OPTIMAL TRANSPORT: THE EULERIAN APPROACH

Yann Brenier[*]

1 Generalized geodesics and Optimal Transport

1.1 Generalized geodesics

Although we could consider the general framework of a Riemannian manifold, we only address the case of a subset D of the Euclidean space $\mathbf{R}^d$, and we assume D to be the closure of a convex open bounded set. Given two points X_0 and X_1 in D, the geodesic curve

$$X(s) = (1-s)X_0 + sX_1 \tag{1}$$

achieves

$$\inf_X \int_0^1 k(X'(s))ds, \tag{2}$$

for all continuous convex even function k on $\mathbf{R}^d$, in particular for the 'quadratic cost' $k(v) = |v|^2/2$, among all smooth paths $s \in [0,1] \to X(s) \in D$ such that $X(1) = X_1$, $X(0) = X_0$. This immediately follows from Jensen's inequality. In the spirit of Young's generalized functions [Yo], [Ta], let us now associate to each admissible path X the following pair of (Borel) measures (ρ, m) defined on the compact set $[0,1] \times D$ by

$$\rho(s,x) = \delta(x - X(s)), \quad m(s,x) = X'(s)\delta(x - X(s)), \quad (s,x) \in [0,1] \times D. \tag{3}$$

They satisfy the following compatibility condition in the sense of distributions

$$\partial_s \rho + \nabla \cdot m = 0, \;\; \rho(0,\cdot) = \rho_0, \;\; \rho(1,\cdot) = \rho_1, \tag{4}$$

where

$$\rho_0(x) = \delta(x - X_0), \quad \rho_1(x) = \delta(x - X_1). \tag{5}$$

[*]CNRS, LMO,Université Paris-Saclay, yann.brenier@universite-paris-saclay.fr

Indeed, we have

$$-\int_D \int_0^1 (\partial_s \phi(s,x) d\rho(s,x) + \nabla \phi(s,x) \cdot dm(s,x)) \tag{6}$$
$$+\int_D (\phi(1,x) d\rho_1(x) - \phi(0,x) d\rho_0(x)) = 0,$$

for all smooth functions $\phi(s,x)$ defined on $[0,1] \times \mathbf{R}^d$. (This also implies, in a weak sense, that m is parallel to the boundary ∂D.) We notice that m is absolutely continuous with respect to ρ and, by Jensen's inequality,

$$\int_0^1 k(X'(s)) ds \tag{7}$$

is bounded from below by

$$K(\rho, m) = \int k(v) d\rho, \tag{8}$$

where $v(s,x)$ is the Radon-Nikodym derivative of E with respect to ρ. A more precise definition of K can be given in terms of the Legendre-Fenchel transform of k denoted by k^* and defined by

$$k^*(y) = \sup_{x \in \mathbb{R}^d} x \cdot y - k(x), \tag{9}$$

where $\cdot$ denotes the inner product in $\mathbb{R}^d$. We assume k^* to be continuous on $\mathbb{R}^d$. Typically

$$k(x) = \frac{|x|^p}{p}, \quad k^*(y) = \frac{|y|^q}{q}, \quad \frac{1}{p} + \frac{1}{q} = 1, \quad 1 < p, q < +\infty,$$

where $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^d$. We have

$$K(\rho, m) = \sup_{\alpha, \beta} \int_D \int_0^1 \alpha(s,x) d\rho(s,x) + \beta(s,x) \cdot dm(s,x), \tag{10}$$

where the supremum is performed over all pair (α, β) of respectively real and vector valued continuous defined on $[0,1] \times D$ subject to satisfy

$$\alpha(s,x) + k^*(\beta(s,x)) \leq 0 \tag{11}$$

pointwise. (Indeed, it can be easily checked that, with this definition, $K(\rho, m)$ is infinite unless i) ρ is nonnegative, ii) m absolutely continuous with respect to ρ and has a Radon-Nikodym density v, ii) $K(\rho, m)$ is just the ρ integral of $k(v)$.) Notice that K is a convex functional, valued in $[0, +\infty]$.

It is now natural to consider the infimum, denoted by $\inf K$, of functional K defined by (10) among *all* pairs (ρ, m) that satisfy compatibility conditions (4), with data (5), and *not only* among those which are of form (3). This new minimization problem is convex (as the original one). Since the class of admissible solutions has been enlarged, the following upper bound follows

$$\inf K \le k(X_1 - X_0) \tag{12}$$

(by using (1) and (3) as an admissible pair). It turns out that there is no gap between the original infimum and the relaxed one.

Theorem 1.1 *The infimum of functional K, defined by (10), among all pair $(\rho, m)(s, x)$ of measures on $[0, 1] \times D$, satisfying (4), with boundary conditions*

$$\rho(0, x) = \delta(x - X_0), \quad \rho(1, x) = \delta(x - X_1), \tag{13}$$

is achieved by the one associated, through (3), to the straight path between the end points $X(s) = (1 - s)X_0 + sX_1$.

Proof

The proof is obtained through the following simple, and typical, duality argument that will be used several times subsequently in these lecture notes. First, we use (6) to relax constraint (4) and write

$$\inf K = \inf_{\rho,m} \sup_{\alpha,\beta,\phi} \int_D \int_0^1 (\alpha(s, x) - \partial_s \phi(s, x)) d\rho(s, x) \tag{14}$$

$$+(\beta(s, x) - \nabla\phi(s, x)) \cdot dm(s, x) + \int_D (\phi(1, x)d\rho_1(x) - \phi(0, x)d\rho_0(x)),$$

where (α, β) are subject to (11), and ϕ should be considered as a Lagrange multiplier for condition (4).

The *formal* optimality conditions for (α, β, ϕ) are

$$\alpha = \partial_s \phi, \;\; \beta = \nabla\phi, \;\; \alpha + k^*(\beta) = 0, \tag{15}$$

which leads to the Hamilton-Jacobi equation

$$\partial_s \phi + k^*(\nabla\phi) = 0. \tag{16}$$

Thus, a good guess for (α, β, ϕ) is

$$\phi(s,x) = x \cdot y - sk^*(y), \quad \alpha = \partial_s \phi, \quad \beta = \nabla \phi, \tag{17}$$

where $y \in \mathbb{R}^d$ will be chosen later. From definition (14), we deduce, with such a guess,

$$\inf K \geq \phi(1, X_1) - \phi(0, X_0) = (X_1 - X_0) \cdot y - k^*(y),$$

for all $y \in \mathbb{R}^d$. Optimizing in y and using that

$$k(x) = \sup_{y \in \mathbb{R}^d} x \cdot y - k^*(y),$$

we get

$$\inf K \geq k(X_1 - X_0),$$

i.e. the reverse inequality of (12), which concludes the proof.

1.2 Extension to probability measures

The main advantage of the concept of generalized geodesics (ρ, m) as minimizers of $K(\rho, m)$ subject to (4) is that (ρ, m) can achieve boundary data

$$\rho(s = 0, \cdot) = \rho_0, \quad \rho(s = 1, \cdot) = \rho_1, \tag{18}$$

that are (Borel) probability measures defined on the subset D. Probability measures should be seen in this context, as generalized (or fuzzy) points.

Theorem 1.2 *Let (ρ_0, ρ_1) a pair of probability measures on D. Then* $\inf K$ *is always finite and does not differ from the Monge-Kantorovich generalized distance between ρ_0 and ρ_1 usually defined by*

$$I_k(\rho_0, \rho_1) =: \inf \int_{D^2} k(x - y) d\mu(x, y), \tag{19}$$

where the infimum is performed on all nonnegative measures μ on $D \times D$ with projections ρ_0 and ρ_1 on each copy of D.

This alternative 'Eulerian' formulation of the OT problem has been established (in the special and most important case $k(v) = |v|^2/2$) in [BB] essentially for numerical purposes and inspired by Fluid Mechanics. It has been used since for a lot of different purposes (including the theory of gradient flows as in [AGS]) and generalized in many different directions (OT on graphs, quantum OT, Boltzmann equations, matrix-valued OT, etc...).

Proof

The proof requires the following well-known result of OT theory, known as Kantorovich duality [RR, Sa, Vi] :

$$I_k(\rho_0, \rho_1) = \sup \int_D (\phi_1(x) d\rho_1(x) - \phi_0(x) d\rho_0(x)), \tag{20}$$

where ϕ_1 and ϕ_0 are continuous functions on D subject to

$$\phi_1(y) \le k(x-y) + \phi_0(x), \quad \forall x,\ y \in D. \tag{21}$$

From definition (19), there is always a minimizer μ, so that

$$\int_{D^2} k(b-a) d\mu(a,b) = I_k(\rho_0, \rho_1).$$

Let us introduce, for this μ,

$$\rho(s,x) = \int_{D^2} \delta(x - X(s,a,b)) d\mu(a,b), \tag{22}$$

$$m(s,x) = \int_{D^2} \partial_s X(s,a,b) \delta(x - X(s,a,b)) d\mu(a,b), \tag{23}$$

where

$$X(s,a,b) = (1-s)a + sb. \tag{24}$$

Just as in the proof of Theorem 1.1, compatibility condition (4) is satisfied and, by Jensen's inequality,

$$\int_{D^2} k(b-a) d\mu(a,b) \ge K(\rho, m) \tag{25}$$

$$\ge \inf K \ge \sup_\phi \int_D (\phi(1,x) d\rho_1(x) - \phi(0,x) d\rho_0(x)),$$

where

$$\partial_s \phi + k^*(\nabla \phi) \le 0. \tag{26}$$

So we can choose ϕ to be a solution of the Hamilton-Jacobi equation (16) on D. Using the Hopf formula to solve (16) (see [Ba], [Li]), we get

$$\phi(s,x) = \inf_{y \in D} (\phi(0, x + s(y-x)) + sk(y-x)),$$

for all $s \geq 0$. Thus, from (25), we finally get

$$I_k(\rho_0, \rho_1) = \int_{D^2} k(b-a) d\mu(a,b) \geq \inf K \geq \sup_\phi \int_D (\phi(1,x) d\rho_1(x) - \phi(0,x) d\rho_0(x)),$$

where

$$\phi(1,x) = \inf_{y \in D} (\phi(0,y) + k(y-x)).$$

Thus, we conclude, using Kantorovich duality (20), that there is no difference between $I_k(\rho_0, \rho_1)$ and $\inf K$, which concludes the proof.

Remark

Strictly speaking, this proof is not complete. Indeed the Hopf formula does not provide, in general, a C^1 function but only a locally Lipschitz one (which means that the partial derivatives on ϕ, in the sense of distributions, are just locally L^∞ functions with respect to the Lebesgue measure). To complete the proof, a regularization argument is convenient: from the Hopf solution ϕ, we can get by convolution a C^1 approximation ϕ_ϵ which satisfies the Hamilton-Jacobi *convex inequality* $\partial_s \phi_\epsilon + k^*(\nabla \phi_\epsilon) \leq 0$ and does not affect too much the value $\int_D (\phi(1,x) d\rho_1(x) - \phi(0,x) d\rho_0(x))$. We leave to the reader the technical details (which require some careful truncations before performing convolutions, due to the boundedness of the domain $[0,1] \times D$).

1.3 A decomposition result

From the proof of Theorem 1.2, we immediately get the following *decomposition* result that asserts that generalized geodesics are mixtures of classical geodesics.

Theorem 1.3 *Each pair (ρ_0, ρ_1) of (Borel) probability measures on D admits a generalized geodesic (ρ, E) linking them with the following structure*

$$\rho(s,x) = \int_{D^2} \delta(x - X(s,a,b)) d\mu(a,b), \tag{27}$$

$$E(s,x) = \int_{D^2} \partial_s X(s,a,b) \delta(x - X(s,a,b)) d\mu(a,b), \tag{28}$$

where $X(\cdot, a, b)$ is the shortest path between a and b in D

$$X(s,a,b) = (1-s)a + sb, \tag{29}$$

and μ is a probability measure on D^2 with projections ρ_0 and ρ_1 on each copy of D.

1.4 Duality theory

Our proof has established that there is no gap between the 'inf-sup' and the 'sup-inf'. However this result has been indirectly obtained through the Monge-Kantorovich approach to the OT problem. Therefore, it is useful to establish *directly* the duality result *without* relying on the MK theory. This can be easily done thanks to the Fenchel-Rockafellar duality theorem, which can be used in many problems of convex optimization. For simplicity, we limit ourself to the quadratic cost $k(v) = |v|^2/2$.

The Fenchel-Rockafellar duality theorem

Theorem 1.4 *Let E be a real Banach space and consider two functions $K_1,\ K_2 : E \to \mathbb{R} \cup \{+\infty\}$ which are both convex. Assume that there exists a point $u_0 \in E$ such that both K_1 and K_2 are finite at u_0 while K_2 is continuous at u_0. Then we have the duality equality*

$$\sup_{u \in E} \left(-K_1(u) - K_2(u)\right) = \inf_{f \in E'} \left(K_1^*(-f) + K_2^*(f)\right),$$

where E' is the dual of E and the Legendre-Fenchel dual $K^* : E' \to \mathbb{R} \cup \{+\infty\}$ of a function $K : E \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$K^*(f) = \sup_{u \in E} \left[\langle f, u\rangle_{E',E} - K(u)\right].$$

Moreover, the infimum in the duality equality is achieved by some point $f \in E'$.

Remark.
Surprisingly enough, this duality theorem (for which a proof is given in [Brz]) is quite similar to the Plancherel formula in harmonic analysis. Indeed, at least formally, one can consider the correspondence between the algebraic structures with operations, respectively, $[+, \cdot]$ and $[\max, +]$ (sometimes in this correspondence, inequalities can show up instead of equalities). Then, the Legendre-Fenchel transform is analogous to the Fourier transform and the duality equality just corresponds to the Plancherel formula:

$$\int u \cdot v = \int \hat{u} \cdot \hat{v},$$

where $u \to \hat{u}$ stands for the Fourier transform. This "Fenchel-Fourier" dictionary is now well established in Mathematics ("Tropical Geometry" in Algebraic Geometry being one of the best known examples).

Application of the Fenchel-Rockafellar theorem

We introduce

$$E = C^0([0,1] \times D; \mathbb{R} \times \mathbb{R}^d),$$

which is a Banach space for the sup norm, and define two convex functions K_1 and K_1 on E, valued in $]0, +\infty]$, as follows. We first set

$$K_1(A, B) = -\int_{D^2} \phi(1,x) d\rho_1(x) - \phi(0,x) d\rho_0(x),$$

whenever there are $\phi \in C^1([0,1] \times D)$ such that

$$A(t,x) = \partial_t \phi(t,x), \quad B(t,x) = \nabla \phi(t,x),$$

and $K_1(A, B) = +\infty$ otherwise. Then, we define

$$K_2(A, B) = 0, \;\text{ if }\; A(t,x) + \frac{|B(t,x)|^2}{2} \leq 0, \quad \forall (t,x) \in [0,1] \times D.$$

and $K_2(A, B) = +\infty$ otherwise.

Notice that the first definition is consistent, in the sense that if A, B are represented as above by two different ϕ, then the value of $K_1(A, B)$ is unchanged.

Lemma 1.5 *The functionals $K_1, K_2 : E \to \mathbb{R} \cup \{+\infty\}$ verify the hypotheses of Theorem 1.4.*

Proof

The convexity condition is clear. Next, we have to find a function u_0 in E having the required properties in the Fenchel-Rockafellar Theorem. We observe here that there is no chance that K_1 is continuous (for the C^0-norm) because arbitrarily near any function where $K_1 < +\infty$ there is some function with $K_1 = +\infty$. On the other side, in the point $(A_0, B_0) = (-1, 0)$ we have $A_0 = \partial_t \phi_0, B_0 = \nabla \phi_0$ for $\phi_0 = -t$, so K_1 is finite at this point. On the other side, $K_2(A_0, B_0) = 0$ and this condition is preserved for small perturbations of (A_0, B_0) in the C^0-norm. Therefore the assumptions of Theorem 1.4 are satisfied, which completes the proof.

We now want to exploit Theorem 1.4 in our setting. We start by noticing that

$$K_2^*(\rho, m) = K(\rho, m),$$

where K is nothing but the functional introduced at the beginning of this chapter. Let us now compute $K_1^*(-\rho,-m)$. By definition,

$$K_1^*(-\rho,-m) = \sup_{\phi,p} \int_{t,x} -\partial_t \phi(t,x)\rho(t,x) - \nabla\phi(t,x)\cdot m(t,x)$$

$$+\int_x \phi(1,x)d\rho_1(x) - \phi(0,x)d\rho_0(x).$$

This exactly means that $K_1^*(-\rho,-m)$ takes value ∞ unless

$$\partial_t\rho + \nabla\cdot m = 0, \quad \rho(0,x) = \rho_0(x), \quad \rho(1,x) = \rho_1(x),$$

in which case $K_1^*(-\rho,-m) = 0$. So, we conclude that

$$\sup_{\rho,m} K_1^*(-\rho,-m) + K_2^*(\rho,m) = K_{opt}(\rho_0,\rho_1)$$

which corresponds to the inf-sup problem while

$$\sup_{A,B} -K_1(A,B) - K_2(A,B)$$

is (almost by definition) just the value of the sup-inf problem that we have computed earlier. So the inf-sup and the sup-inf have the same optimal value and we can state:

Theorem 1.6 *The Eulerian OT problem can be successively written in primal (sup) and dual (inf) form:*

$$\sup_\phi \int_D \phi(1,x)d\rho_1(x) - \phi(0,x)d\rho_0(x),$$

subject to

$$\partial_t\phi(t,x) + \frac{|\nabla\phi(t,x)|^2}{2} \leq 0, \quad \forall (t,x) \in [0,1]\times D$$

and

$$\inf_{\rho,m} K(\rho,m), \quad K(\rho,m) = \frac{1}{2}\int_{t,x} |v(t,x)|^2\rho(t,x), \quad m = \rho v,$$

subject to

$$\partial_t\rho + \nabla\cdot m = 0, \;\; \rho(0,x) = \rho_0(x), \;\; \rho(1,x) = \rho_1(x),$$

(m being parallel to the boundary ∂D) and there is at least an optimal solution (ρ,m) to the second one.

2 Incompressible OT

As seen earlier, the OT on a subset D of $\mathbb{R}^d$ (still assumed to be the closure of a bounded convex open set), is just a theory of generalized geodesics, or, equivalently, following Otto's point of view [Ot2], a theory of geodesics on the "manifold" $Prob(D)$ of all probability measures on D. It is therefore natural to extend this idea to more complex (convex) "manifolds". The most interesting case, in our opinion, is the the set $DS(D)$ of all doubly stochastic probability measures on D, namely the set of all Borel measures μ on $D \times D$ having as projection on each copy of D the (normalized) Lebesgue measure on D, which means

$$\int_{D\times D} f(x)d\mu(x,y) = \int_{D\times D} f(y)d\mu(x,y) = \int_D f(x)dx,$$

for all continuous functions f on D. As $d > 1$, this compact convex set turns out to be just the weak closure of the group of orientation and volume preserving diffeomorphisms of D, usually denoted by $SDiff(D)$ [AK], [Ne], through the following embedding

$$g \in SDiff(D) \to \mu_g \in DS(D), \quad \mu_g(x,y) = \delta(y - g(x)).$$

This group is of particular importance because it is the configuration space of incompressible fluids. $SDiff(D)$ is naturally embedded in the space $L^2(D, \mathbb{R}^d)$ of all square Lebesgue integrable maps from D to $\mathbb{R}^d$. Therefore, $SDiff(D)$ inherits the L^2 metric. Then, as pointed out by Arnold [AK], the equations of geodesic curves along $SDiff(D)$ exactly are the Euler equations of incompressible inviscid fluids (see also [MP] and [Br3]).
In our framework, it is very easy to define generalized geodesic curves on $DS(D)$.

Definition 2.1 *Given μ_0, μ_1 in $DS(D)$, we define a (minimizing) generalized geodesic curve joining μ_0 and μ_1 to be a pair (μ, E) of (Borel) measures defined on $Q = [0,1] \times D \times D$ and valued in $\mathbb{R}_+ \times \mathbb{R}^d$ such that*

$$\int_Q \partial_s f(s,x,y)d\mu(s,x,y) + \nabla_x f(s,x,y) \cdot dE(s,x,y) \tag{30}$$

$$= \int_{D^2} f(1,x,y)d\mu_1(x,y) - \int_{D^2} f(0,x,y)d\mu_0(x,y),$$

for all smooth function f *on* $[0,1] \times D^2$*, and*

$$\int_Q f(s,x) d\mu(s,x,y) = \int_0^1 \int_D f(s,x) dx ds, \tag{31}$$

for all continuous function f *on* $[0,1] \times D$*, that minimizes*

$$K(\mu, E) = \sup_{\alpha,\beta} \int_Q \alpha(s,x,y) d\mu(s,x,y) + \beta(s,x,y) \cdot dE(s,x,y) \tag{32}$$

where the supremum is performed over all pair (α, β) *of continuous functions defined on* Q *respectively valued in* $\mathbb{R}$ *and* $\mathbb{R}^{md}$*, subject to satisfy*

$$\alpha(s,x,y) + \frac{1}{2}|\beta(s,x,y)|^2 \leq 0. \tag{33}$$

pointwise.

This can be seen as a multiphasic OT problem, where to each point $y \in D$ is attached a "phase" described by $\mu(\cdot,\cdot,y)$ and $E(\cdot,\cdot,y)$. These phases are coupled by constraint (31) which forces the different phases to share the volume available in D during their motion. Not surprisingly, this makes the optimality equations more subtle than in the classical OT. Indeed, there is a Lagrange multiplier corresponding to constraint (31) that physically speaking is the pressure $p(s,x)$ of the fluid at each point $x \in D$ and each $s \in [0,1]$. The formal optimality conditions read

$$E(s,x,y)/\mu(s,x,y) = e(s,x,y), \quad e(s,x,y) = \nabla_x \phi(s,x,y), \tag{34}$$

$$\partial_s \phi(s,x,y) + \frac{1}{2}|\nabla_x \phi(s,x,y)|^2 + p(t,x) = 0.$$

This multiphasic OT problem has been studied in details in [Br3, AF1, AF2] and related to the classical Euler equations. (In some cases it is shown that generalized geodesics can be approximated by classical solutions to the Euler equations with vanishing forcing.) To motivate further researches, let us just quote two results that are true for any pair of data (μ_0, μ_1) in $DS(D)$. First, ∇p is uniquely defined (although there may be several generalized minimizing geodesic between μ_0 and μ_1). This fact follows easily from convex duality, but is rather surprising from the classical fluid mechanics point of view. Next, ∇p has some limited regularity [Br3, AF1, AF2]: it is a locally bounded measures in the interior of $[0,1] \times D$, which is not obvious and follows from the minimization principle and its duality properties. Further regularity can therefore be expected. We conjecture that p is semi-concave in x (which implies that the *second* order derivatives in space of p are measures and not only the first order ones, as already established), in particular as the data μ_0 and μ_1 are absolutely continuous with respect to the Lebesgue measure on D^2 with smooth positive density, as in Caffarelli's regularity theory of the (quadratic) OT [Ca].

References

[AF1] L. Ambrosio, A. Figalli, *On the regularity of the pressure field of Brenier's weak solutions to incompressible Euler equations,* Calc. Var. Partial Differential Equations 31 (2008) 497-509.

[AF2] L. Ambrosio, A. Figalli, *Geodesics in the space of measure-preserving maps and plans,* Archive for Rational Mechanics and Analysis 194 (2009) 421-462.

[AGS] L. Ambrosio, N.Gigli, G. Savaré *Gradient flows in metric spaces and the Wasserstein spaces of probability measures,* Lectures in Mathematics, ETH Zurich, Birkhäuser, 2005.

[AK] V.I.Arnold, B.Khesin, *Topological methods in Hydrodynamics, Springer Verlag, 1998.*

[Ba] G. Barles, *Solutions de viscosité des équations d'Hamilton-Jacobi,* Mathématiques et applications 17, Springer, 1994.

[BB] J.-D. Benamou, Y. Brenier, *A Computational Fluid Mechanics solution to the Monge-Kantorovich mass transfer problem, to appear in Numerische Math.*

[Br0] Y.Brenier, *Décomposition polaire et réarrangement monotone des champs de vecteurs, C. R. Acad. Sci. Paris Sér. I Math. 305 (1987), no. 19, 805–808.*

[Br1] Y.Brenier, *Polar factorization and monotone rearrangement of vector-valued functions, Comm. Pure Appl. Math. 64 (1991) 375-417.*

[Br3] Y.Brenier, *Minimal geodesics on groups of volume-preserving maps, Comm. Pure Appl. Math. 52 (1999) 411-452.*

[Brz] H. Brezis, *Analyse fonctionnelle appliquée,* Masson, 1974.

[Ca] L.A. Caffarelli, *Boundary regularity of maps with convex potentials. Ann. of Math. (2) 144 (1996), no. 3, 453-496.*

[EGn] L.C. Evans, W. Gangbo, *Differential equations methods for the Monge-Kantorovich mass transfer problem, Mem. Amer. Math. Soc. 137 (1999), no. 653.*

[EGr] L.C. Evans, Gariepy, R. *Measure theory and fine properties of functions, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1992.*

[GM] W. Gangbo, R. McCann, *The geometry of optimal transportation,* Acta *Math. 177 (1996) 113-161.*

[GMS] M. Giaquinta, G. Modica, J. Souček, *Cartesian currents in the calculus of variations. II. Variational integrals,* Springer-Verlag, Berlin, 1998.

[JKO] D. Kinderlehrer, R. Jordan, F. Otto, *The variational formulation of the Fokker-Planck equation, SIAM J. Math. Anal. 29 (1998), no. 1,1–17.*

[Ka] L.V. Kantorovich, *On a problem of Monge, Uspekhi Mat. Nauk. 3 (1948), 225–226.*

[KS] M. Knott, C.S. Smith, *On the optimal mapping of distributions, J. Optim. Theory Appl. 43 (1984), no. 1, 39–49.*

[Li] P.-L. Lions, *Generalized solutions of Hamilton Jacobi equations, Research Notes in Mathematics, 69. Pitman, Boston, Mass.-London, 1982.*

[Mc1] R. McCann, *A convexity principle for interacting gases,* Adv. Math. *128 (1997) 153-179.*

[Mc2] R. McCann, *Polar factorization of maps on Riemannian manifolds, Geom. Funct. Anal. 11 (2001) 589-608.*

[MP] C. Marchioro, M. Pulvirenti, *Mathematical theory of incompressible nonviscous fluids, Springer, New York, 1994.*

[Ne] Y. Neretin, *Categories of bistochastic measures and representations of some infinite-dimensional groups, Sb. 183 (1992), no. 2, 52-76.*

[Ot1] F. Otto, *Evolution of microstructure in unstable porous media flow: a relaxational approach,. Comm. Pure Appl. Math. 52 (1999) 873-915.*

[Ot2] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation, Comm. Partial Differential Equations 26 (2001) 101-174.*

[OV] F. Otto, C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality, J. Funct. Anal. 173 (2000) 361-400.*

[RR] S.T. Rachev, L. Rüschendorf, *Mass transportation problems, Vol. I and II. Probability and its Applications, Springer-Verlag, New York, 1998.*T

[Sa] F. Santambrogio, *Optimal transport for applied mathematicians,* Birkhäuser/Springer, 2015.

[Su] V.N. Sudakov, *Geometric problems in the theory of infinite-dimensional probability distributions, Proceedings of the Steklov Institute 141 (1979) 1–178.*

[Ta] L. Tartar, *The compensated compactness method applied to systems of conservation laws. Systems of nonlinear PDE, NATO ASI series, Reidel,Dordecht, 1983.*

[Vi] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, 58, AMS, Providence, 2003.

[Yo] L. C. Young, *Lectures on the calculus of variations. Chelsea,New York, 1980.*