Brenier's paper [4]. We refer to the books [15] and [16] for a modern account of the theory, its applications and complete references.

In the present paper, we are interested in transportation costs based on Bregman distances. Given a strictly convex and differentiable function ϕ on $\mathbb{R}^n$, the Bregman distance associated to ϕ is:

$$B_\phi(x, y) := \phi(x) - \phi(y) - \langle \nabla\phi(y), x - y \rangle, \ \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n. \qquad (1)$$

For a general convex function ϕ (neither smooth nor even everywhere finite), such functions (which are not symmetric in general hence not distances in the usual sense) have been extensively used in optimization. Indeed, Bregman distances are closely related to relative entropy functions, they are well designed to modelling dissimilarity features in imaging, statistics and decision sciences. Since their systematic introduction by Bregman [3], Bregman distances have been the focus of an active stream of research in optimization and convex programming due to their use in general proximal schemes and interior point methods (see for instance [8], [2], [1] and the references therein).

If we take B_ϕ as transportation cost, we can remark that for any transport plan γ, one has:

$$\int_{\mathbb{R}^n\times\mathbb{R}^n} B_\phi d\gamma = \int_{\mathbb{R}^n} \phi d(\mu-\nu) + \int_{\mathbb{R}^n} \langle y, \nabla\phi(y)\rangle\, d\nu(y) - \int_{\mathbb{R}^n\times\mathbb{R}^n} \langle x, \nabla\phi(y)\rangle\, d\gamma(x,y).$$

Since only the last term depends on γ, the corresponding Monge-Kantorovich problem amounts to maximizing $\int_{\mathbb{R}^n\times\mathbb{R}^n} \langle x, \nabla\phi(y)\rangle\, d\gamma(x,y)$ over $\Pi(\mu,\nu)$. Up to a change of variables, the previous problem has been solved by Brenier in [4] and is very well-known nowadays. We shall therefore not insist on this case.

If we consider now as transportation cost, the symmetric part of the Bregman distance B_ϕ:

$$2c_\phi(x,y) := \langle \nabla\phi(x) - \nabla\phi(y), x-y\rangle\,,$$

dropping as previously the terms that depend only on x or y we see that the Monge-Kantorovich problem with cost c_ϕ amounts to:

$$(\mathcal{M}K) \qquad \sup_{\gamma\in\Pi(\mu,\nu)} \int_{\mathbb{R}^n\times\mathbb{R}^n} H_\phi(x,y)\ d\gamma(x,y).$$

Here H_ϕ is defined by:

$$H_\phi(x,y) = \langle \nabla\phi(x), y\rangle + \langle \nabla\phi(y), x\rangle\,,\ \forall (x,y)\in\mathbb{R}^n\times\mathbb{R}^n. \tag{2}$$

Note that $(\mathcal{MK})$ is the relaxation of the Monge's problem:

$$(\mathcal{M}) \qquad \sup_{s\in\Delta(\mu,\nu)} \int_{\mathbb{R}^n} H_\phi(x,s(x))\ d\mu(x)$$

where $\Delta(\mu,\nu)$ stands for the set of transport maps, i.e., the set of Borel maps s such that $s\sharp\mu=\nu$. Our aim is to prove existence of an optimal transportation map (i.e. a solution to $(\mathcal{M})$) and to characterize it. More precisely, we shall prove under some regularity assumptions, that $(\mathcal{M})$ admits a unique solution (up to μ-a.e. equivalence of course) s that is of the form:

$$s(x) = \nabla(\phi + \frac{1}{2}d_x^2\phi)^\star(\nabla f(x)) \tag{3}$$

for some ϕ-convex potential f (see definition 1). Conversely, if s has the form (3) and is a transport map then s actually solves $(\mathcal{M})$.

In section 2, we introduce some convexity tools related to the duality theory for $(\mathcal{M}K)$. Section 3 is devoted to the existence, uniqueness and announced characterization of optimal transportation maps.

2 Preliminaries

In the remainder of the paper, we will always assume the following:

$$(\mathcal{H}) \left\{ \begin{array}{ll} (i) & \text{there is some closed ball } B \text{ such that } \mu(B) = \nu(B) = 1, \\ (ii) & \mu \text{ is absolutely continuous with respect to the} \\ & \ n\text{-dimensional Lebesgue's measure,} \\ (iii) & \phi \text{ is a strictly convex function of class } C^2 \text{ on } B. \end{array} \right.$$

Our focus is then on the Monge problem:

$$(\mathcal{M}) \qquad \sup_{s\in\Delta(\mu,\nu)} \int_{B\times B} H_\phi(x, s(x))\ d\mu(x)$$

where $\Delta(\mu,\nu) := \{s \ \text{ Borel} : \ B \to B \ : s\sharp\mu = \nu\}$ and H_ϕ is defined by (2).

2.1 Dual problem and ϕ-convex functions

By standard linear programming duality, $(\mathcal{M}K)$ is the dual problem of:

$$(\mathcal{D}) \quad \inf \left\{ \int_B f d\mu + \int_B g d\nu \ : \ f(x) + g(y) \geq H_\phi(x,y), \ \forall (x,y) \in B^2 \right\},$$

where the infimum is taken over all pairs $(f,g) \in C^0(B,\mathbb{R})^2$. We are going to establish the existence of solutions for $(\mathcal{D})$ and:

$$\inf(\mathcal{D}) = \max(\mathcal{M}) = \max(\mathcal{M}K). \tag{4}$$

The existence of a solution for $(\mathcal{M}K)$ can be easily obtained using a direct compactness argument method. Note also that obviously:

$$\sup(\mathcal{M}K) \geq \sup(\mathcal{M}) \tag{5}$$

(this comes from the fact that if s is admissible for $(\mathcal{M})$ then $\gamma = \mu \otimes \delta_{s(x)}$ is admissible for $(\mathcal{M}K)$) and by the duality theorem (see [9]):

$$\sup(\mathcal{M}K) = \min(\mathcal{D}). \tag{6}$$

Let us note that if (f,g) is admissible for $(\mathcal{D})$ then the following holds:

$$g(y) \geq \sup_{x\in B} \{\langle \nabla\phi(x), y\rangle + \langle \nabla\phi(y), x\rangle - f(x)\}. \tag{7}$$

This leads us to introduce the ϕ-Fenchel transform of f:

$$f^{\phi}(x) = \sup_{y \in B}\{\langle \nabla\phi(x), y\rangle + \langle \nabla\phi(y), x\rangle - f(y)\}.$$

Let us remark that (f, f^{ϕ}) satisfies the Young inequality:

$$f(x) + f^{\phi}(y) \geq H_{\phi}(x, y), \tag{8}$$

hence the pair (f, f^{ϕ}) is admissible for $(\mathcal{D})$. Let us also remark that due to the regularity of ϕ there exists a constant $C = C(\phi, B)$ such that:

$$H_{\phi}(x, y) - H_{\phi}(x, y') \leq C|y - y'|, \quad \forall (x, y, y') \in B^3. \tag{9}$$

Let us introduce some definitions related to ϕ-Fenchel transforms:

Definition 1 *A function f is said to be ϕ-convex if there exists a nonempty subset A of $B \times \mathbb{R}$ such that:*

$$f(x) = \sup_{(y,t) \in A}\{H_{\phi}(x, y) + t\}.$$

If f is ϕ-convex, the $\phi-$subdifferential, $\partial^{\phi} f(x)$ of f at $x \in B$ is defined by:

$$\partial^{\phi} f(x) = \{y \in B \ : f(x') - f(x) \geq \left\langle \nabla\phi(y), x' - x\right\rangle + \left\langle \nabla\phi(x') - \nabla\phi(x), y\right\rangle\}.$$

We gather some useful results in the next proposition:

Proposition 1 *For every $f \in C^0(B, \mathbb{R})$, then we have:*

1. *f^{ϕ} and $(f^{\phi})^{\phi}$ are ϕ-convex and C-Lipschitz;*
2. *if f is ϕ-convex then $f(x) + f^{\phi}(y) = H_{\phi}(x, y) \Leftrightarrow x \in \partial^{\phi} f^{\phi}(y)$;*
3. *$f \geq (f^{\phi})^{\phi}$ and f is ϕ-convex if and only if $f = (f^{\phi})^{\phi}$.*

We refer to [7] for a proof of these basic properties. Existence of solutions to $(\mathcal{D})$ is now guaranteed by:

Lemma 1 *The problem $(\mathcal{D})$ admits a pair of solutions $(\overline{f}, \overline{g})$ satisfying:*

$$\overline{f}^{\phi} = \overline{g}, \quad \overline{g}^{\phi} = \overline{f}.$$

Proof. As we have already seen, if (f, g) is admissible for $(\mathcal{D})$, then (f, f^{ϕ}) is also admissible thanks to the Young inequality, and $g \geq f^{\phi}$, hence:

$$\int_B f(x)\ d\mu(x) + \int_B g(y)\ d\nu(y) \geq \int_B f(x)\ d\mu(x) + \int_B f^{\phi}(y)\ d\nu(y).$$

These facts shows (f, f^{ϕ}) is a better candidate for $(\mathcal{D})$. With the same arguments we get $((f^{\phi})^{\phi}, f^{\phi})$ is even better.

Let $(f_n, g_n)_n$ be a minimizing sequence for $(\mathcal{D})$. Since $(\tilde{f}_n, \tilde{g}_n) = ((f_n^\phi)^\phi, f_n^\phi)$ is a better candidate than (f_n, g_n), without loss of generality, we may replace (f_n, g_n) by $(\tilde{f}_n, \tilde{g}_n)$. As μ and ν have the same total mass, for any $a \in \mathbb{R}$ we have:

$$\int_B \tilde{f}_n(x)\, d\mu(x) + \int_B \tilde{g}_n(y)\, d\nu(y) = \int_B (\tilde{f}_n(x) - a)\, d\mu(x) + \int_B (\tilde{g}_n(y) + a)\, d\nu(y)$$

so we may assume $\min_B \tilde{g}_n = 0$. Since $\tilde{f}_n = \tilde{g}_n^\phi$, using the fact that H_ϕ is bounded on B, we easily get that $(\tilde{g}_n)$ and $(\tilde{f}_n)$ are uniformly bounded. Since those families are also equi-Lipschitz by Proposition 1, according to Ascoli theorem, we may assume (up to a subsequence) that they converge uniformly respectively to some limits $\overline{g}$ and $\overline{f}$. It is immediate to check that $\overline{g} = \overline{f}^\phi$, $\overline{f} = \overline{g}^\phi$ and that $(\overline{f}, \overline{g})$ solves $(\mathcal{D})$. $\square$

Lemma 2 *If f is ϕ-convex and finite, then $\partial^\phi f(x) \neq \emptyset$ for all $x \in B$. Moreover f is differentiable almost everywhere and at every point of differentiability x of f one has $\partial^\phi f(x) = \{s_f(x)\}$ where s_f is the Borel map defined by*

$$s_f(x) = \nabla(\phi + \frac{1}{2} d_x^2 \phi)^\star (\nabla f(x))$$

($d_x^2\phi$ denoting the quadratic form $d_x^2\phi(y) := \phi''(x)(y, y)$).

Proof. As f is ϕ-convex, there exists a subset C of $B \times \mathbb{R}$ such that:

$$f(x) = \sup_{(y,t) \in C} \{H_\phi(x, y) + t\}, \quad \forall x \in B.$$

Let $x \in B$ and let (y_n, t_n) be a maximizing sequence for $f(x)$. Up to a subsequence, we may assume that y_n admits a limit y, hence t_n converges to $t = f(x) - H_\phi(x, y)$. We then have:

$$f(x') - f(x) \geq \left\langle \nabla\phi(y), x' - x \right\rangle + \left\langle \nabla\phi(x') - \nabla\phi(x), y \right\rangle$$

for all $x' \in B$ so that y belongs to $\partial^\phi f(x)$. By (9), one also gets that for all x' in B:

$$f(x) - f(x') \leq H_\phi(x, y) + t - H_\phi(x', y) - t \leq C|x - x'|.$$

This proves that f is Lipschitz hence, by Rademacher theorem, differentiable almost everywhere.

Let x be a point (in the interior of B) of differentiability of f and let $y \in \partial^\phi f(x)$. For all $k \in \mathbb{R}^n$ and t sufficiently small, $x + tk$ is in B and we have:

$$\langle \nabla\phi(y), tk \rangle + \langle \nabla\phi(x + tk) - \nabla\phi(x), y \rangle \leq f(x + tk) - f(x) = t \langle \nabla f(x), k \rangle + o(t)$$

which yields, after dividing by t and letting $t \to 0^+$

$$\langle \nabla f(x), k \rangle \geq \langle \nabla \phi(y), k \rangle + \phi''(x)(y, k).$$

Since $k \in \mathbb{R}^n$ is arbitrary, we then get:

$$\nabla f(x) = \nabla \phi(y) + \phi''(x)(y) = \nabla(\phi + \frac{1}{2} d_x^2 \phi)(y).$$

By assumption $\nabla(\phi + \frac{1}{2} d_x^2 \phi)$ is injective so that the previous leads to:

$$y = s_f(x) := \nabla(\phi + \frac{1}{2} d_x^2 \phi)^\star (\nabla f(x)). \tag{10}$$

$\square$

2.2 Getting back to $(\mathcal{M})$

We shall prove that if a pair of ϕ-convex functions $(\overline{f}, \overline{g})$ solves $(\mathcal{D})$ then the map defined by $\overline{s}(x) = s_{\overline{f}} = \nabla(\phi + \frac{1}{2} d_x^2 \phi)^\star (\nabla \overline{f}(x))$ pushes forward μ and actually solves $(\mathcal{M})$. First, we state the following result (for the sake of completeness, we give the proof of this result which can be also found in [7] and is based on a key idea of [11]):

Lemma 3 *Let g be $\phi-$convex and $h \in C^0(B, \mathbb{R})$. Let*

$$f_0 := g^\phi \quad \text{and for all } t \in \mathbb{R}, \quad f_t := (g + th)^\phi.$$

Then for any point x where f_0 is differentiable, and for s_{f_0} given by (10), we have:

$$\lim_{t \to 0^+} \frac{1}{t} [f_t(x) - f_0(x)] = -h(s_{f_0}(x)).$$

Proof. We know that:

$$f_0(x) = H_\phi(x, s_{f_0}(x)) - g(s_{f_0}(x)). \tag{11}$$

For $t > 0$ let $y_t \in B$ be such that

$$f_t(x) = H_\phi(x, y_t) - g(y_t) - th(y_t). \tag{12}$$

Since $f_0 = g^\phi$ and $f_t = (g + th)^\phi$ we have:

$$-h(s_{f_0}(x)) \leq \frac{1}{t} [f_t(x) - f_0(x)] \leq -h(y_t). \tag{13}$$

If we prove that y_t converges to $s_{f_0}(x)$ as t goes to 0^+, the desired result will then follow. Assume that for some sequence t_n tending to 0^+, y_{t_n} converges to some $y \in B$, then passing to the limit in (12) yields

$$f_0(x) = H_\phi(x, y) - g(y) = H_\phi(x, y) - (f_0)^\phi(y)$$

so that $y \in \partial^\phi f_0(x) = \{s_{f_0}(x)\}$. By compactness of B this implies that y_t converges to $s_{f_0}(x)$ as t goes to 0^+.

$\square$

We then deduce the following:

Proposition 2 *Let* $(\overline{f}, \overline{g})$ *be a solution of* $(\mathcal{D})$ *and define for point of differentiability* x *of* $\overline{f}$*:*

$$\overline{s}(x) := \nabla(\phi + \frac{1}{2}d_x^2\phi)^\star(\nabla\overline{f}(x)).$$

Then $\overline{s}$ *pushes forward* μ *to* ν *and:*

$$\int_B H_\phi(x, \overline{s}(x))d\mu(x) = \max(\mathcal{MK}) = \max(\mathcal{M}). \tag{14}$$

Proof. Let $h \in C^0(B, \mathbb{R})$, and define for $t \in \mathbb{R}$ $(f_t, g_t) := ((\overline{g}+th)^\phi, \overline{g}+th)$. Then (f_t, g_t) is admissible for $(\mathcal{D})$ by Young inequality, hence, for any $t > 0$:

$$\int_B \frac{f_t - \overline{f}}{t} d\mu + \int_B h d\nu \geq 0.$$

Using Lemma 3, and the dominated convergence theorem we get

$$\int_B h(\overline{s}(x))d\mu(x) \leq \int_B h d\nu,$$

and since h is arbitrary in the previous, we get $\overline{s}\sharp\mu = \nu$. To prove that $\overline{s}$ solves $(\mathcal{M})$, recall that $\partial^\phi\overline{f}(x) = \{\overline{s}(x)\}$ for μ-a.e. x, hence $H_\phi(x, \overline{s}(x)) = \overline{f}(x) + (\overline{f})^\phi(\overline{s}(x))$. Integrating with respect to μ and using $\overline{s}\sharp\mu = \nu$, we get, according to (6) and (5):

$$\begin{aligned}\int_B H_\phi(x, \overline{s}(x))d\mu(x) &= \int_B \overline{f} d\mu + \int_B \overline{f}^\phi d\nu = \min(\mathcal{D}) \\ &= \max(\mathcal{MK}) \geq \max(\mathcal{M}).\end{aligned}$$

$\square$

3 Characterization of optimal transportation maps

We are now in position to prove:

Theorem 1 *Under the general assumptions* $(\mathcal{H})$*, one has:*

1. $(\mathcal{MK})$ *and* $(\mathcal{M})$ *have the same value,*
2. $(\mathcal{M})$ *admits a unique solution* $\overline{s}$ *(up to* μ*-a.e. equivalence),*

3. *this solution is the unique map s such that $s\sharp\mu = \nu$ and*

$$s(x) = \nabla(\phi + \frac{1}{2}d_x^2\phi)^\star(\nabla f(x))$$

for some ϕ-convex function f.

Proof. 1. follows from Proposition 2.

2. Let $(\overline{f}, \overline{f}^\phi)$ be a pair of ϕ-convex potentials solving $(\mathcal{D})$ and let $\overline{s} := s_{\overline{f}}$ be defined as in Proposition 2. We know from Proposition 2 that $\overline{s}$ solves $(\mathcal{M})$. Assuming σ solves $(\mathcal{M})$, we then have:

$$\int_B H_\phi(x, \sigma(x))d\mu(x) = \inf(\mathcal{D}) = \int_B (\overline{f}(x) + \overline{f}^\phi(\sigma(x))d\mu(x),$$

from Young's inequality, we then have $H_\phi(x, \sigma(x)) = \overline{f}(x) + \overline{f}^\phi(\sigma(x))$ μ-a.e. hence $\sigma(x) \in \partial^\phi \overline{f}(x) = \{\overline{s}(x)\}$ μ-a.e.

3. Assume $s\sharp\mu = \nu$ and $s(x) = \nabla(\phi + \frac{1}{2}d_x^2\phi)^\star(\nabla f(x))$ for some ϕ-convex function f. Then $s(x) \in \partial^\phi f(x)$ for μ-a.e. x and

$$\int_B H_\phi(x, s(x))d\mu(x) = \int_B (f(x) + f^\phi(s(x))d\mu(x) \geq \inf(\mathcal{D}) = \sup(\mathcal{M}),$$

which implies that $s = \overline{s}$ (in the μ-a.e. sense) the unique solution of $(\mathcal{M})$.

$\square$

We can reformulate the previous result in terms of the cost c_ϕ, which is the symmetric part of the Bregman distance B_ϕ:

$$2c_\phi(x, y) := \langle \nabla\phi(x) - \nabla\phi(y), x - y \rangle\,.$$

Indeed, we have seen in the introduction that $(\mathcal{M})$ is equivalent to the *minimization* problem:

$$\inf_{s \in \Delta(\mu,\nu)} \int_B c_\phi(x, s(x))d\mu(x). \tag{15}$$

Hence, Theorem 1 means that (15) admits a unique solution $\overline{s}$, which is characterized by $\overline{s}\sharp\mu = \nu$ and $\overline{s} = s_f$ for some ϕ-convex function f.

Let us conclude the paper with a few remarks.

- When $\mu = \nu$, the optimal transportation is of course the identity map. In this straightforward case the infimum in the dual problem is attained for $f(x) = g(x) = \langle x, \nabla\phi(x)\rangle$.

- The optimal transport map s (or equivalently the optimal ϕ-convex potential f) is formally characterized by the equation of Monge-Ampère type:
$$|\det(Ds_f(x))|\nu(s_f(x)) = \mu(x) \text{ with } f\ \phi\text{-convex}.$$

- In the present paper, we made strong assumptions on the data ϕ (assumed to be smooth and strictly convex) and the measures μ and ν (assumed to have compact support). Although it is possible to consider more general measures (satisfying some integrability conditions related to ϕ say) or less regular functions ϕ, it seems, to the best of our knowledge, very difficult to treat the case of a general convex function ϕ. The case where ϕ is not strictly convex for instance leads to the same difficulties as for the transportation cost $|x - y|$ (see for instance [10] or [6] for this delicate problem).

Acknowledgements: The idea of this paper originated with a stimulating discussion with Hedy Attouch on Bregman distances, the authors are therefore particularly grateful to him.

References

[1] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, *Bregman monotone optimization algorithms*, SIAM Journal on Control and Optimization, **42**, no. 2, pp. 596-636 (2003).

[2] H. H. Bauschke, D. Noll, *The method of forward projections*, Journal of Nonlinear and Convex Analysis **3**, no. 2, pp. 191-205 (2002).

[3] L. M. Bregman, *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming*, U.S.S.R Computational Math. and Math. Physics, **3**, pp. 200-217 (1967).

[4] Y. Brenier, *Polar Factorization and monotone rearrangements of vector valued functions*, Comm. Pure and Applied Math., **44**, pp. 375–417 (1991).

[5] Y. Brenier, *The dual least action problem for an ideal, incompressible fluid*, Arch. Rational Mech. Anal. ,**122**, no. 4, pp. 323–351 (1993).

[6] L. Caffarelli, M. Feldman, R.J. McCann, *Constructing optimal maps for Monge's transport problem as limit of strictly convex cost*, J. Amer. Math. Soc., **15**, pp.1-26 (2002).

[7] G. Carlier, *Duality and existence for a class of mass transportation problems and economic applications*, Adv. in Math. Econ., **5**, pp. 1–21 (2003).

[8] G. Chen, M. Teboulle, *Convergence analysis of a proximal-like algorithm using Bregman functions*, SIAM J. Optimization, **3**, pp. 538–543 (1993).

[9] I. Ekeland, R. Temam, *Convex Analysis and Variational Problems*, North-Holland (1972).

[10] L.C. Evans, W. Gangbo, *Differential Equation Methods for the Monge-Kantorovich Mass Transfer Problem.* Memoirs AMS, 658, (1999).

[11] W. Gangbo, *An elementary proof of the polar factorization of vector-valued functions*, Arch. Rational Mech. Anal., **128**, pp. 381–399 (1994).

[12] W. Gangbo, R.J. McCann, *The Geometry of Optimal Transportation*, Acta Math., **177**, pp. 113–161 (1996).

[13] L.V. Kantorovich, *On the transfer of masses*, Dokl. Akad. Nauk. SSR, **37**, pp. 227-229 (1942).

[14] G. Monge, *Mémoire sur la théorie des déblais et remblais*, Histoire de l' académie des sciences, Paris (1781).

[15] S.T. Rachev, L. Rüschendorf, *Mass Transportation Problems. Vol. I: Theory; Vol. II : Applications*, Springer-Verlag (1998).

[16] C. Villani, *Topics in Optimal transportation*, Graduate studies in Mathematics, 58, AMS (2003).