On the weak closure of convex sets of Probability measures

Nicola Gigli*
University of Bordeaux

November 2008

Abstract

We prove that a closed, geodesically convex subset $\mathcal{C}$ of $\mathscr{P}_2^r(\mathbb{R}^d)$ is closed with respect to weak convergence in $\mathscr{P}_2^r(\mathbb{R}^d)$. This means that if $(\mu_n) \subset \mathcal{C}$ is such that $\mu_n \rightharpoonup \mu$ in duality with continuous bounded functions and $\sup_n \int |x|^2 d\mu_n < \infty$, then $\mu \in \mathcal{C}$ as well.

1 Introduction

The aim of this paper is to study the weak closure properties of geodesically convex sets. The reasons of such an interest come from the fact that the distance W was recently studied because of the strict relations with some evolution PDE's which may be interpreted as curves of maximal slope of certain *geodesically* convex functionals, i.e. functionals that are convex along geodesics. Such an approach, introduced by Otto in [8] and then further analyzed by several authors Carrillo-McCann-Villani in [4], by Agueh in [1] and by the author together with Ambrosio and Savaré in [2] (see [2] for more detailed references), leads to the study of the problem of existence and uniqueness of those curves: the theory of minimizing movements introduced by De Giorgi ([6]) provides a satisfactory answer to these questions under only weak compactness assumptions. In [2] there are mainly two theorems on existence of curves of maximal slope for geodesically convex functionals which rely on two different kind of assumptions on the functional F:

i) F is lower semicontinuous w.r.t. the *weak* topology and satisfies

$$F(\boldsymbol{\gamma}[t]) \leq (1-t)F(\boldsymbol{\gamma}[0]) + tF(\boldsymbol{\gamma}[1])$$

for any *optimal* plan $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})$ (see Corollaries 2.4.11 and 2.4.12 of [2]),

ii) F is lower semicontinuous w.r.t. the *strong* topology and satisfies

$$F(\boldsymbol{\gamma}[t]) \leq (1-t)F(\boldsymbol{\gamma}[0]) + tF(\boldsymbol{\gamma}[1])$$

for *any* plan $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})$ (see Theorem 4.0.4 of [2]).

*Nicola.Gigli@math.u-bordeaux1.fr

Here and in the following *strong* topology stands for the topology induced by W_2, and *weak* topology stands for a[1] topology for which a sequence (μ_n) is converging to μ if and only if (μ_n) converges to μ in duality with continuous and bounded functions, and $\sup_n \int |x|^2 d\mu_n < \infty$.

The two notions of convexity along geodesics just introduced are strictly related to the following notions of geodesic convexity for sets:

Definition 1.1 (Geodesically convex sets) *We say that a set* $\mathcal{C} \subset \mathscr{P}_2(\mathbb{R}^d)$ *is* geodesically convex *if for any* $\mu_1, \mu_0 \in \mathcal{C}$ *there exists a* $\boldsymbol{\gamma} \in \mathcal{O}pt(\mu_0, \mu_1)$ *such that the whole segment joining* μ_0 *to* μ_1 *through* $\boldsymbol{\gamma}$ *belongs to* $\mathcal{C}$*, that is:*

$$\big((1-t)\pi^1 + t\pi^2\big)_{\#}\boldsymbol{\gamma} \in \mathcal{C}, \quad \forall t \in [0,1].$$

Definition 1.2 (Strongly geodesically convex sets) *We will say that a set* $\mathcal{C} \subset \mathscr{P}_2(\mathbb{R}^d)$ *is* strongly geodesically convex *if for any* $\mu_1, \mu_0 \in \mathcal{C}$ *and every* $\boldsymbol{\gamma} \in \mathcal{A}dm(\mu_0, \mu_1)$ *the whole segment joining* μ_0 *to* μ_1 *through* $\boldsymbol{\gamma}$ *belongs to* $\mathcal{C}$*, that is:*

$$\big((1-t)\pi^1 + t\pi^2\big)_{\#}\boldsymbol{\gamma} \in \mathcal{C}, \quad \forall t \in [0,1].$$

It is easy to check that if a functional F l.s.c. w.r.t. the W_2-topology is convex along geodesics in the sense of (i) (respectively, (ii)), then its sublevels are geodesically convex (respectively, strongly geodesically convex).

The main result of this work is to show that in case (i) the assumption of lower semicontinuity w.r.t. the weak topology is redundant and may be substituted with semicontinuity w.r.t. W_2 *provided we know that the functional attains the value* $+\infty$ *at non regular measures.* In order to prove this we will show that any W_2-closed geodesically convex subset of $\mathscr{P}_2(\mathbb{R}^d)$ is closed w.r.t. weak convergence of measures in $\mathscr{P}_2(\mathbb{R}^d)$.

The idea comes from functional analysis: indeed it is well known that a closed convex subset of an Hilbert space is weakly closed, as it may be written as intersection of a family of halfspaces. Here we first introduce the notion *halfspace* in $\mathscr{P}_2(\mathbb{R}^d)$ and show that an halfspace is weakly closed; then we prove that W_2-closed geodesically convex subsets of $\mathscr{P}_2^r(\mathbb{R}^d)$ are intersections of a family of halfspaces, and thus weakly closed as well.

A technical issue arises when dealing with non regular measures, the author doesn't know whether the same result holds for general measures or not.

2 Preliminaries

In this section we recall the basic facts of optimal transport theory we will need in the rest of the paper. This introduction is very far from being exhaustive, the interested reader may look at [2] and [10] for proofs and generalizations.

[1] We will state and prove our result in term of sequential closure of geodesically convex sets, as the introduction of the weak topology in $\mathscr{P}_2(\mathbb{R}^d)$ (i.e. the natural topology for which converging sequences are those weakly converging in the sense of definition 2.1), is a bit technical and does not add really much to our understanding of the geometry of $\mathscr{P}_2(\mathbb{R}^d)$. The interested reader may have a look at Chapters 2 and 5 of [7] for a detailed discussion.

We will denote by $\mathscr{P}_2(\mathbb{R}^d)$ the set of probability measures with finite second moment, i.e.:

$$\mathscr{P}_2(\mathbb{R}^d) := \Big\{ \mu \in \mathscr{P}(\mathbb{R}^d) \ : \ \int |x|^2 d\mu < \infty \Big\},$$

and by $\mathscr{P}_2^r(\mathbb{R}^d)$ its subset made of *regular* measures, which are those measures which give 0 mass to $n-1$ rectifiable sets.

We endow $\mathscr{P}_2(\mathbb{R}^d)$ with the quadratic Wasserstein distance, defined as:

$$W_2(\mu,\nu) := \sqrt{\inf \int |x-y|^2 d\boldsymbol{\gamma}},$$

where the infimum is taken among all *admissible* plans $\boldsymbol{\gamma} \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d)$ satisfying $\pi^1_\# \boldsymbol{\gamma} = \mu$ and $\pi^2_\# \boldsymbol{\gamma} = \nu$, where π^1, π^2 are the projection onto the first and second coordinate respectively. A plan which realizes the minimum is called *optimal*.

Definition 2.1 (Convergences in $\mathscr{P}_2(\mathbb{R}^d)$) *We will say that a sequence (μ_n) converges strongly to μ if $W(\mu_n,\mu) \to 0$ as $n \to \infty$ and that it converges weakly if $\int \psi d\mu_n \to \int \psi d\mu$ as $n \to \infty$ for every $\psi \in C_b(\mathbb{R}^d)$ and $\sup_n \int |x|^2 d\mu_n < \infty$.*

The following celebrated result is due to Brenier.

Theorem 2.2 *Let $\mu \in \mathscr{P}_2^r(\mathbb{R}^d)$ and $\nu \in \mathscr{P}_2(\mathbb{R}^d)$. Then there exists only one optimal plan $\boldsymbol{\gamma}$ and this plan is induced by an* optimal map*. Furthermore, this map is the gradient of a convex function. That is, there exists a convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$ such that $\boldsymbol{\gamma} = (Id, \nabla\varphi)_\# \mu$.*

We will denote by T_μ^ν the optimal map given by Brenier's theorem. For a given $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ we will write L^2_μ for the set of measurable maps $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $\|T\|^2_\mu := \int |T(x)|^2 d\mu(x) < \infty$. The space L^2_μ is endowed with a natural inner product: $\langle T, S \rangle_\mu := \int \langle T(x), S(x) \rangle d\mu(x)$.

The following is a well known stability result of optimal maps.

Proposition 2.3 *Let $\nu, \nu_n \in \mathscr{P}_2(\mathbb{R}^d)$, $n \in \mathbb{N}$, and $\mu \in \mathscr{P}_2^r(\mathbb{R}^d)$. Then the sequence (ν_n) converges strongly (resp. weakly) to ν if and only if the sequence $(T_\mu^{\nu_n})$ converges strongly (resp. weakly) to T_μ^ν in L^2_μ.*

Recall that if $\mu \in \mathscr{P}_2^r(\mathbb{R}^d)$ and $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ the unique constant speed geodesic on $[0,1]$ starting from μ and finishing at ν is given by $t \mapsto \mu_t := (Id + t(T_\mu^\nu - Id))_\# \mu$, where Id is the identity map. In this case it is said that the geodesic is *induced* by T_μ^ν. It is known that if μ_t^1 and μ_t^2 are two constant spees geodesics starting from μ and induced by T, S respectively, then it holds:

$$\lim_{t \downarrow 0} \frac{W_2(\mu_t^1, \mu_t^2)}{t} = \|T - S\|_\mu. \tag{2.1}$$

For a proof of this fact see Appendix of [2] or Chapter 4 of [7].

Finally recall that, with the same notation as above, it holds

$$\lim_{t \downarrow 0} \frac{W_2(\mu_t^1, \nu)}{t} = -2\langle T - Id, T_\mu^\nu - Id \rangle_\mu, \qquad \forall \nu \in \mathscr{P}_2(\mathbb{R}^d), \tag{2.2}$$

see Proposition 7.3.6. of [2] for a proof of this fact.

3 The result

The basic object we will need for our result is the following:

Definition 3.1 (Halfspace) *Let $\mu \in \mathscr{P}_2^r(\mathbb{R}^d)$, $v \in L^2_\mu$ and $C \in \mathbb{R}$. The two halfspaces $\mathscr{H}^+_{v;C}$ and $\mathscr{H}^-_{v;C}$ identified by v, C are:*

$$\mathscr{H}^+_{v;C} := \left\{ \nu \ : \ \langle T^\nu_\mu - Id, v \rangle_\mu \geq C \right\},$$
$$\mathscr{H}^-_{v;C} := \left\{ \nu \ : \ \langle T^\nu_\mu - Id, v \rangle_\mu \leq C \right\}.$$

As said, we are going to study only sequential closure of sets: the following proposition is the enabler of the theory.

Proposition 3.2 *Let $\mu \in \mathscr{P}_2^r(\mathbb{R}^d)$, $v \in L^2_\mu$ and $C \in \mathbb{R}$. Then the two halfspaces $\mathscr{H}^+_{v;C}$ and $\mathscr{H}^-_{v;C}$ are weakly sequentially closed.*

Proof. Consider a sequence $(\nu_n) \subset \mathscr{P}_2(\mathbb{R}^d)$ which weakly converges to ν. By proposition 2.3 we know that the sequence of optimal transport maps $(T^{\nu_n}_\mu)$ weakly converges to the optimal transport map T^ν_μ. Thus the bound $\langle T^{\nu_n}_\mu - Id, v \rangle_\mu \geq C$ (or $\langle T^{\nu_n}_\mu - Id, v \rangle_\mu \leq C$) passes to the limit. □

We want to prove that any geodesically convex subset $\mathcal{C}$ of $\mathscr{P}_2^r(\mathbb{R}^d)$ is the intersection of a family of halfspaces. The idea is to find, for every $\nu \notin \mathcal{C}$, an halfspace which contains $\mathcal{C}$ and does not contain ν. Observe that if we knew a priori the existence of a measure $\mu \in \mathcal{C}$ which realizes the minimum distance from ν to $\mathcal{C}$, then the halfspace $\mathscr{H}^-_{T^\nu_\mu - Id;0}$ has the needed property. Indeed, pick any measure $\sigma \in \mathcal{C}$ and let $\mu_t := (Id + t(T^\sigma_\mu - Id))_\# \mu$ and $\nu_t := (Id + t(T^\sigma_\mu - Id))_\# \mu$ be the two geodesics connecting μ to σ and ν respectively. From the minimality of μ and the geodesic convexity of $\mathcal{C}$, it follows that $W(\mu_t, \nu) \geq W(\mu, \nu)$, thus from formula (2.2) we obtain

$$-2\langle T^\sigma_\mu - Id, T^\nu_\mu - Id \rangle_\mu = \lim_{t \downarrow 0} \frac{W^2(\mu_t, \nu) - W^2(\mu, \nu)}{t} \geq 0.$$

However, a priori we don't know that such μ exists (we will know this fact a posteriori, once weak closure will be estabilished), so we need to procede proving the existence of quasi-minima, and then showing that that the above argument still applies to quasi-minima.

The key lemma we will need is the following.

Lemma 3.3 *Let (E, d) be a complete geodesic metric space, $C \subset E$ a closed set, $P \in E \setminus C$, and $0 \leq a < 1$. Then there exists a point $Q \in C$ such that*

$$\frac{d(Q_t, C)}{t} \geq a, \quad \forall t \in (0, d(P, Q)], \tag{3.1}$$

where $Q_t : [0, d(P, Q)] \to E$ is any choice of a geodesic connecting Q to P ($Q_0 = Q$, $Q_1 = P$) parameterized by arc length.

Proof. Let we fix a notation: for any point $R \in C$ let $\mathbf{R}$ be the set of geodesics connecting R to P parameterized by arc length, and let $R_t \in \mathbf{R}$ be a generic element of this set.

We will say that a point $R \in C$ has the property $\mathcal{G}$ iff for every $R_t \in \mathbf{R}$ it holds

$$\frac{d(R_t, C)}{t} \geq a, \quad \forall t \in (0, d(P, R)].$$

Our aim is to prove that a point with the property $\mathcal{G}$ exists.

Start choosing any point $R \in C$ and suppose that it doesn't have the property $\mathcal{G}$. Then there exists a point $R' \in C$ such that

$$d(R_t, R') < at, \qquad \text{for some } t > 0 \text{ and some } R_t \in \mathbf{R}. \tag{3.2}$$

From this we get

$$d(R, R') \leq d(R, R_t) + d(R_t, R') < t(a+1),$$

and $d(R', P) \leq d(R', R_t) + d(R_t, P) < at + d(R, P) - t$ from which it follows

$$d(R, P) - d(R', P) > t(1-a). \tag{3.3}$$

Putting together the last two inequalities we get the key estimate

$$d(R, R') < \frac{1+a}{1-a} \big(d(R, P) - d(R', P) \big). \tag{3.4}$$

This inequality is all we need to prove the thesis: we will proceed by transfinite induction by using its telescopic property.

Let Ω be the first uncountable ordinal. Define a function

$$\begin{aligned} &\Omega \to C \\ &\alpha \to R_\alpha, \end{aligned}$$

beginning by choosing $R_0 \in C$ in any way. Then if α is the successor of some ordinal, we have two cases:

i) $R_{\alpha-1}$ has the property $\mathcal{G}$,

ii) $R_{\alpha-1}$ does not have the property $\mathcal{G}$.

In the first case we put $R_\alpha := R_{\alpha-1}$, in the second one we choose R_α among those points R' satisfying (3.2) with $R = R_{\alpha-1}$. Finally, if α is a limit ordinal we let R_α be the limit of $R_{\alpha'}$ with $\alpha' < \alpha$.

We have to prove that this is a good definition, we will do this by proving at the same time that the following "extended" version of (3.4) holds:

$$d(R_\alpha, R_\beta) < \frac{1+a}{1-a} \left(d(R_\alpha, P) - d(R_\beta, P) \right), \quad \forall \alpha \leq \beta. \tag{3.5}$$

We prove this inequality by transfinite induction on β: it is true for 0, and it is easy to see that if it holds for β then it holds for $\beta+1$. Indeed by construction and from the first part of the proof, $R_{\beta+1}$ satisfies (3.4) with $R = R_\beta$, $R' = R_{\beta+1}$, therefore combining (3.4) and (3.5) we get

$$\begin{aligned} d(R_\alpha, R_{\beta+1}) &\le d(R_\alpha, R_\beta) + d(R_\beta, R_{\beta+1}) \\ &< \frac{1+a}{1-a}\left(d(R_\alpha, P) - d(R_\beta, P) + d(R_\beta, P) - d(R_{\beta+1}, P)\right) \\ &= \frac{1+a}{1-a}\left(d(R_\alpha, P) - d(R_{\beta+1}, P)\right), \quad \forall \alpha \le \beta. \end{aligned}$$

Given that the case $\alpha = \beta + 1$ is obvious, we get the claim.

Now let β be a limit ordinal, observe that we can't write inequality (3.5) for such a β, yet, since we have still to prove that R_β exists: we are going to prove at the same time that R_β is well defined and that for this point (3.5) holds. Since $\beta < \Omega$ there exists an increasing sequence (α_n) converging to β; for every α_n the inequality (3.5) holds, therefore we have

$$d(R_{\alpha_m}, R_{\alpha_n}) < \frac{1+a}{1-a}\left(d(R_{\alpha_m}, P) - d(R_{\alpha_n}, P)\right), \quad \forall m \le n.$$

Being the sequence $d(R_{\alpha_n}, P)$ non increasing (by equation (3.3)) and bounded from below, it is a Cauchy sequence and the previous inequality shows that the same is true for the sequence R_{α_n}, which therefore converges to some point we call R_β. Since the previous argument applies to every increasing sequence α_n, showing that the corrispond points R_{α_n} form a Cauchy sequence, we get that R_β is well defined (i.e. it does not depend on the particular sequence (α_n) chosen), that the function $\alpha \to R_\alpha$ is continuous (with respect to the order topology) and that (3.5) holds for any $\beta < \Omega$.

Observe that from inequality (3.3) it follows that if $R_{\alpha+1} \neq R_\alpha$, then $d(R_{\alpha+1}, P)$ is strictly less than $d(R_\alpha, P)$. We are almost done: since there is no strictly decreasing function from Ω to $\mathbb{R}$, we have that the map $\alpha \to R_\alpha$ has to be eventually constant, therefore for some α we have $R_\alpha = R_{\alpha+1}$, which means by construction that the point $Q = R_\alpha$ satisfies the thesis. □

Note that this proposition is a generalization of the Drop Theorem of Daneš valid in Banach spaces, see [5] for further reference.

This lemma is closely related to the Ekeland-Bishop-Phelps principle. Actually a shorter proof may be given with a direct application of the EBP principle: we present here one found by B.Kirchheim. Use EBP principle to find $Q \in C$ which is a minimizer of

$$x \to f(x) := d(x, P) + \frac{1-a}{1+a} d(x, Q).$$

Then such a Q has the claimed property. Indeed, if this is not the case, there exists $R \in C$ and $0 \le t \le d(P, Q)$ such that $d(R, Q_t) < at$. For such R we have the following bounds

$$\begin{aligned} d(R, P) &\le d(R, Q_t) + d(Q_t, P) < at + d(P, Q) - t, \\ d(R, Q) &\le d(R, Q_t) + d(Q_t, Q) < at + t. \end{aligned}$$

Therefore it holds

$$f(Q) = d(P,Q) = d(P,Q) - t(1-a) + \frac{1-a}{1+a}t(a+1) > d(R,P) + \frac{1-a}{1+a}d(R,Q) = f(R),$$

which contradicts the minimality of Q.

Now we have all the elements to prove our main result.

Theorem 3.4 *Let $\mathcal{C}$ be a strongly closed, geodesically closed subset of $\mathscr{P}_2^r(\mathbb{R}^d)$. Then $\mathcal{C}$ is sequentially weakly closed.*

Proof. Given the structure of weakly converging sequences in $\mathscr{P}_2(\mathbb{R}^d)$, we can assume without loss of generality, that $\mathcal{C}$ is bounded; let R be its diameter. Choose any measure $\nu \notin \mathcal{C}$: the claim will be achieved if we show that there exist a measure $\mu \in \mathcal{C}$ and a constant $c \in \mathbb{R}$ such that

$$\langle T_\mu^\nu - Id, T_\mu^\sigma - Id \rangle \leq c < \|T_\mu^\nu - Id\|_\mu^2 = W^2(\mu,\nu), \quad \forall \sigma \in \mathcal{C}.$$

Indeed in this case the set $\mathcal{C}$ would be included in the halfspace $\mathscr{H}^-_{T_\mu^\nu - Id;c}$ which is weakly closed by proposition 3.2 and does not contain ν. By the arbitrariness of ν we can conclude.

Let us prove our claim. Fix $a < 1$ and apply proposition 3.3 with $P = \nu$ to find a measure μ^a satisfying

$$W(\mu_t^a, \mathcal{C}) \geq atW(\nu, \mu_t^a),$$

where $\mu_t^a := (Id + t(T_{\mu^a}^\nu - Id))_\# \mu^a$. Now fix $\sigma \in \mathcal{C}$ and define $v^a = T_\mu^\nu - Id$, $w = T_\mu^\sigma - Id$. Observe that for small $t > 0$ it holds

$$\sigma_t := \left(Id + t\sqrt{1-a^2}\frac{\|v^a\|_{\mu^a}}{\|w\|_{\mu^a}} w \right)_\# \mu^a \in \mathcal{C},$$

therefore we know that

$$W^2(\mu_t^a, \sigma_t) \geq a^2 t^2 \|v^a\|_{\mu^a}^2.$$

Recalling equation (2.1), we get

$$\lim_{t \to 0^+} \frac{W^2(\mu_t^a, \sigma_t)}{t^2} = \|v^a - w\|_\mu^2,$$

we obtain

$$\left\| v^a - \sqrt{1-a^2}\frac{\|v^a\|_{\mu^a}}{\|w\|_{\mu^a}} w \right\|_{\mu^a}^2 \geq a^2 \|v^a\|_{\mu^a}^2.$$

Some algebraic manipulations show that the previous inequality implies

$$\langle v^a, \sqrt{1-a^2}\frac{\|v^a\|_{\mu^a}}{\|w\|_{\mu^a}} w \rangle_{\mu^a} \leq \|v^a\|_{\mu^a} \|w\|_{\mu^a} \sqrt{1-a^2} \leq R\|v^a\|_{\mu^a} \sqrt{1-a^2}.$$

By choosing a near to 1 and observing that $\|v^a\|_{\mu^a} \leq W(\nu, C) + R$ we get that the last term of the previous inequality is close to 0. Therefore it is smaller than $d^2(\nu, \mathcal{C})$, which in turn is smaller then $W^2(\nu, \mu^a)$ and the claim is achieved. □

References

[1] M. Agueh, *Asymptotic behavior for doubly degenerate parabolic equations*, C.R. Math. Acad. Sci. Paris, 337 (2003), pp. 331–336.

[2] L. Ambrosio, N. Gigli and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Second edition, Birkhäuser Verlag (2008).

[3] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math., 44 (1991), pp. 375–417.

[4] J.A. Carrillo, R.J. McCann and C.Villani, *Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates*, Rev. Mat. Iberoamericana, 19 (2003), pp. 971–1018.

[5] A. Brondsted, *On a lemma of Bishop and Phelps*, in Partial Differential Equations and Related Topics, J. A. Goldstein, ed., vol. 446 of Lecture Notes in Mathematics, Spinger, Berlin, 1975, pp. 65–74.

[6] E. De Giorgi, *New problems on minimizing movements*, in Boundary Value Problems for PDE and Applications, C. Baiocchi and J.L. Lions, eds. Massons, 1993, pp. 81–98.

[7] N. Gigli, *On the geometry of the space of measures in $\mathbb{R}^d$ endowed with the quadratic optimal transportation distance*, Scuola Normale Superiore, Pisa, PhD thesis, 2008, available at: http://cvgmt.sns.it/people/gigli/.

[8] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174.

[9] C. Villani, *Topics in optimal transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2003.

[10] C. Villani, *Optimal transport, old and new*, Springer Verlag, 2008.