Optimal Transport and Applications

Summer School, Lake Arrowhead*

Oct 6th - Oct 11th 2013

Organizers:

Nestor Guillen
University of California, Los Angeles, USA

Dimitri Shlyakhtenko
University of California, Los Angeles, USA

Christoph Thiele
University of California, Los Angeles, USA / University of Bonn, Germany

*supported by NSF grand DMS 1001535

Contents

1 Stability in the anisotropic isoperimetric inequality

after A. Figalli, F. Maggi and A. Pratelli [1]
A summary written by Marcos Charalambides

Abstract

We characterize near-minimizers for the anisotropic isoperimetric inequality. As an application, we obtain a sharp, stable version of the Brunn-Minkowski inequality for convex bodies.

1.1 The anisotropic isoperimetric inequality

Fix a dimension $n \geq 2$ and let K be an open, bounded, convex subset of $\mathbb{R}^n$ which contains the origin. The *anisotropic perimeter* of a 'nice' (open with smooth boundary, say) set $E \subset \mathbb{R}^n$ with outer unit normal vector ν_E is

$$P_K(E) := \int_{\partial E} \|\nu_E\|_* d\mathcal{H}^{n-1} \tag{1}$$

where $\mathcal{H}^{n-1}$ is $(n-1)$-dimensional Hausdorff measure and, for $\nu \in S^{n-1}$, $\|\nu\|_*$ denotes the support function

$$\|\nu\|_* := \sup\{x \cdot \nu \ : \ x \in K\}. \tag{2}$$

In the case when K is the unit ball centered at the origin, this notion of perimeter agrees with the Euclidean perimeter.

In general, we have the scaling law

$$P_K(\lambda E) = \lambda^{n-1} P_K(E), \lambda > 0 \tag{3}$$

but, by contrast with the Euclidean perimeter, the anisotropic perimeter is not invariant under the action of $SO(\mathbb{R}^n)$.

We have the following isoperimetric inequality which generalizes the classical Euclidean one.

Theorem 1. *Let K and E be as above. Then,*

$$P_K(E) \geq n|K|^{\frac{1}{n}}|E|^{\frac{n-1}{n}}. \tag{4}$$

This may be deduced from the Brunn-Minkowski inequality. Indeed, for $\epsilon > 0$,

$$\frac{|E + \epsilon K| - |E|}{\epsilon} \geq \frac{(|E|^{\frac{1}{n}} + \epsilon |K|^{\frac{1}{n}})^n - |E|}{\epsilon} \tag{5}$$

and, as $\epsilon \to 0^+$, the inequality converges to (4) (provided E is nice enough).

An alternative proof of Theorem 1 based on mass transport techniques, was given by Gromov [2]. By using Gromov's argument as a starting point, the authors prove a stable version of the anisotropic isoperimetric inequality, i.e. they show that sets E for which $P_K(E)$ is close to optimal are actually close to K (up to translation and scaling).

Theorem 2. *Let K and E be as above with $0 < |E| < \infty$. If*

$$P_K(E) \leq (1 + \delta) n |K|^{\frac{1}{n}} |E|^{\frac{n-1}{n}}, \tag{6}$$

then there exists $x_0 \in \mathbb{R}^n$ such that, for $r > 0$ defined by $r^n|K| = |E|$,

$$|E \triangle (x_0 + rK)| \leq C_n \delta^{\frac{1}{2}} |E|. \tag{7}$$

This result substantially improves the previously best known bound [3] which has a worse exponent for δ of around n^{-2} and a constant C_n which, in addition to depending on n, also depends on K.

As an application, the authors obtain a sharp stable version of the Brunn-Minkowski inequality for convex bodies; this is discussed in Section 1.4.

The proof of Theorem 2 begins by applying Gromov's argument with a different transport map than the Knothe map considered by Gromov. Instead, the Brenier map [4] is used, which allows for a more direct argument classifying minimizers in the isoperimetric inequality. Combining the argument with a stable version of the arithmetic-geometric mean inequality and an appropriate Sobolev-Poincaré inequality then gives good control for the near-minimal case.

1.2 Gromov's proof revisited

Given a smooth bounded set $E \subset \mathbb{R}^n$, the Brenier-McCann Theorem [4], [5] implies that there is a convex function $\phi : \mathbb{R}^n \to \mathbb{R}$ whose gradient $T = \nabla \phi$ (the Brenier transport map) pushes forward the probability measure $|E|^{-1}\mathbf{1}_E(x)dx$ to the probabilty measure $|K|^{-1}\mathbf{1}_K(x)dx$.

In this section, we sketch Gromov's argument applied to this transport map. We assume, by rescaling if necessary, that $|E| = |K|$.

Remark 3. *We will assume, in particular, that T is smooth whenever convenient. To make the proof rigorous, we can use results on functions of bounded variation; this is done in full detail in [1].*

From the measure transportation property we may deduce, in particular, that T maps E into K and, for $x \in E$,

$$\det \nabla T(x) = 1. \tag{8}$$

Since ϕ is convex, ∇T is positive definite and symmetric so we may diagonalize it. Writing $\lambda_1(x) \leq \ldots \leq \lambda_n(x)$ for the (measurable) eigenvalues, it follows that $\det \nabla T = \prod_{j=1}^n \lambda_j$ and $\operatorname{div} T = \sum_{j=1}^n \lambda_j$.

By the arithmetic-geometric mean inequality, we deduce that

$$(\det \nabla T)^{\frac{1}{n}} \leq \frac{\operatorname{div} T}{n}. \tag{9}$$

We define

$$\|x\| := \inf\{r > 0 \ : \ r^{-1}x \in K\}. \tag{10}$$

Then $\|x\| \leq 1$ if and only if x lies in the closure of K. We immediately deduce the inequality

$$x \cdot y \leq \|x\| \|y\|_*. \tag{11}$$

Combining (11), (8), (9), the Divergence Theorem and the fact that $\|T(x)\| \leq 1$ for $x \in \partial E$, we obtain

$$\begin{aligned} n|K|^{\frac{1}{n}}|E|^{\frac{n-1}{n}} &= \int_E n(\det \nabla T)^{\frac{1}{n}} \\ &\leq \int_E \operatorname{div} T \\ &= \int_{\partial E} T \cdot \nu_E \, d\mathcal{H}^{n-1} \\ &\leq \int_{\partial E} \|T\| \|\nu_E\|_* \, d\mathcal{H}^{n-1} \\ &\leq P_K(E), \end{aligned} \tag{12}$$

which proves Theorem 1.

Using the Brenier map, it is (formally) easy to characterize the minimizers of the inequality. Indeed, equality implies that $n(\det \nabla T)^{\frac{1}{n}} = \operatorname{div} T$ on E. Thus, we have equality in the arithmetic-geometric mean inequality which in turn imples that $\lambda_1 = \ldots = \lambda_n$ on E. By (8), we deduce that $\nabla T = \mathrm{Id}$. Consequently, E must be a translate of K.

1.3 From equality to near-equality

By (12), if E satisfies $P_K(E) \leq (1+\delta)n|K|^{\frac{1}{n}}|E|^{\frac{n-1}{n}}$ for small $\delta > 0$, then

$$\int_{\partial E}(1 - \|T\|)\|\nu_E\|_* d\mathcal{H}^{n-1}. \tag{13}$$

Furthermore, $\nabla T \approx \mathrm{Id}$ and we have the following quantitative bound:

$$\int_E |\nabla T - \mathrm{Id}| \leq Cn^2|K|\delta^{\frac{1}{2}}. \tag{14}$$

Here, we have endowed the space of $n \times n$ matrices with the trace norm $|S| = (tr(S^T S))^{\frac{1}{2}}$. The bound is a consequence of the following stable version of the arithmetic-geometric mean inequality.

Lemma 4. *Suppose that $0 < \lambda_1 \leq \ldots \leq \lambda_n$ and let λ_A and λ_G denote the arithmetic and geometric mean of $\{\lambda_1, \ldots, \lambda_n\}$ respectively. Then there exists $C > 0$ such that*

$$\frac{1}{\lambda_n}\sum_j (\lambda_j - \lambda_G)^2 \leq Cn^2(\lambda_A - \lambda_G). \tag{15}$$

Armed with this lemma, we calculate

$$\begin{aligned}
(\int_E |\nabla T - \mathrm{Id}|)^2 &= \left(\int_E \sqrt{\sum_j (\lambda_j - 1)^2}\right)^2 \\
&\leq \|\lambda_n\|_{L^1(E)} \left(\int_E \sum_j \frac{(\lambda_j - \lambda_G)^2}{\lambda_n}\right) \\
&\leq \|\lambda_n\|_{L^1(E)} \int_E Cn^2(\lambda_A - \lambda_G) \\
&= Cn^2 \|\lambda_n\|_{L^1(E)} \int_E \frac{\mathrm{div} T}{n} - (\det \nabla T)^{\frac{1}{n}} \\
&\leq Cn^2 \|\lambda_n\|_{L^1(E)} \delta |K|.
\end{aligned}$$

Furthermore, $|\lambda_n - 1| \leq |\nabla T - \mathrm{Id}|$ so $\|\lambda_n\|_{L^1(E)} \leq |K| + \int_E |\nabla T - \mathrm{Id}|$ from which we may bound $\|\lambda_n\|_{L^1(E)}$ by $Cn^2|K|$ and deduce (14).

The idea now is to apply a Sobolev-Poincaré inequality on E to control an appropriate norm of $T - \mathrm{Id}$ which would in turn control $|E\Delta K|$. But

E may not be connected or regular enough for this to work. Nonetheless, the authors prove that, for sufficiently small δ, there is a large subset of E for which a Sobolev-Poincaré inequality does hold and it is then a simple reduction to replace E by this subset.

Remark 5. *One difficulty now is that this subset may not inherit much of the smoothness of E. As a consequence, all of the inequalities need to be generalized to the setting of sets of finite perimeter using more sophisticated geometric measure theory. Full details may be found in [1].*

With this reduction, the Sobolev-Poincare inequality takes the form

$$\| - \nabla f\|_*(E) \geq C_{n,K} \inf_c \int_{\partial E} \|f - c\| \|\nu_E\|_* d\mathcal{H}^{n-1}. \tag{16}$$

Applying this to $f = T - \mathrm{Id}$ and using inequality (14) yields

$$\int_{\partial E} \|T - \mathrm{Id}\| \|\nu_E\|_* d\mathcal{H}^{n-1} \leq C_{n,K} \delta^{\frac{1}{2}} \tag{17}$$

up to translation of E.

By (13), (17) and the triangle inequality,

$$\int_{\partial E} |1 - \|x\|| \|\nu_E\|_* d\mathcal{H}^{n-1}(x) \leq C_{n,K} \delta^{\frac{1}{2}}. \tag{18}$$

It is not difficult to see that the left hand side controls $|E \Delta K|$; see [1, Lemma 3.5]. The proof is then completed by using a renormalization argument to show that, in fact, $C_{n,K}$ can be chosen to be independent of K.

1.4 Stability in the Brunn-Minkowski inequality

Suppose that E and F are open, bounded, convex subsets of $\mathbb{R}^n$ containing the origin.

Observe that, for nice G,

$$P_{E+F}(G) = P_E(G) + P_F(G). \tag{19}$$

Therefore, by Theorem 1,

$$\begin{aligned} n|E + F| &= P_{E+F}(E + F) \qquad (20) \\ &= P_E(E + F) + P_F(E + F) \\ &\geq n|E + F|^{1-\frac{1}{n}} (|E|^{\frac{1}{n}} + |F|^{\frac{1}{n}}) \end{aligned}$$

which gives another proof of the Brunn-Minkowski inequality. The characterization of minimizers for the anisotropic isoperimetric inequality immediately implies that equality in the Brunn-Minkowski inequality holds if and only if E and F are homothetic.

If we apply Theorem 2 on $P_E(E+F)$ and $P_F(E+F)$, a straightforward argument then yields the following stable version.

Theorem 6. *Let E and F be as above and suppose that*

$$|E+F|^{\frac{1}{n}} \leq (1+\delta)(|E|^{\frac{1}{n}} + |F|^{\frac{1}{n}}). \tag{21}$$

Then there exists $x_0 \in \mathbb{R}^n$ such that, for $r > 0$ defined by $|E| = r^n|F|$ and $\sigma = \max\{|E|/|F|, |F|/|E|\}$,

$$|E\Delta(x_0 + rF)| \leq C_n \delta^{\frac{1}{2}} \sigma^{\frac{1}{2n}} |E|. \tag{22}$$

The exponents of δ and σ in this inequality are sharp. For σ, this can be seen by taking E to be the unit cube and F to be a sequence of balls whose radius tends to zero. For δ, one can take E to be the unit ball and F to be a sequence of ellipsoids approximating E. See [1, Section 4] for more details.

References

[1] Figalli, M., Maggi, F. and Pratelli, A. *A mass transportation approach to quantitative isoperimetric inequalities.* Inventiones mathematicae **182** (1) (2010), 167–211.

[2] Milman, V. D. and Schechtman, G. *Asymptotic theory of finite-dimensional normed spaces. With an appendix by M. Gromov.* Lecture Notes in Mathematics, 1200. Springer-Verlag, Berlin, 1986. viii+156 pp.

[3] Esposito, L., Fusco, N. and Trombetti, C. *A quantitative version of the isoperimetric inequality: the anisotropic case.* Ann. Sc. Norm. Super. Pisa Cl. Sci. **4** (2005), no. 4, 619–651

[4] Brenier, Y. *Polar factorization and monotone rearrangement of vector-valued functions.* Comm. Pure Appl. Math. **44** (4) (1991), 375–417.

[5] McCann, R.J. *Existence and uniqueness of monotone measure-preserving maps.* Duke Math. J. **80** (2) (1995), 309–323.

Marcos Charalambides, UC Berkeley
email: **marcos@math.berkeley.edu**

2 Hölder regularity of optimal transport maps, and underlying inequalities from convex geometry

after L. Caffarelli [3], N. Guillen and J. Kitagawa [6]
A summary written by Nicholas Cook

Abstract

In [2], Caffarelli considers convex solutions to the Monge-Ampère equation, and uses the method of barriers to prove estimates on their growth away from a supporting hyperplane. He then uses these estimates to establish localization and regularity properties of solutions. Here we describe an route to Caffarelli's estimates which goes through convex geometry rather than the method of barriers. This is carried out in [6] for general costs; in these notes we simplify the exposition by focusing on the quadratic cost function.

2.1 Introduction

Given subsets $\Omega, \widetilde{\Omega} \subset \mathbb{R}^d$ with measures $d\mu = \rho dx, d\tilde{\mu} = \tilde{\rho} dx$ absolutely continuous with respect to Lebesgue measure, and a cost funtion $c : \Omega \times \widetilde{\Omega} \to \mathbb{R}$, the optimal transport problem seeks a measurable map $T : \Omega \to \widetilde{\Omega}$ such that $T_{\#}\mu = \tilde{\mu}$, and

$$\int_\Omega c(x, T(x)) d\mu(x) = \inf_{S_{\#}\mu = \tilde{\mu}} \int_\Omega c(x, S(x)) d\mu(x). \tag{1}$$

For the *quadratic cost* function $c(x, \tilde{x}) = -x \cdot \tilde{x}$ it was shown by Brenier that there exists a unique solution $T = Du$ which is the gradient of some convex potential function u (see [1]). This result was extended to more general cost functions by Gangbo and McCann [4].

Due to the characterization $T = Du$, it can be seen that solutions to (1) satisfy an equation of Monge-Ampère type:

$$\det(D^2 u(x) + D^2_{x,\tilde{x}} c(x, T(x))) = |\det D^2_{x,\tilde{x}} c(x, T(x))| \rho(x)/\tilde{\rho}(T(x)). \tag{2}$$

For the case of the quadratic cost function this reduces to the classical Monge-Ampère equation

$$\det(D^2 u) = \rho(x)/\tilde{\rho}(Du(x)). \tag{3}$$

With the assumption that the marginals $\mu, \tilde{\mu}$ satisfy

$$\lambda \leq \frac{\rho(x)}{\tilde{\rho}(\tilde{x})} \leq \Lambda \tag{4}$$

for a.e. $(x, \tilde{x}) \in \mathrm{spt}\rho \times \mathrm{spt}\tilde{\rho}$, and some constants $0 < \lambda \leq \Lambda < \infty$, we concern ourselves with solutions to the inequalities

$$\lambda \leq \det(D^2 u) \leq \Lambda. \tag{5}$$

In [2], Caffarelli introduced a geometric approach to show that *weak* solutions to (5) in fact possess $C^{1,\alpha}$ regularity (and hence an optimal transport map T is α-Hölder continuous). At the core of his proofs are two lemmas providing upper and lower bounds for the growth of a convex solution u away from an affine function ℓ that slices the graph of u.

Lemma 1. *Suppose u is convex in Ω and $\det D^2 u \leq \Lambda < \infty$. Then for any affine function ℓ we have*

$$\ell(x) - u(x) \ll_{\Lambda,n} |\{u \leq \ell\}|^{2/n} d(x, \Pi)^{2/n} \tag{6}$$

where Π is any supporting hyperplane to the sublevel set $\{u \leq \ell\}$.

(Here and in the sequel, $f(x) \ll_p g(x)$ means $f(x) \leq C(p) g(x)$.)

Lemma 2. *Suppose that u is convex in Ω and $\det D^2 u \geq \lambda > 0$. Then for any affine function ℓ we have*

$$\sup_{\{u \leq \ell\}} \ell - u \gg_{\lambda,n} |\{u \leq \ell\}|^{2/n}. \tag{7}$$

Combining these lemmas we have

Corollary 3. *Let u be a convex solution of $\lambda \leq \det D^2 u \leq \Lambda$ in Ω, and let ℓ be an affine function. Fix some $\delta \in (0,1)$. Then for any point $x_0 \in \{u \leq \ell\}$ such that*

$$u(x_0) - \ell(x_0) \leq \delta \inf_{\{u \leq \ell\}} u - \ell \tag{8}$$

and any hyperplane Π supporting the convex set $\{u \leq \ell\}$, we have

$$d(x_0, \Pi) \gg_{\delta,\lambda,\Lambda,n} 1. \tag{9}$$

Caffarelli goes on to establish 1) that convex solutions u to (5) are either strictly convex, or else their contact set with some affine function crosses the domain; 2) at each point in the domain, a strictly convex solution will have a unique supporting hyperplane; 3) strictly convex solutions are in $C^{1,\alpha}$. The proofs proceed by contradiction, showing how the negation of any of these implies the existence of a family of affine functions ℓ_ε and supporting hyperplanes Π_ε to $\{u \leq \ell_\varepsilon\}$, as well as points on the interior x_ε satisfying (8) for which $d(x_\varepsilon, \Pi)$ can be made arbitrarily small, contradicting (9); see [3] for details.

Caffarelli established Lemmas 1 and 2 by using the affine invariance of the Monge-Ampère equation and the method of barriers. Affine invariance is not available when dealing with the equation (2) for general cost function c, which motivates searching for an alternative approach. In [6] it was noted that these lemmas in fact follow from basic inequalities from convex geometry, and analogous lemmas were obtained for the case of general cost function.

The purpose of these notes is to describe the convex geometric approach to the above lemmas, for the simplest case of the quadratic cost function. We will actually establish a variant of Lemma 1, which is still sufficient to obtain Corollary 3:

Lemma 4. *Suppose u is convex in Ω and $\det D^2u \leq \Lambda < \infty$. Then for any affine function ℓ we have*

$$\ell(x) - u(x) \ll_{\Lambda,n} |\{u \leq \ell\}|^{2/n} \left(\frac{d(x, \Pi_+ \cup \Pi_-)}{d(\Pi_+, \Pi_-)} \right)^{1/n} \tag{10}$$

where Π_+, Π_- are parallel supporting hyperplanes to the sublevel set $\{u \leq \ell\}$.

2.2 Geometric approach to Lemmas 1 and 4

Let A be an origin-symmetric convex body in $\mathbb{R}^n$, i.e. a compact convex set with non-empty interior. We define its polar body by

$$A^o := \{\xi \in \mathbb{R}^n : |\langle \xi, x \rangle| \leq 1 \text{ for all } x \in A\}.$$

Denote by B_p the unit ball in ℓ_p^n. One can verify with Hölder's inequality that $B_p^o = B_{p'}$, where p' is the dual exponent to p.

An alternative description of the polar body is as follows. For a convex body A, define the cone over A with *height* $h > 0$, and *center* $x_0 \in K$ by

$$K_{A,x_0,h}(x) := \sup \ell(x)$$

where the supremum is taken over all affine functions such that $\ell(x_0) \leq h$ and $\ell(x)) \leq 0$ on ∂A. Then for A origin symmetric, we have

$$A^o = \partial K_{A,0,1} \tag{11}$$

where ∂f denotes the subdifferential of f. This motivates extending the definition of polar body for non-origin symmetric convex bodies:

$$A^o_{x_0,h} = \partial K_{A,x_0,h}. \tag{12}$$

An important affine-invariant of an origin-symmetric convex body is the *Mahler volume*

$$M(A) := |A||A^o|.$$

It was proven by Santaló in [8] that the Mahler volume is maximized by the euclidean unit ball (and hence all ellipsoids). The Mahler conjecture is that it is minimized by the cube, i.e. the unit ball in ℓ_n^∞ (and hence also by its dual, the cross-polytope). Lower bounds on the Mahler volume are known as reverse-Santaló inequalities. For generalized polar bodies we have the following inequality of reverse-Santaló type:

Lemma 5. *Let A be a convex subset of $\mathbb{R}^n$ and let Π_+, Π_- denote parallel supporting hyperplanes to A. Then for $x \in A$, $h > 0$ we have*

$$|A||A^o_{x,h}| \geq n^{-n}\lambda^n \frac{d(\Pi_+, \Pi_-)}{d(x, \Pi_+ \cup \Pi_-)}. \tag{13}$$

Lemma 5 quickly implies Lemma 4. Indeed, taking $A = \{u \leq \ell\}$ and $h = \ell(x) - u(x)$, we have

$$\ell(x) - u(x) \leq n \left(\frac{d(x, \Pi_+ \cup \Pi_-)}{d(\Pi_+, \Pi_-)} \right)^{1/n} |\{u \leq \ell\}|^{1/n} |\partial K_{A,x,\ell(x)-u(x)\}}|^{1/n} \tag{14}$$

where we have used (12). Now from convexity of u it follows that

$$\partial K_{A,x,\ell(x)-u(x)\}}(x) = \partial K_{A,x,\ell(x)-u(x)\}}(A) \subset \partial u(A) \tag{15}$$

where $\partial u(A) := \cup_{x \in A} \partial u(x)$. Now since u weakly solves (5), $\partial u(A) \leq \Lambda |A|$. Combining this with (15) and (14) proves (10).

Lemma 1 can similarly be established using an inequality of Santaló type − see [6] for details. We briefly remark on the proof of Lemma 5. Recall the following useful theorem of Fritz John [7]:

Theorem 6. *Let A be a convex body in $\mathbb{R}^n$. Then there is an ellipsoid $\mathcal{E}$ with center at the origin and a point $p \in \mathbb{R}^n$ such that*

$$p + \mathcal{E} \subset A \subset p + n\mathcal{E}.$$

This allows one to reduce to the case that $B(0,1) \subset A \subset B(0,n)$. Then if we let $S(\Pi_+, \Pi_-)$ denote the slab bounded by Π_+, Π_-, the containment $A \subset B(0,n) \cap S(\Pi_+, \Pi_-)$ allows one to locate a sufficiently large subset of $A^o_{x,h}$, leading to the estimate (13).

References

[1] Brienier, Y., *Décomposition polaire et réarrangement monotone des chaps de vecteurs.* C.R. Acad. Sci. Paris Sér. I Math., 305(19):805-808, 1987.

[2] Caffarelli, L.A., *A localization property of viscosity solutions to the Monge-Ampère equation and their strict convexity.* Ann. of Math. (2), 131(1):129-134, 1990.

[3] Caffarelli, L.A., *A priori estimates and the geometry of the Monge-Ampère equation.* In: *Nonlinear partial differential equations in differential geometry, IAS/Park City Math. Ser* 2 (1992), pp. 5-63.

[4] Gangbo, W. and McCann, R.J., *The geometry of optimal transportation.* Acta Math., 177(2):113-161, 1996.

[5] McCann, R.J. and Guillen, N., *Geometry, Regularity and Applications.* In: *Analysis and Geometry of Metric Measure Spaces: Lecture Notes of the 50th Séminaire de Mathématiques Supérieures (SMS), Montréal, 2011* 56 (2013), p. 145.

[6] Guillen, N. and Kitagawa, J., *On the local geometry of maps with c-convex potentials.*Preprint at arXiv:math/1212.4865v1.

[7] John, F., *Extremum problems with inequalities as subsidary conditions*, Studies and Essays presented to R. Courant on his 60th birthday, Jan 9. 1948, Interscience Publishers Inc., New York, NY 1948, 187-204.

[8] Santaló, L., *Un unvariante afín para los cuerpos convexos del espacio de n dimensiones*, Portugalie Math. **8** (1949), 155-161.

NICHOLAS COOK, UCLA
email: **nickcook@math.ucla.edu**

3 A convexity theory for interacting gases and equilibrium crystals

after R. J. McCann [1]
A summary written by Matías G. Delgadino

Abstract

The notion of displacement interpolation is defined for probabilities in $\mathbb{R}^d$. Using this, the class of displacement convex functionals is defined. The stictly displacement convex functional have a unique minimizer in the space of probabilities, just as in the classical case .

3.1 Introduction

The notion of convexity is elemental and central in several areas in Mathematics. In [1], McCann was able to generalize this notion to some functionals defined in the space of probabilities in $\mathbb{R}^d$. Using a classical result by Brennier [2], he was able to define curves interpolating probabilities, which in fact are geodesics. This curves act as the straight segments in the case of functions defined in $\mathbb{R}^d$. Displacement convexity is proven for some particular functionals that model internal, potential and interaction energy of a gas or fluid. Using a convexity argument, existence and uniqueness of energy minimizers is shown.

3.2 Interpolation of Probability Measures

To begin defining interpolation on $\mathcal{P}(\mathbb{R}^d)$, we need to define how a measurable map acts on measures.

Definition 1. *Given X and Y measure spaces, with a map $T : X \to Y$ measurable. Then, if ω measure in X, then T induces a measure $T_{\#}\omega$ in Y, by the following action*

$$T_{\#}\omega[A] = \omega[T^{-1}(A)].$$

$T_{\#}\omega$ is called the push-forward of ω through T; it is a probability measure if ω is.

One of the first question that naturally arises is, if given $\rho_0, \rho_1 \in \mathcal{P}(\mathcal{R}^d)$, can we find a transformation that pushes ρ_0 onto ρ_1. And the follow up, if there is more than one of these maps, can we find one of them that has particular properties under mild assumptions on ρ_0. This was answered by Brennier [2] and expanded by McCann [1].

Theorem 2. *Let $\rho_0 \in \mathcal{P}_{ac}(\mathbb{R}^d)$ and $\rho_1 \in \mathcal{P}(\mathbb{R}^d)$. There is a convex function ψ, whose gradient $\nabla\psi$ pushes forward ρ_0 to ρ_1. Moreover, $\nabla\psi$ is uniquely determined ρ_0-almost everywhere.*

With Theorem 2 we are able to give the following definition of an interpolant between two probability measures.

Definition 3. *Given $\rho_0,\ \rho_1 \in \mathcal{P}(\mathbb{R}^d)$, with ρ_0 absolutely continuous. At time $t \in [0,1]$, the displacement interpolant $\rho_t \in \mathcal{P}(\mathbb{R}^d)$ between ρ_0 and ρ_1 is defined by*

$$\rho_t = [(1-t)I + t\nabla\psi]_{\#}\rho_0, \tag{1}$$

where I is the identity map and $\nabla\psi$ is uniquely determined by Theorem 2.

Remark 4. *A displacement interpolant may still be defined even if neither of the end points ρ_0 and ρ_1 are absolutely continuous, though the interpolant might not be unique anymore. We can do this by considering $p \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with cyclically monotone support having ρ_0 and ρ_1 as its marginals. We can always find such a p by solving the Kantorovich's problem with quadratic cost. Let $t \in [0,1]$ and define*

$$\Pi_t(x,y) = (1-t)x + ty,$$

on $\mathbb{R}^d \times \mathbb{R}^d$. Then,

$$\rho_t = \Pi_{t\#}p. \tag{2}$$

This definition coincides with the latter one, due to the characterization of cyclically monotone sets as graphs of convex functions. For further reference see Chapter 2 [3].

Lets see what are the properties of the displacement interpolant.

Proposition 5. *Let $\rho_0, \rho_1 \in \mathcal{P}(\mathbb{R}^d)$ be probability measures with $\rho_0 \in \mathcal{P}_{ac}(\mathbb{R}^d)$. For $t \in [0,1]$, the displacement interpolant $\rho_t = \rho_0 \xrightarrow{t} \rho_1$ satisfies:*

1. *ρ_t is uniquely determined;*
2. *ρ_t is absolutely continuous for $t < 1$;*

3. $\rho_0 \xrightarrow{t} \rho_1 = \rho_1 \xrightarrow{1-t} \rho_0$;

4. *If* $t, t' \in [0,1]$, *then* $\rho_t \xrightarrow{s} \rho_{t'} = \rho_0 \xrightarrow{(1-s)t+st'} \rho_1$.

Remark 6. *Calculating explicitly,* $\mathcal{W}_2(\rho_0, \rho_t) = t\mathcal{W}_2(\rho_0, \rho_1)$. *Together with the previous properties, we see that the displacement interpolant is a constant speed geodesics in* $\mathcal{P}_{ac}(\mathbb{R}^d)$ *with the* $\mathcal{W}_2$ *distance.*

Remark 7. *Item 3. can be interpreted as* $\mathcal{P}_{ac}(\mathbb{R}^d)$ *is a displacement convex subset of* $\mathcal{P}(\mathbb{R}^d)$, *and the remaining measures lie on its boundary.*

3.3 Displacement Convexity

Now we are able to give a new notion of convexity:

Definition 8. *Given a map* $J : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$, *we will call it displacement convex, if for every* $\rho_0, \rho_1 \in \mathcal{P}_{ac}(\mathbb{R}^d)$, $J(\rho_t)$ *is a convex function of* t *on* $[0,1]$, *where* ρ_t *is the displacement interpolant between* ρ_0 *and* ρ_1.

We are going to consider three basic examples:

- **Internal Energy**

$$\mathcal{U}(\rho) = \int_{\mathbb{R}^d} U(\rho(x))dx, \quad U : \mathbb{R}_+ \to \mathbb{R} \cup \{+\infty\} \ \text{ measurable;} \tag{3}$$

- **Potential Energy**

$$\mathcal{V}(\rho) = \int_{\mathbb{R}^d} V(x)d\rho(x), \quad V : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\} \ \text{ measurable;} \tag{4}$$

- **Interaction Energy**

$$\mathcal{W}(\rho) = \frac{1}{2}\int_{\mathbb{R}^d \times \mathbb{R}^d} W(x-y)d\rho(x)d\rho(y), \quad W : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\} \ \text{ measurable;} \tag{5}$$

Theorem 9. *Let* $\mathcal{P}$ *be a displacement convex subset of* $\mathcal{P}(\mathbb{R}^d)$, *on which* $\mathcal{U}$, $\mathcal{V}$ *and* $\mathcal{W}$ *are well-defined with values in* $\mathbb{R} \cup \{+\infty\}$. *Then,* *(i) If U satisfies $U(0) = 0$ and*

$$\Psi : r \to r^n U(r^{-n}) \text{ is convex nonincresing on } (0, +\infty), \tag{6}$$

then $\mathcal{U}$ is displacement convex on $\mathcal{P}$. Conversely, if Ψ is nonincresing and $\mathcal{U}$ is displacement convex on $\mathcal{P}_{ac}(\mathbb{R}^d)$, then Ψ is convex.

(ii) If V is convex, then $\mathcal{V}$ is displacement convex on $\mathcal{P}$. Conversely, if $\mathcal{V}$ is displacement convex on $\mathcal{P}_2(\mathbb{R}^d)$, then V is convex.

(iii) If W is convex, then $\mathcal{W}$ is displacement convex on $\mathcal{P}$. If W is strictly-convex, then for all $m \in \mathbb{R}^d$ $\mathcal{W}$ is strictly-displacement convex on $\mathcal{P}_m$, the subset of $\mathcal{P}$ with prescribed mean m. Conversely, if $\mathcal{W}$ is displacement convex on $\mathcal{P}_2(\mathbb{R}^d)$, then W is convex.

Remark 10. *Condition (6) can be interpreted as $\mathcal{U}$ being displacement convex under dilations. Moreover, if U is differentiable we can reformulate the condition in terms of the thermodynamical pressure ($P(\rho) = \rho U'(\rho) - U(\rho)$) by,*

$$\rho \to \frac{P(\rho)}{\rho^{1-\frac{1}{n}}} \text{ is nondecreasing.} \tag{7}$$

Remark 11. *For items (ii) and (iii), we can exchange convexity in V and W by strictly convex, semi-convexity and λ-uniformly convex, obtaining respective convexity properties for the Energies.*

3.3.1 Applications of Convexity

Using this new notion of convexity, we can prove existence and uniqueness of minimizers to the functionals assuming strict convexity, similarly to the case when we try to minimize a strictly convex function in $\mathbb{R}$.

Theorem 12. *Take $\mathcal{F} = \mathcal{V} + \mathcal{W} + \mathcal{U}$, such that $\mathcal{V}$, $\mathcal{W}$ and $\mathcal{U}$ are displacement convex in $\mathcal{P}_{ac}$, and either $\mathcal{V}$ or $\mathcal{W}$ is strictly displacement convex. Then, there is a unique minimizer $\rho_0 \in \mathcal{P}_{ac}$ of $\mathcal{F}(\rho)$, up to translation, if $\mathcal{V}$ is not strictly convex.*

The proof of the theorem uses displacement convexity both for proving existence and uniqueness of the minimizer.

Remark 13. ***Open Problem:***

Is there another examples of displacement convex functionals, that are not of the form $\mathcal{U}$, $\mathcal{V}$ or $\mathcal{W}$? For example, any functional that depends on the derivatives of the measure?

References

[1] McCann, Robert John. *A convexity theory for interacting gases and equilibrium crystals.* Diss. Princeton University, 1994.

[2] Brenier, Yann. *Polar factorization and monotone rearrangement of vector?valued functions.* Communications on pure and applied mathematics 44.4 (1991): 375-417.

[3] Villani, C?dric. *Topics in optimal transportation.* Vol. 58. AMS Bookstore, 2003.

Matías G. Delgadino, University of Maryland
email: **mdelga@math.umd.edu**

4 Differential equations methods for the Monge-Kantorovich mass transfer problem

after L.C. Evans and W. Gangbo [3]
A summary written by Taryn C. Flock

Abstract

The authors of [3] construct a solution to the classical Monge-Kantorovich Problem by studying the p-Laplacian equation in the limit as $p \to \infty$. We summarize their results.

4.1 Introduction

4.1.1 The Monge problem

At issue is Monge's optimal transport problem with the cost function $|x - y|$ for measures which are absolutely continuous with respect to Lebesgue measure. Given are two nonnegative L^1 functions f^+, f^-, satisfying $\int_{\mathbb{R}^n} f^+ dx = \int_{\mathbb{R}^n} f^- dy$. Let $\mu^+ = f^+ dx$ and $\mu^- = f^- dy$. The goal is to find a function s which transports μ^+ to μ^- optimally. More precisely, a measurable function r is said to transport μ^+ to μ^- if

$$\int h(x) f^+(x) dx = \int h(r(x)) f^-(r(x)) dx \qquad \forall \text{ continuous functions } h(x).$$

Such functions r will be referred to as mass transfer maps and the set of all mass transfer maps will be referred to by $\mathcal{A}$. A mass transfer map s is optimal if

$$\int |x - s(x)| f^+(x) dx = \inf_{r \in \mathcal{A}} \int |x - r(x)| f^+(x) dx.$$

4.1.2 The Monge-Kantorovich problem

Kantorovich in [4],[5] proposed the following relaxation of the problem which has led to many advances in the subject. Let

$$\mathcal{M} = \{\text{probablity measures } q \text{ on } \mathbb{R}^n \times \mathbb{R}^n \text{ such that } proj_x q = \mu^+ \text{ and } proj_y q = \mu^-\}$$

We now seek $p \in \mathcal{M}$ which is optimal in the sense that

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |x-y| dp(x,y) = \min_{q \in \mathcal{M}} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |x-y| dq(x,y).$$

The Monge-Kantorovich problem is a relaxation of the original Monge problem because for each $r \in \mathcal{A}$, $q_r(E) = \int_{x|(x,r(x)) \in E} f^+(x)dx \in \mathcal{M}$.

Among the advantages of Monge-Kantorovich problem is the existence of a dual maximization problem.

4.1.3 The Monge-Kantorovich dual problem

We give the statement of the dual problem presented in [3]. Let

$$\mathcal{L} = \{w : \mathbb{R}^n \to \mathbb{R}^n : \sup_{x,y} \frac{w(x) - w(y)}{|x-y|} \leq 1\}.$$

We seek $u \in \mathcal{L}$ which is optimal in the sense that

$$\int_{\mathbb{R}^n} u \left(f^+ - f^-\right) dz = \max_{w \in \mathcal{L}} \int_{\mathbb{R}^n} w(f^+ - f^-)dz.$$

This problem is dual to the original in the sense that

$$\max_{w \in \mathcal{L}} \int_{\mathbb{R}^n} w(f^+ - f^-)dz = \min_{q \in \mathcal{M}} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |x-y| dq(x,y). \tag{1}$$

4.1.4 Solving the Monge problem

Monge determined that if an optimal mass transfer map s exists, then there also exists a scalar potential function u such that

$$\frac{s(x) - x}{|s(x) - x|} = -Du(x)$$

i.e. u determines the direction the optimal transport map should move x. A solution u of the Monge-Kantorovich dual problem can be interpreted as this potential u. We could build a solution to the Monge problem from a solution u of the dual problem by simply solving for s, if in addition to u, we knew $|s(x) - x|$, the distance s moves x, for all x.

The insight of Evans and Gangbo in [3] is that this information can be found and used to construct a solution s by studying the p-Laplician.

4.2 The p-Laplacian

The p-Laplacian PDE is

$$-\text{div}(|Du_p|^{p-2}Du_p) = f \qquad (n+1 \leq p < \infty).$$

It arises as the Euler-Lagrange equation for the problem of minimizing

$$\int_{\mathbb{R}^n} \frac{1}{p}|Dw|^p - wf dx.$$

The connection to the Monge-Kantorovich dual problem is that if there exists u, satisfying $u_p \to u$ as $p \to \infty$ for $\{u_p\}$ the weak solutions of the p-Laplacian PDE, then $u \in \mathcal{L}$ and u maximizes $\int_{\mathbb{R}} wf dz$. Thus taking the initial data to be $f = f^+ - f^-$, u constructed in this manner is optimal for the Monge-Kantorovich dual problem. Further, u solves

$$-div(aDu) = f^+ - f^-,$$

where a is determined by $|Du_p|^{p-2}Du_p \rightharpoonup aDu$. This function a, known as the **transport density**, contains the information $|s(x) - x|$ for all x.

4.3 Sketch of the construction in [3]

For technical reasons, several further restrictions are placed on the Monge problem: f^+, f^- are required to be Lipschitz functions with compact support. Setting $X = supp(f^+)$ and $Y = supp(f^-)$, these supports are required to be disjoint, contained in a large ball B(0,R), and have smooth boundaries. Finally, f^+ and f^- are required to be strictly positive on the interior of their supports.

Making use of previous work showing existence of and (uniform) bounds for solutions of the p-Laplacian PDE, [3] shows for some $S > R$ to be specified:

Theorem 1. *There exists a function u, which is Lipschitz on B(0,S) and function $a \in L^\infty(B(0,S))$ such that*

$$-div(aDu) = f^+ - f^- \text{ in the weak sense on } B(0,S)$$

$$|Du| \leq 1 \text{ a.e and } a \geq 0 \text{ a.e.}$$

$$\text{for a.e. } z \quad a(z) > 0 \text{ implies } |Du| = 1$$

And further u is maximal for the Monge-Kantorovich dual problem.

This function u is used to construct the **transport set**: $T = \{z \in \mathbb{R}^n : \exists x \in X, y \in Y$ such that $u(x) - u(z) = |x - z|$ and $u(z) - u(y) = |z - y|\}$. As $u \in \mathcal{L}$, $|x - z| + |z - y| = u(x) - u(y) \leq |x - y|$, and hence the points x, y, z are colinear. In fact, the transport set can be decomposed into families of colinear points. For $z_0 \in T$, define the **transport ray** through z_0, by R_{z_0} :

$$R_{z_0} = \{z \in \mathbb{R}^n : |u(z_0) - u(z)| = |z - z_0|\}$$

R_{z_0} is the set containing z_0 along which u changes at the maximum rate 1.

Lemma 2. *Properties of the transport set*

1. $X \cup Y \subseteq T$.
2. *For almost every $z_0 \in T$, there exists a unique transport ray R_{z_0} through z_0 with an upper endpoint $a_0 \in X$ and lower endpoint $b_0 \in Y$ along which u decreases at rate 1.*
3. $|Du| = 1$ *a.e. on T.*

By optimality of u and the duality principle (1), if s is a mass transfer map such that for each $x \in X$, $s(x) \in Y \cap R_x$ then s is optimal.

Lemma 3. *Relation of a to the transport set*

1. *supp*$(a) \subseteq T$
2. *a vanishes at the endpoints of transport rays*

The idea of [3] is to define an optimal map s by $s(x) = z(1, x)$ where $z(t, z_0)$ is a solution to the ODE (cf. Dacorogna-Moser [2])

$$\dot{z}(t) = \frac{-a(z(t))}{tf^-(z) + (1-t)f^+(z(t))} Du(z(t)) \qquad z(0) = z_0.$$

Intuitively, because $|Du| = 1$ on T, a is supported on T, and a vanishes at the endpoints of transport rays, for almost every $x \in X$, $s(x) \in Y \cap R_x$. Hence if s transfers μ^+ to μ^-, then it is optimal.

Proving this rigorously and showing that s is a mass transfer map is done by an approximation argument. Let a_ϵ be the mollification a by η_ϵ (chosen carefully in [3]). Define ν_ϵ so that $a_\epsilon \nu_\epsilon$ is the mollficiation of $-(aDu)$. Moreover, let f_ϵ^+ and f_ϵ^- be the mollifications of f^+ and f^-. Define the approximate mass transfer map $s_{\epsilon,\delta}$ by $s_{\epsilon,\delta}(z_0) = z_{\epsilon,\delta}(1, z_0)$ where $z_{\epsilon,\delta}(t, z_0)$ is a solution to the ODE

$$\dot{z}_{\epsilon,\delta}(t) = \frac{-a_\epsilon(z(t))}{tf_\epsilon^-(z) + (1-t)f_\epsilon^+(z(t)) + \delta}\nu_\epsilon(z(t)) \qquad z_{\epsilon,\delta}(0) = z_0.$$

Lemma 4. *$s_{\epsilon,\delta}$ is an* **approximate mass transfer map**

$$f_\epsilon^+(z) + \delta = (f_\epsilon^-(s_{\epsilon,\delta}(z)) + \delta) \ det \ Ds_{\epsilon,\delta}(z).$$

This is proved following the method of Dacorogna-Moser [2], and requires the smoothness introduced by the mollification. The main step is proving that

$$\frac{\partial}{\partial_t}\left[(tf_\epsilon^- + (1-t)f_\epsilon^+ + \delta) \text{ det } Ds_{\epsilon,\delta}\right] = 0.$$

It then remains to check that the approximate mass transfer maps converge almost everywhere. This is rather subtle because a and Du are in not in general continuous. A key observation is that a and u are well behaved when restricted to transport rays. In particular for almost every $z_0 \in T$ $a|_{R_{z_0}}$ is locally Lipschitz along R_{z_0} and

Proposition 5. *For almost every transport ray R, for every $\sigma > 0$, there exists $C > 0$ and a tubular neighborhood N of R^σ (the set of points in R which are distance at least σ from an endpoint) such that for each point $z \in N \cap T$ where $D(u)$ exists*

$$|Du(z) - Du(\hat{z})| \leq C|z - \hat{z}|$$

where $\hat{z}$ is the projection of z onto R.

By studying the behavior of the approximate mass transfer maps along transport rays, these estimates provide sufficient control to show that approximations do in fact converge almost everywhere.

4.4 Historical Remark

[3] was the first to show existence of an optimal transport in $\mathbb{R}^n$ for $n \geq 3$. Later works construct optimal transport maps in greater generality, using approximate transport maps coming from solutions of the Monge problem with the strictly convex costs $|x - y|^p$ for $p > 1$ and then showing that these maps also converge to an optimal mass transport map by studying their behavior along transport rays. ([1], [6]).

References

[1] Caffarelli L.A., Feldman M., and McCann R.J., *Constructing optimal maps for Monge???s transport problem as a limit of strictly convex costs.* J. Amer. Math. Soc., **15**, 2002. 1???26

[2] Dacorogna, B. and Moser, J., *On a partial differential equation involving the Jacobian determinant*. Ann. Inst. H. Poincare, **7** (1990) 1-26

[3] Evans, L. C. and Gangbo, W., *Differential equations methods for the Monge-Kantorovich mass transfer problem.* **653** American Mathematical Soc., 1999;

[4] Kantorovich, L, *On the transfer of masses* Dokl. Akad. Nauk. SSSR, **37** (1942) 227-229

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

[6] Trudinger, N.S, and Wang, X.-J., *On the Monge mass transfer problem.* Calc. Var. Paritial Differential Equations,**13** 2001 19???31,

Taryn C Flock, UC Berkeley
email: **taryn.flock@berkeley.edu**

5 Benamou-Brenier's approach for OTT

after J.D. Benamou and Y. Brenier [2]
A summary written by Augusto Gerolin

Abstract

In this survey, we are going to expose the main ideas around the Benamou-Brenier Formula, which is a dynamical formulation of Optimal Transportation Theory (OTT). First, we introduce the main arguments of Benamou-Brenier's paper [2] and we summarize the main steps of the proof of the general case. The last section, we will be devoted for a theoretical view on the numerical aspects of that problem.

5.1 Introduction

Let $\Omega \subset \mathbb{R}^d$ be an open set and $p \geq 1$. We denote by $\mathcal{P}_p(\Omega)$, the space of probability measures with finite p^{th} moment. We know that $\mathcal{P}_p(\Omega)$ is naturally endowed by the so-called Wasserstein[1] distance W_p,

$$W_p(\mu,\nu)^p = \inf \int |T(x) - x|^p d\mu \tag{1}$$

which the infimum is taken among all maps T transporting μ to ν. When the minimum is achieved by some map T, we say that T is an optimal map and solves the L^p Monge-Kantorovich Problem (MKP).

Roughly speaking, Benamou-Brenier [2] discover, for the case $p = 2$, that the Optimal Transport Problem

$$\min \Big\{ \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |x-y|^p d\gamma(x,y) : \gamma \in \Pi(\mu,\nu) \Big\} \tag{2}$$

where the set $\Pi(\mu,\nu)$ is the set of transport plans having μ and ν as maginals, is equivalent to (see Theorem 10)

$$\inf \left\{ \int_0^1 \int_\Omega \|v_t\|^p_{L^p} d\mu_t dt : \partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0, \mu_0 = \mu, \mu_1 = \nu \right\} \tag{3}$$

[1] The name "Wasserstein/Vasershtein distance" was coined by R. L. Dobrushin in 1970, after the Russian mathematician Leonid Nasonovich Vasershtein who introduced the concept in 1969.

where the infimum is taken among all weakly continuous distributional solutions of the continuity equation having $\mu_0 = \mu$ and $\mu_1 = \nu$.

The expression (3) can be interpreted as a dynamical formulation of Optimal Transportation Theory: many trajectories of transportation given by velocity fields are admissible. That can be seen as an *Eulerian point of view* of OTT. While the Kantorovich formulation (2) is *static*, depending only on the optimal coupling of (μ, ν).

5.2 Two Motivations

1. More powerful numerical methods: In Benamou-Brenier's paper ([2]) there a mention about the needed the development of new numerical methods relate to the MKP. Let us recall the argument here. Given two compact supported probabily densities ρ_0 and ρ_1 on a domain $\Omega \subset \mathbb{R}^n$, there is an unique optimal transport map T from the support of ρ_0 to ρ_1 which can be written as the gradient of some convex funcion ψ, $T(x) = \nabla\psi(x)$. In addition, if ψ is - for instance - an Alexandrov's solution of the Monge-Ampére Equation[2],

$$det(H\psi(x))\rho_1(\nabla\psi(x)) = \rho_0(x) \tag{4}$$

where $H\psi$ represents the Hessian matrix of ψ, then ψ inherits the smoothness of both dentisities ρ_0 and ρ_1 under some additional hypothesis (see [4]).

From the Optimal Transportation Theory, the equation (4) is the natural computation solution of Monge-Kantorovich Problem in L^2. Unfortunately, by one hand, the equation (4) is fully non-linear second order elliptic equation and, by the other hand, the mass transportation problem involving (4) is not a standard boundary value problem even in the case when the density ρ_0 vanishes along a smooth subset of $\mathbb{R}^n$ [4]. Benamou-Brenier introduced an alternative numerical method for the Monge-Kantorovich Problem in L^2.

2. Non-linear Evolution Models: Several mass transportation problems comes from diffusion PDEs of the type

$$\partial_t\mu + \nabla \cdot (h(\mu)v) = 0$$

where $\mu = \rho(x)\mathcal{L}^n$ is a measure and $h(\mu)$ is a convex or concave function. That kind of equation was widely explored in the literature in applied models as crowd motion, traffic congestion, swarming models ([3], [6]).

[2]for a complete explanation about this subject see [5].

5.3 Heuristics

We are going to argue heuristically in order to give some reasons why the equivalence in between (2) and (3) can be true. In our notes, we are going to follow the ideas of [2]. Finally, for simplifying the notation, we will identify absolutely continuous measures μ on $\mathbb{R}^n$ with their respective densities ρ.

Consider measures $\rho_0(x)$ and $\rho_1(y)$. Thanks to Brenier's Theorem there exits an optimal transport map T and a convex function ψ such that $T = \nabla\psi$, $T_\sharp \nu = \mu$. Moreover, we can define a curve ρ_t which solves the continuity equation

$$\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0$$

when v is smooth, this equation captures the evolution of the spatial distribution of particles that are initially distributed according to ρ_0 and whose velocity is v. In other words, $\rho_t = X_{t\sharp}(\rho_0)$ where X_t is a velocity field defined by

$$X_0(x) = x \quad \text{ and } \quad \partial_t X_t(x) = v(t, X_t(x))$$

We are going to prove that $W_2^2(\rho_0, \rho_1) = \int_0^1 \int_\Omega |v(t,x)|^2 \rho_t(x) dx dt$. Consider a pair (ρ, v) satisfying the constraint of (3) with v being as smooth as we need, and let X_t be a velocity flow define above. We have

$$B(\rho_0, v) = \int_0^1 \int_\Omega |v(t,x)|^2 \rho_t(x) dx dt = \int_0^1 \int_\Omega |v(t, X_t(x))|^2 \rho_0(x) dx dt$$

First we proof that the $B(\rho_0, v) \geq W(\rho_0, \rho_1)$.

$$\begin{aligned} B(\rho_0, v) &= \int_0^1 \int_\Omega |v(t, X_t(x))|^2 \rho_0(x) dx dt && (5) \\ &= \int_\Omega \int_0^1 |\partial_t X_t(x)|^2 dt \rho_0(x) dx \quad \textit{(Fubbini's Theorem)} && (6) \\ &= \int_\Omega \left| \int_0^1 \partial_t X_t(x) dt \right|^2 \rho_0(x) dx \quad \textit{(Jensen's Inequality)} && (7) \\ &= \int_\Omega |X_1(x) - x|^2 \rho_0(x) dx \geq W_2(\rho_0, \rho_1)^2 && (8) \end{aligned}$$

since $X_1(x)$ is a transport map from ρ_0 to ρ_1. Now, we are going to show that the equality in (8) can be archived. Consider the Brenier's map $\nabla\psi$ and the McCann's interpolation map given by

$$Y_t(x) = (1-t)x + t\nabla\psi = \nabla((1-t)|x|^2 + t\psi)$$

hence $\rho_t = Y_{t\sharp}\rho_0$. If we write $\psi_t = (1-t)|x|^2 + t\psi$, then the associated vector field is $\partial_t Y_t(x) = \nabla\psi - x = v(t, Y_t(x)) = v(t, \nabla\psi_t(x))$. Since $\nabla\psi_t$ is gradient of a convex function, it can be - at least formally! - inverted and its inverse is $\nabla\psi_t^*$. Thus, $v(t,x) = \nabla(\nabla\psi_t^*(x)) - \nabla\psi_t^*(x)$.

By construction (ρ, v) is an admissible couple for the Benamou-Brenier problem and verify

$$E(\rho, v) = \int_\Omega |\nabla\phi(x) - x|^2 \rho_0(x)dx = W_2(\rho_0, \rho_1)^2$$

5.4 Euler Equation and Optimal Transport

From this section Ω will denote a convex set in $\mathbb{R}^n$. The mail goal in the next two sections is to identify Lipschitz curves in $(\mathcal{P}_p(\Omega), W_p)$ with solutions of the continuity equation with L^p vector fields v_t, and to connect the L^p norm of v_t with the metric derivative $|\mu'|(t)$.

Definition 1. *Let (X, d) be a metric space and $\omega : [0,1] \to X$ be a curve on X. We define the metric derivative of ω at time t, denoted by $|\omega'|(t)$ through*

$$|\omega'|(t) = \lim_{h\to 0} \frac{d(\omega(t+h), \omega(t))}{|h|}$$

provided this limit exits.

We can garantee the existence of metric derivative for Lipschitz curves.

Proposition 2. *Suppose that $\omega :\to X$ is a Lipschitz continuous, then the metric derivative $|\omega'|(t)$ exists for almost every $t \in [0,1]$ and we have*

$$d(\omega(t), \omega(s)) \leq \int_t^s |\omega'|(\tau)d\tau, \qquad \textit{for } t < s$$

The next theorem relates Lipschitz curves in $\mathcal{P}_p$ with the continuity equation.

Theorem 3. *Let $(\mu_t)_{t\in[0,1]}$ be a Lipschitz curve for the distance W_p, $(p > 1)$. Then for almost every $t \in [0,1]$ there exists a vector field $v_t \in L^p(\mu_t, \mathbb{R}^n)$ such that*

(i) the continuity equation $\partial_t\mu_t + \nabla\cdot(v_t\mu_t) = 0$ is satisfied in the weak sense

(ii) for almost every t, we have $\|v_t\|_{L^p(\mu_t)} \leq |\mu'|(t)$.

Conversely, if $(\mu_t)_{t\in[0,1]}$ is a family of measures in $\mathcal{P}_p(\Omega)$ and for each t we have a vector field $v_t \in L^p(\mu_t, \mathbb{R}^n)$ with $\|v_t\|_{L^p(\mu_t)} \leq C$ then $(\mu_t)_{t\in[0,1]}$ is a Lipschitz curve for the W_p distance and for almost everywhere t,$|\mu'|(t) \leq \|v_t\|_{L^p(\mu_t)}$

Under natural hypothesis on the regularity of the family of vector fields v_t, we can show also the uniqueness of solutions in continuity equation.

5.5 Geodesic Spaces

Definition 4 (Geodesics)**.** *Given two points $x_0, x_1 \in X$. We say that a curve $\omega : [0,1] \to X$ is a geodesic in between x_0 and x_1 if it minimizes the lenght amoung all curves such that $\omega(0) = x_0$ and $\omega(1) = x_1$.*

A metric space (X, d) is said to be a *lenght space* if for all $x, y \in X$

$$d(x,y) = \inf\{\text{Lenght}(\omega) : \omega \in Lip, \omega(0) = x, \omega(1) = y\}.$$

A metric space is said to be a *geodesic space* if it is a lenght space and there exist geodesics between arbitrary points.

Remark 5. *Remind the lenght of a curve ω in a metric space is defined as*

$$\text{Lenght}(\omega) = \sup\left\{\sum_{k=0}^{n-1} d(\omega(t_k), \omega(t_{k+1}) : n \geq 1, 0 = t_0 < t_1 < \cdots < t_n = 1\right\}$$

Definition 6 (Constant speed geodesics)**.** *Let is (X, d) be a lenght space, a curve $\omega : [0,1] \to X$ is said to be a constant speed geodesics between $\omega(0) \in X$ and $\omega(1) \in X$ if it satisfies*

$$d(\omega(t), \omega(s)) = |t - s| d(\omega(0), \omega(1)), \quad \textit{for all } t, s \in [0,1]$$

Example 7. *$(\mathcal{P}_p(\Omega), W_p)$ is a geodesic space $(p \geq 1)$[3]. In fact, consider $\mu, \nu \in (\mathcal{P}_p(\Omega, W_p)$ and γ an optimal transport plan in $\Pi(\nu, \mu)$ for the cost $|x-y|^p$. Define $\pi_t : \Omega \times \Omega \to \Omega$ through $\pi_t(x,y) = (1-t)x + ty$. We can verify that the curve $\mu_t = (\pi_t)_\sharp \mu$ is a constant speed geodesic in $\mathcal{P}_p(\Omega)$ connecting $\mu_0 = \mu$ to $\mu_1 = \nu$.*

In the case where μ is absolutely continuous with respect to Lebesgue measure, every constant speed geodesic is obtained as $((1-t)Id + tT)_\sharp \mu$, where T denotes the OT map from μ to ν which the p-distance cost.

[3]Notice that here holds the case p=1

The next proposition gives a kind of converse of the last example. It states that from geodesics it is possible to reconstruct the optimal transport.

Proposition 8. *Let μ_t be a constant speed geodesic between μ and ν, and suppose $p > 1$. Then there exists an optimal plan $\gamma \in \Pi(\mu, \nu)$ such that, for every t, we have $\mu_t = (\pi_t)_\sharp \gamma$.*

Now, we give the last ingredient for the prove of Benamou-Brenier Formula. In fact, in the particular case of $\mathcal{P}_p$ spaces, constant speed geodesics may be found by minimizing $\int |\mu'|^p(t)dt$.

Proposition 9. *Let is (X, d) be a lenght space and $\omega : [0, 1] \to X$ a curve connecting two points $x_0, x_1 \in X$. For a $p > 1$ fixed, the following facts are equivalent*

1. *ω is a constant speed geodesic,*
2. *$|\omega'|(t) = d(\omega(0), \omega(1))$ almost everywhere,*
3. *ω solves $\min\{\int_0^1 |\omega'|^p dt : \omega(0) = x_0, \omega(1) = x_1\}$*

Using the notations of the example (7) and knowing that for a Wasserstein space, $|\mu'|^p(t) = \int |v_t|^p d\mu_t$, we have

Theorem 10 (Benamou-Brenier formula for $(\mathcal{P}_p, W_p)$ spaces)**.** *Let $\Omega \subset \mathbb{R}^n$ be a convex domain, $p > 1$, $\mu, \nu \in \mathcal{P}_p(\Omega)$ compacted support measures and μ_t a constant speed geodesic conecting μ and ν as defined in (7). Then,*

$$W_p(\mu, \nu)^p = \min\left\{ \int_0^1 \int_\Omega \|v_t\|^p d\mu_t dt : \partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0, \mu_0 = \mu, \mu_1 = \nu \right\} \tag{9}$$

5.6 Convex Reformulation of Benamou-Brenier

This section is widely inspired in a lecture notes witten by Filippo Santambrogio [8]. Notice that the minimization problem in the Theorem 10 in the variables (μ_t, v_t) has non-linear constraints (due the product $v_t \mu_t$) and the functional is non-convex (since $(t, x) \mapsto t|x|^p$ is not convex). However, it's possible to transform it into a convex problem. Indeed, we need just to switch variables, from (μ_t, v_t) into (μ_t, E_t) where $E_t = v_t\mu_t$. Then, we introduce a function

$$f_p(x,y) = \begin{cases} \frac{1}{p}|y|^p x^{1-p}, & \text{if } x > 0 \\ 0, & \text{if } (x,y) = (0,0) \\ \infty, & \text{otherwise} \end{cases}$$

and the minimization problem (9) becomes

$$\min\Big\{\int\int_\Omega f_p(\mu_t, E_t)dt : \partial_t\mu_t + \nabla\cdot E_t = 0, \mu_0 = \mu, \mu_1 = \nu\Big\} \qquad (10)$$

In the equation (10), the constrains are linear, the functional is convex[4] and lower semi-continuous. Benamou-Brenier used a dual convex formulation of (10) in order to develop a numerical method, called "augmented Lagrangians", in the framework of Uzawa-type Algorithm.

5.6.1 Dual Formulation

We write the constrains of (10) in the weak form

$$\min_{\mu,E}\int_0^1 dt\int_\Omega f_p(\mu_t, E_t) + \sup_\phi -\int_0^1\int_\Omega \partial_t\phi d\mu_t dt + \int_0^1\int_\Omega \nabla\phi dE_t dt + G(\phi)$$

where $G(\phi) = \int_\Omega \phi(1,x)d\nu(x) - \int_\Omega \phi(0,x)d\mu(x)$, and we notice that the quantity f_p in this variational problem may be expressed as a sup

$$f_p(x,y) = \sup\{a\cdot x + b\cdot y : a,b\in\mathbb{R}^n, |a| + \frac{1}{q}|b|^q \leq 0\}$$

and hence we solve

$$\min_{\mu,E}\sup_{q\in K,\phi}\int_\Omega a(x)d\mu_t dt + \int_\Omega b(x).dE_t dt - \int_0^1\int_\Omega \partial_t\phi d\mu_t dt - \int_0^1\int_\Omega \nabla\phi\cdot dE_t dt + G(\phi)$$

where $K = \{q(t,x) = (a(t,x), b(t,x)) : a(t,x) + \frac{1}{q}|b|^q \leq 0,\ (t,x)\}$. We can now exchange the min sup by sup min using the Legendre-Frenchel conjugate

[4]The functional is not that convex becuase is 1-homogeneous and, hence non-stricly convex. This reduces the efficiency of any gradient descent algorithm in order to solve the problem.

f_p^* (see [7]). In this way we have the following equivalence

$$\frac{1}{p}W_p(\mu_0,\mu_1) = \min\Big\{\int_0^1\int_\Omega \frac{1}{p}\frac{|\phi|^p}{\mu^{p-1}}dt : \partial_t\mu_t + \nabla\cdot(v_t\mu_t) = 0, \mu_0 = \mu, \mu_1 = \nu\Big\}$$
$$= \sup\Big\{\int_\Omega \phi(1,x)d\nu(x) - \int_\Omega \phi(0,x)d\mu(x) : \partial_t\phi + \frac{1}{q}|\nabla\phi|^q \leq 0\Big\}$$

References

[1] Ambrosio L., Gigli, N. and Savaré G., *Gradient flows in metric spaces and in the spaces of probability measures.* Birkhauser, 2005;

[2] Benamou, J.D. and Brenier, Y., *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem.* Numerische Mathematik January 2000, Volume 84, Issue 3, pp 375-393;

[3] Buttazzo G., Jimenez C., Oudet E., *An optimization problem for mass transportation with congested dynamics.* SIAM J. Control Optim., 48.

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

[5] De Phillips G., *Regularity of optimal transport maps and applications.* Birkhauser Mathematics, Publications of the SNS, 2013.

[6] Maury B., Roudneff-Chupin A., Santambrogio F., *A macroscopic crowd motion model of the gradient flow type* Math. Models Methods Appl. Sci., 20 (10) (2010).

[7] Rockafellar, R. T., *Convex Analysis* Princeton University Press. ISBN 0-691-01586-4 (1996).

[8] Santambrogio F., *Lecture notes on Optimal Transportation* available in http://www.math.u-psud.fr/∼santambr/teaching.html (Sep, 2013).

Augusto Gerolin, Università? di Pisa
email: **gerolin@mail.dm.unipi.it**

6 Polar Factorization and Monotone Rearrangment of Vector-Valued Functions

after Y. Brenier [1]
A summary written by Jordan Greenblatt

Abstract

Given an arbitrary $u \in L^p(X, \mu)$ for a probability space (X, μ) isomorphic to the unit interval (in a sense to be formalized below), it is well known that there is an essentially unique non-decreasing rearrangement $u^\# \in L^p([0,1], dx)$. We generalize this result to show that for any bounded and connected open domain $\Omega \subset \mathbb{R}^d$ with Lebesgue negligible boundary, Borel probability measure β on $\overline{\Omega}$ equivalent to Lebesgue measure, and $u \in L^p(X, \mu; \mathbb{R}^d)$, u has an essentially unique rearrangement of the form $\nabla\psi$ with $\psi \in W^{1,p}(\Omega, d\beta)$ convex on Ω. Additionally, if u does not map any positive measure sets in X to negligible sets in $\mathbb{R}^d$, there exists an essentially unique measure preserveing map $s : (X, \mu) \to (\overline{\Omega}, \beta)$ such that $u = \nabla\psi \circ s$ μ-a.e. and the maps sending u to ψ and s are continuous in the appropriate sense. Moreover, the theorem unifies seemingly distinct classical results.

6.1 Introduction

A few definitions are necessary to clarify the abstract and summarize the proof. Here (X, μ) and (Y, ν) are probability spaces.

Definition 1 (Measure preserving map). *A map $s : X \to Y$ is called measure preserving if, for all measurable $E \subset Y$, $s^{-1}(E)$ is measureable in X and $\nu(E) = \mu(s^{-1}(E))$. If there exists an injective measure preserving map from X to Y, the two spaces are said to be isomorphic.*

Definition 2 (L^p rearrangment). *For any $u \in L^p(X, \mu; \mathbb{R}^d)$, a function $v \in L^p(Y, \nu; \mathbb{R}^d)$ is called an L^p rearrangement (or simply rearrangment based on context) if for all $f \in C(\mathbb{R}^d)$ such that $|f(x)| \lesssim (1 + |x|^p)$ (all functions called f will be of this type unless otherwise stated), $\int_X f \circ u(x)\, d\mu = \int_Y f \circ v(y)\, d\nu$.*

Although not all of these assumptions will used overtly, we will assume for the remainder of the summary that $\beta(\partial\Omega) = 0$ and that β is represented by a non-negative function also called β bounded away from 0 on all compact sets. All L^p and Sobolev spaces will values in $\mathbb{R}^d$. We will further assume that all measures are Borel and $W^{1,p}(\Omega,\beta)$ embeds compactly into $L^p(\Omega,\beta)$. Also, $K \subset L^p(\Omega,\beta)$ will denote $\{\nabla\psi : \psi \in W^{1,p}(\Omega,\beta), \psi \text{ convex}\}$ and uniqueness will describe uniqueness up to a set of measure 0 where the measure is determined by context.

Theorem 3 (Rearrangement and polar factorization)**.** *For any $u \in L^p(X,\mu)$, there exists a unique rearrangement $u^\#$ of u in the set K and the map $u \mapsto u^\#$ is continuous from $L^p(X,\mu)$ to $L^p(\Omega,\beta)$. If u is non-degenerate in the sense that it does not map any positive measure sets in X to negligible sets in $\mathbb{R}^d$, there is a unique measure preserveing map $s : (X,\mu) \to (\mathbb{R}^d,\beta)$ such that $u = \nabla\psi \circ s$. Also, s is the unique maximizer among measure preserving maps of the functional $\tilde{s} \mapsto \int_X u(x)\cdot\tilde{s}(x)\,d\mu$. Finally the mapping $u \mapsto s$ is continuous from the non-degenerate elements of $L^p(X,\mu)$ to $L^q(X,\mu)$ for all $q \in [1,\infty)$.*

The existence, uniqueness, and continuity of rearrangments in K will come as a corollary of the following more general result.

Theorem 4. *For any probability measure α on $\mathbb{R}^d$ such that $\int(1+|y|^p)\,d\alpha < \infty$, there is a unique $u^\# = \nabla\psi \in K$ such that for all f, $\int f(y)\,d\alpha = \int f \circ u^\#(z)\,d\beta$. Moreover, if $\{\alpha_n\}$ is a sequence of probability measures on $\mathbb{R}^d$ such that $\int f(y)\,d\alpha_n \to \int f(y)\,d\alpha$ for all f, $\psi_n \to \psi$ in $W^{1,p}(\Omega,\beta)$ up to an additive constant.*

Setting α to be the measure given by $\alpha(E) := \mu(u^{-1}(E))$, the existence, uniqueness, and continuity of $u \mapsto u^\#$ follow quickly.

Theorem 3 generalizes a few a priori unrelated classical results. The standard non-decreacing rearrangement on $[0,1]$ is simply the case where (Ω,β) is $(0,1)$ with Lebesgue measure. This illustrates that on intervals in $\mathbb{R}$, K coincides with the set of non-decreasing functions. The matrix polar factorization on $GL_d(\mathbb{R})$ given by $A = RU$ for $R, U \in GL_d(\mathbb{R})$ positive and orthogonal respectively matches the polar factorization in theorem 3. In this application, (X,μ) and (Ω,β) are both the unit ball with normalized Lebesgue measure and $u(x) := Ax$. Moreover, the non-degeneracy condition on A in the matrix factorization theorem (i.e. $\det(A) \neq 0$) that forces U to be unique is easily seen to match the measure non-degeneracy condition on u in theorem 3 that has the same effect.

The Helmholtz decomposition theorem says the following: Let $z \in C^\infty(\overline{\Omega};\mathbb{R}^d)$ for $\Omega \subset \mathbb{R}^d$ open, bounded, connected, and having smooth boundary. Then

z can be written as $z(x) = w(x) + \nabla p(x)$ where p is a smooth real valued function and w is a smooth divergence free vector field parallel to $\partial\Omega$ with w unique and p unique up to an additive constant (the uniqueness is clear). To view the existence of the decomposition as an example of theorem 3, let $u_\epsilon := x + \epsilon z(x)$ and write $\psi_\epsilon(x) = |x|^2/2 + \epsilon p(x) + \epsilon^2 e_\psi(x, \epsilon)$ and $s_\epsilon(x) = x + \epsilon w(x) + \epsilon^2 e_s(x, \epsilon)$. Evaluating $\partial_\epsilon[\psi_\epsilon \circ s_\epsilon]_{\epsilon=0}$ provides the decomposition and applying $\int_\Omega f \circ s(x)\, dx = \int_\Omega f(x)\, dx$ to arbitrary $f \in C^\infty(\overline{\Omega})$ provides the divergence and boundary conditions on w.

6.2 Monge-Kantorovich problems

The basic strategy for proving Theorem 4 is to reduce it to common Monge-Kantorovich problems whose solutions can be described more precisely. Given a probability space $(\mathbb{R}^d, \alpha)$ with $\int (1 + |y|^p)\, d\alpha < \infty$, the three relevant problems are as follows:

Primal MKP: Find $\phi \in C(\mathbb{R}^d) \cap L^p(\mathbb{R}^d, \alpha)$ and $\psi \in C(\Omega) \cap L^p(\Omega, \beta)$ minimizing $\int \phi\, d\alpha$ subject to the constraints that $\int \psi\, d\beta = 0$ and $\phi(y) + \psi(z) \geq y \cdot z$ for all $(y, z) \in \mathbb{R}^d \times \Omega$.

Dual MKP: Find a probability measure m with marginals α and β on $\mathbb{R}^d \times \overline{\Omega}$ that maximizes $\int y \cdot z\, dm(y, z)$.

Mixed MKP: Find $\phi \in C(\mathbb{R}^d) \cap L^p(\mathbb{R}^d, \alpha)$ and $\psi \in C(\Omega) \cap L^p(\Omega, \beta)$ and a probability measure m with marginals α and β on $\mathbb{R}^d \times \overline{\Omega}$ such that $\phi(y) + \psi(z) \geq y \cdot z$ for all $(y, z) \in \mathbb{R}^d \times \Omega$, $\int \psi\, d\beta = 0$, and $\int \phi\, d\alpha \leq \int y \cdot z\, dm(y, x)$.

The thrust of the proof is showing that the mixed MKP has a unique solution and finding necessary conditions on it. These are mostly straightforward or classical and will therefore be taken as black boxes. Before listing the conditions we need the following definition:

Definition 5 (Legendre or Legendre-Frenchel Transform)**.** *For any $\theta : \Omega \to \mathbb{R}$, its Legendre transform $\theta^* : \mathbb{R}^d \to \overline{\mathbb{R}}$ is given by $\theta^*(y) := \sup_{z \in \Omega}[y \cdot z - \theta(z)]$.*

Lemma 6. *The following hold for any solution (ϕ, ψ, m) to the mixed MKP:*

1. *(ϕ, ψ) and m are the unique solutions to the primal and dual MKPs.*

2. *ψ and ϕ are convex and $\phi = \psi^*, \psi = \phi^*$ α- and β-a.e. Moreover $\psi \in W^{1,p}(\Omega, \beta)$ and $dm(y, z) = \delta[y - \nabla\psi(z)]\, d\beta(z)$.*

3. *Let $\{\alpha_n\}$ be a sequence of probability measures on $\mathbb{R}^d$ such that for any f, $\int f(y)\,d\alpha_n \to \int f(y)\,d\alpha$. If (ϕ_n, ψ_n, m_n) are solutions to the respective MKPs, then $\phi_n \to \phi$ uniformly on compact subsets of $\mathbb{R}^d$, $\psi_n \to \psi$ in $W^{1,p}$, and $\int f(y,z)\,dm_n \to \int f(y,z)\,dm$ for all f.*

4. *If α is absolutely continuous, then $\nabla\phi$ and $\nabla\psi$ are inverses α- and β-a.e. and $dm(y,z) = \delta[z - \nabla\phi(y)]\,d\alpha(y)$.*

Then, assuming the MKP has a solution, we show that theorem 4 follows: By lemma 6, if (ϕ, ψ, m) solves the MKP, $dm = \delta[y - \nabla\psi(z)]\,\beta(z)$ solves the dual MKP so for any f, $\int f(y)\,d\alpha = \int f(y)\,dm(y,z) = \int f(\nabla\psi(z))d\beta$ with $\psi \in W^{1,p}$ convex. Setting $u^{\#} = \nabla\psi \in K$ proves the first portion of the theorem. The continuity statement in the theorem follows directly from the third item of lemma 6.

6.3 Existence and uniqueness of solutions to the mixed MKP

Uniqueness follows directly from the first item in lemma 6. For existence, we initially assume that the support of α is contained in B_R and will then extend to arbitrary α. More precisely, given such an α, let α_n be the normalized restriction to $E_n \subset \mathbb{R}^d$ where $\{E_n\}$ is an increasing exhaustion of $\mathbb{R}^d$. Then the weak convergence of $\{\alpha_n\}$ to α in conjunction with the continuity property in lemma 6 gives the desired result.

We will use as a black box the classical convex analysis result that there exists a probability measure m with marginals α and β such that $\int y \cdot z\,dm(y,z)$ attains the infemum of the set

$$\left\{ \int \phi\,d\alpha + \int \psi\,d\beta : \phi \in C(\overline{B_R}), \psi \in C(\overline{\Omega}), \phi(y) + \psi(z) \geq y \cdot z \ \forall (y,z) \in \overline{B_R} \times \overline{\Omega} \right\}$$

Based on this result, if the infemum above is attained by some pair (ϕ, ψ) and m is the solution measure, (ϕ, ψ, m) satisfies the mixed MKP. In order to find this minimizing pair, first let $\{(\phi_n, \psi_n)\}$ be an infemizing sequence. Without loss of generality we can substract a constant from ϕ_n and add it to ψ_n to enforce $\min_{x \in \overline{B_R}} \phi_n(x) = 0$ for all n. Then let $\tilde{\psi}_n(z) := \sup_{y \in \overline{B_R}} \{y \cdot z - \phi_n(y)\}$ and $\tilde{\phi}_n(y) := \sup_{z \in \overline{\Omega}} \{y \cdot z - \tilde{\psi}_n(z)\}$. A simple calculation shows $\tilde{\psi}_n(0) = 0$, $\mathrm{Lip}(\tilde{\psi}_n) \leq R$, and $\tilde{\psi}_n(z) \leq \psi_n(z)$ for all n. In particular, this means the sequence $\{\tilde{\psi}_n\}$ is uniformly bounded, uniformly Lipschitz, and dominated by $\{\psi_n\}$. Similar calculations reveal the same about $\{\tilde{\phi}_n\}$ (replacing $\{\psi_n\}$ with $\{\phi_n\}$ and R with r where $\overline{\Omega} \subset B_r$).

Then, by the Arzela-Ascoli theorem, passing to a subsequence there is a pair $(\phi, \psi) \in C(\overline{B_R}) \times C(\overline{\Omega})$ such that $\tilde{\phi}_n \to \phi$ and $\tilde{\psi}_n \to \psi$ uniformly. As $\{\tilde{\phi}_n, \tilde{\psi}_n\}$ is an infemizing sequence for the functional $(g, h) \mapsto \int g\, d\alpha + \int h\, d\beta$ and both measures are finite, this uniform convergence shows that (ϕ, ψ) is indeed a minimizing pair and thus the mixed MKP has a solution.

6.4 Proof of polar factorization theorem

Letting α be the measure on $\mathbb{R}^d$ given by $\alpha(E) = \mu(u^{-1}(E))$, the non-degeneracy condition on u is equivalent to α being absolutely continuous. Then, if (ϕ, ψ, m) is the solution to the mixed MKP associated with α, the map $s := \nabla\phi \circ u$ preserves measure (this can be checked directly). Also by lemma 6, $\nabla\psi \circ \nabla\phi(y) = y$ α-a.e. so, by the definition of α, $\nabla\psi \circ s(x) = (\nabla\psi \circ \nabla\phi) \circ u(x) = u(x)$ μ-a.e. As $\nabla\psi \in K$, this completes the existence proof. The following lemma for the uniqueness proof is straightforward and will be taken as a black box:

Lemma 7. *If $u^{\#} \in K$ is an L^p rearrangement of u and ψ is the function in $W^{1,p}(\Omega, d\beta)$ such that $\int \psi\, d\beta = 0$ and $u^{\#} = \nabla\psi$, then (ϕ, ψ, m) is the solution to the mixed MKP where $\phi = \psi^*$ and $dm(y, z) = \delta[y - \nabla\psi(z)]\, d\beta(z)$.*

This lemma implies that if $\nabla\psi \circ s$ and $\nabla\tilde{\psi} \circ \tilde{s}$ are two polar factorizations of u, then $\nabla\psi = \nabla\tilde{\psi}$ β-a.e. Thus $s = \nabla\phi \circ u = (\nabla\phi \circ \nabla\psi) \circ \tilde{s} = \tilde{s}$ μ-a.e. by lemma 6 and the non-degeneracy of $\tilde{s}$.

The next part of theorem 3 is that the map $u \mapsto s$ is continuous for non-degenerate u from $L^p(X, \mu)$ to $L^q(X, \mu)$ for all $q \in [1, \infty)$. Letting $\{u_n\}, u$ be non-degenerate in $L^p(X, \mu)$ such that $u_n \to u$ in L^p, by theorem 4, $\nabla\psi_n \to \nabla\psi$ in L^p. Thus, for any $f \in C_c(\mathbb{R}^d \times \Omega)$, $\int f(u_n(x), s_n(x))\, d\mu = \int f(\nabla\psi_n(z), z)\, d\beta \to \int f(\nabla\psi(z), z)\, d\beta = \int f(u(x), s(x))\, d\mu$.

Using the boundedness and uniform continuity of f, the finitude of $\mu(X)$, and a decomposition of X based on Chebyshev's inequality, it can be shown that $\int |f(u_n(x), s_n(x)) - f(u(x), s_n(x))|\, d\mu \to 0$ and in particular $\int f(u(x), s_n(x))\, d\mu \to \int f(u(x), s(x))\, d\mu$. This limit can be extended to $f(y, z)$ of the form $g(y) \cdot h(z)$ where $g \in L^1(\mathbb{R}^d, \alpha), h \in C(\overline{\Omega})$ by a density argument. Note that $f(y, z) := \nabla\phi(y) \cdot z$ is one such function as $\nabla\phi$ maps the probability space $(\mathbb{R}^d, \alpha)$ into Ω, i.e. is a bounded function on a finite measure set and is thus integrable.

The limit statement then becomes $\int \nabla\phi \circ u(x) \cdot s_n(x)\, d\mu \to \int \nabla\phi \circ u(x) \cdot s(x)\, d\mu$ or $\int s(x) \cdot s_n(x)\, d\mu \to \int s(x) \cdot s(x)\, d\mu$. Moreover, by the measure preserving property, $\int |s_n(x)|^2\, d\mu = \int |z|^2\, d\beta = \int |s(x)|^2\, d\mu$ for all n. The conclusion of these calculations is that $\int |s_n(x) - s(x)|^2\, d\mu \to 0$ so $s_n \to s$ in $L^2(X, \mu)$. However, because the 's_n's are essentially bounded uniformly by $\max_{\overline{\Omega}} |z| < \infty$, the sequence converges in L^q for $q \in (2, \infty)$ and because (X, μ) is a probability space, it also converges in L^q for $q \in [1, 2)$.

To see that s is a maximzer for the functional $\tilde{s} \mapsto \int u \cdot \tilde{s}\, d\mu$ among measure preserving maps from (X, μ) to (Ω, β), let s' be another such map. Note that the convexity inequality $\psi(s'(x)) \geq \psi(s(x)) + \nabla\psi(s(x)) \cdot (s'(x) - s(x))$ holds μ-a.e. by the non-degeneracy of s. Because $\nabla\psi \circ s = u$ μ-a.e., rearranging and integrating the inequality over (X, μ) yields $\int u(x) \cdot (s'(x) - s(x))\, d\mu \leq \int \psi \circ s'(x)\, d\mu - \int \psi \circ s(x)\, d\mu$. However, because s and s' are measure preserving, the two integrals are both equal to $\int \psi(z)\, d\beta$ so the right side of the inequality is 0, thus showing that s is indeed a maximizer.

To see that s is unique, let s' be a maximizer of the functional among measure preserving maps and let m' be the probability measure on $\mathbb{R}^d \times \overline{\Omega}$ defined for any f by $\int f(y, z)\, dm' = \int f(u(x), s'(x))\, d\mu$. It is straightforward to verify that m' has marginals α and β and solves the dual MKP. Thus $m' = \delta[y - \nabla\psi(z)]\, d\beta(z)$. Therefore $\int f(u(x), s'(x))\, d\mu = \int f(u(x), s(x))\, d\mu$ for all $f \in C_c(\mathbb{R}^d, \Omega)$ and, as before, this can be extended to $f(y, z) = \nabla\phi(y) \cdot z$ to show that $\int s(x) \cdot s'(x)\, d\mu = \int s(x) \cdot s(x)\, d\mu$. As $\|s'\|_{L^2(d\mu)} = \|s\|_{L^2(d\mu)}$, this shows that $s = s'$ in the $L^2(X, \mu)$ sense (i.e. μ-a.e.) so s is unique.

References

[1] Brenier, Y., *Polar Factorization and Monotone Rearrangement of Vector-Valued Functions.* Comm. Pure Appl. Math., 44:375???417, 1991.

Jordan Greenblatt, UCLA
email: **jsg66@math.ucla.edu**

7 An Elementary Introduction to Monotone Transportation

after K. Ball [3]
A summary written by Paata Ivanisvili

Abstract

We outline existence of the Brenier map. As an application we present simple proofs of the multiplicative form of the Brunn–Minkowski inequality and the Marton–Talagrand inequality.

7.1 Introduction

Given any two probability measures μ and ν on the Euclidian space $\mathbb{R}^n$ we say that a map $T : \mathbb{R}^n \to \mathbb{R}^n$ transports μ to ν if for each measurable set $A \subseteq \mathbb{R}^n$ we have

$$\mu(T^{-1}(A)) = \nu(A). \tag{1}$$

Condition (1) is equivalent to the following one: for any bounded continuous real-valued function f we have

$$\int_{\mathbb{R}^n} f(T(x))d\mu(x) = \int_{\mathbb{R}^n} f(x)d\nu(x). \tag{2}$$

It is worth mentioning that such map does not exist for arbitrary probability measures μ and ν. For example, if μ is one point mass and ν is supported on two different points then we can easily see that condition (1) can not be fulfilled.

Problem of mass transportation at first time was like this: for the given two probability measures μ and ν we need to minimize functional

$$\int \|x - Tx\| d\mu(x) \tag{3}$$

over all possible choices of T which transports μ to ν. The norm $\|\cdot\|$ represents usual Euclidian distance in $\mathbb{R}^n$.

One can easily see that such *optimal* map T which minimizes (3) is not unique in general. The problem of mass transportation itself is very difficult, for example, because of requirement (1). In order to avoid such strong requirement one can consider the following mass transportation problem: given two probability measures μ and ν we need to minimize the functional

$$\int \|x - y\| d\gamma(x, y) \tag{4}$$

over all possible choices of the measure γ on the product $\mathbb{R}^n \times \mathbb{R}^n$ such that for all measurable sets $A, B \subseteq \mathbb{R}^n$ we have

$$\gamma(A \times \mathbb{R}^n) = \mu(A), \tag{5}$$

$$\gamma(\mathbb{R}^n \times B) = \nu(B). \tag{6}$$

We should mention that in the case when both of the measures μ and ν are discrete, then the problem of minimizing (4) with conditions (5),(6) and the fact that both of the measures ν, μ are probability measures is nothing more than just a problem of linear programming. So the existence of measure γ in this particular case follows immediately.

From the point of view of linear programming it is quite natural to replace the integrand in (4) by some arbitrary real-valued function $c(x, y)$. In this general case we can treat the value $c(x, y)$ as a cost of moving the point x to y.

Henceforth, we will pay attention to the optimal transportation map T which transports μ to ν (see (1)) and minimizes

$$\int c(x, Tx) d\mu(x).$$

It is known that if c is a strictly convex function of the distance $\|x - y\|$ then the optimal transportation T is unique. In [2], Brenier explained that for $c(x, y) = \|x - y\|^2$ the optimal map T is a gradient of some convex function $\varphi : \mathbb{R}^n \to \mathbb{R}$ and vice versa, if ϕ is convex function and $\nabla\phi$ transports μ to ν then $T = \nabla\phi$ is optimal transportation. Such a map T will be called *Brenier map*. The property $T = \nabla\varphi$ allows us to use Brenier map for a wide range of applicatons (see subsections 7.3,7.4)

7.2 A construction of the Brenier Map

In the next theorem Brenier map will be constructed for some special measures.

Theorem 1. *If μ and ν are probability measures on $\mathbb{R}^n$, ν has compact support and μ assigns no mass to any set of Hausdorff dimension $(n-1)$ then there is a convex function $\varphi : \mathbb{R}^n \to \mathbb{R}$, so that $T = \nabla\varphi$ transports μ to ν.*

We sketch the proof of the theorem. First we consider the case when the measure ν is atomic i.e.

$$\nu = \sum_1^n \alpha_j \delta_{u_j}$$

For such a measure we find a convex function of the form $\varphi(x) = \max_j\{\langle x, u_j\rangle - s_j\}$ with some apropriate nambers s_j, such that it satisfyes the required property. In general, we approximate measure ν weakly by atomic measures v_k. It turns out that we can choose coresponding convex functions φ_k so that they converge locally uniformly to some convex function φ, moreover

$$\nabla\varphi_k \to \nabla\varphi$$

except for some set. Finally, by standart weak limit arguments we can see that the map $\nabla\varphi$ transports the measure μ to ν.

Having this theorem, it is worth mentioning the following relation between the measures μ and ν. Since $T = \nabla\varphi$, therefore, derivative of T i.e. Hessian of φ is positive semi-definite symmetric map. This means that T is essentialy 1-1. So, if μ and ν have densities f and g rspectively, then one can easily see that condition (1) turns into the following one

$$f(x) = g(Tx)\det(T'(x)). \tag{7}$$

This relation will be useful for our applications.

7.3 The Brunn–Minkowski Inequality

Classical Brunn-Minkovski inequality estimates the volume of the convex sum of nonempty sets in the Euclidian space from below. Namely, let A and B be non-measurable subset of $\mathbb{R}^n$. For $\lambda \in (0,1)$ we define

$$(1-\lambda)A + \lambda B = \{(1-\lambda)a + \lambda b : a \in A, b \in B\}.$$

Then

$$|(1-\lambda)A+\lambda B|^{1/n} \geq (1-\lambda)|A|^{1/n}+\lambda|B|^{1/n} \tag{8}$$

where $|A|$ denotes the n dimensional Lebesgue measure (volume) of the set A.

The first applications of Brenier map in proving Brunn–Minkowski inequality was found by McCann [4]. Barthe [1] used the Brenier map and gave a very clear proof of the Brascamp-Lieb inequality.

We restrict ourselves to a weak version of inequality (8), the so called multiplicative form of Brunn–Minkowski inequality, namely

$$|(1-\lambda)A+\lambda B| \geq |A|^{1-\lambda}|B|^{\lambda}. \tag{9}$$

The idea of using Brenier map in proving inequality (9) is the following: we consider the Brenier map T for the probability measures $\chi_A/|A|$ and $\chi_B/|B|$. Then the image of the map $T_\lambda = (1-\lambda)x+\lambda T(x)$ lies in $(1-\lambda)A+\lambda B$. So, using (7), one can see that inequality (9) follows from the estimate

$$\det((1-\lambda)I+\lambda T'(x)) \geq (\det T'(x))^{\lambda}$$

which is true for every positive semi-definite symmetric matrix $T'(x)$.

7.4 The Marton–Talagrand Inequality

In this subsection we present Marton–Talagrand inequality which was firstly observed by Marton [5]. The idea of proving this inequality is based on the existence of Brenier map.

Let γ be the standard Gaussian measure on $\mathbb{R}^n$ with density

$$g(x) = \frac{1}{(2\pi)^{n/2}}e^{-|x|^2/2}.$$

For a density f on $\mathbb{R}^n$ we define the relative entropy of f to be

$$\mathrm{Ent}(f||\gamma) = \int_{\mathbb{R}^n} f\log(f/g)dx.$$

The cost of transporting measure γ to the measure with density f is defined as

$$C(g,f) = \int |x-T(x)|^2 d\gamma,$$

T is the Brenier map transporting γ to the measure with density f.

Theorem 2. *With the notation above*

$$\frac{1}{2}C(g,f) \le \operatorname{Ent}(f||\gamma).$$

One of the important corollaries of the Marton–Talagrand inequality are probabilistic deviation inequalities. Consider measurable set $A \subset \mathbb{R}^n$. Let A_ε be a ε neighborhood of A. Set $B = \mathbb{R}^n \setminus A_\varepsilon$.

Then we have

$$\gamma(B) \le e^{-\gamma(A)\varepsilon^2}.$$

Indeed, take $f = \chi_B g(x)/\gamma(B)$. Then the relative entropy of f will be $-\log\gamma(B)$. By Marton–Talagrand inequality we have $C(g,f) \le -2\log\gamma(B)$. However,

$$C(g,f) = \int_{\mathbb{R}^n} \|x - T(x)\|^2 d\gamma \ge \int_A \|x - T(x)\|^2 d\gamma \ge \varepsilon^2\gamma(A)$$

So we obtain the desired result.

References

[1] Barthe, F., *Inégalités de Brascamp–Lieb et convexité.* C.R. Acad. Sci. Paris, 324, 885-888 (1997).

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

[3] K. Ball, *An Elementary Introduction to Monotone Transportation.* In: Geometric Aspects of Functional Analysis. Springer, 2004, pp. 41-52.

[4] McCann, R. J., *A convexity principle for interacting gases* Adv. Math., 228, 153-179 (1997)

[5] Marton, K., *Bounding $\bar{d}$-distance by informational divergence: A method to prove measure concentration.* The Annals of Probability, 24, 857-866 (1996)

Paata Ivanisvili, MSU
email: **ivanishvili.paata@gmail.com**

8 A Priori Estimates and the Geometry of the Monge-Ampère Equation

after L. Caffarelli
A summary written by Sajjad Lakzian

Abstract

We will very briefly touch on Caffarelli's regularity theory for fully nonlinear PDE and the geometry and regularity of the Monge-Ampère equation.

8.1 Introduction

The well known regularity results for small perturbation of linear equations are as follows:

(I) [Cordes-Nienberg Type Estimates]. Let $0 < \alpha < 1$ and u a bounded solution on B_1 of $Lu = a_{ij}D_{ij}u = f$, $|a_{ij} - \delta_{ij}| \leq \delta_0(\alpha)$ small enough, and f bounded; then,

$$||u||_{C^{1,\alpha}(B_{1/2})} \leq C\left(||u||_{L^\infty(B_1)} + ||f||_{L^\infty}\right). \tag{1}$$

(II) [Calderon-Zygmund]. if $f \in L^p$ for some $1 < p < \infty$ and $\delta_0(p)$ small enough, then

$$||u||_{W^{2,p}(B_{1/2})} \leq C\left(||u||_{L^\infty(B_1)} + ||f||_{L^p}\right). \tag{2}$$

(III) [Schauder]. If a_{ij} and f are of class C^α then,

$$||u||_{C^{2,\alpha}(B_{1/2})} \leq C\left(||u||_{L^\infty(B_1)} + ||f||_{C^\alpha}\right). \tag{3}$$

Caffarelli [1] has generalized these type of estimates to Fully nonlinear uniformly elliptic PDEs.

8.2 Fully Nonlinear Uniformly Elliptic PDE

The equation is of the form

$$F(D^2u, x) = f(x) \tag{4}$$

Uniform ellipticity in D^2 of equation 4 means that there exist λ and Λ such that for any matrix $N \in M_{n\times n}$ and any $N \in S^+$ we have

$$0 < \lambda||M|| < F(N+M, x) - F(N, x) \leq \Lambda||M|| \tag{5}$$

Definition 1 (Viscosity Solution)**.** *The continuous function u is called a C^2-viscosity solution of (4) if for any C^2-subsolution (resp. supersolution) ϕ, $u - \phi$ cannot have an interior minimum (resp. maximum).*

Let S denote the symmetric matrices then, $\beta(x)$, the oscillation of F in the variable x is given by:

$$\beta(x) = \sup_{M \in S} \frac{F(M,x) - F(M,0)}{||M||} \tag{6}$$

Caffarelli's results are as follows:

8.3 Caffrelli's Main Results

Theorem 2 ($W^{2,p}$ Regularity)**.** *Let u be a bounded viscosity solution of $F(D^2u, x) = f(x)$ in B_1 and assume that solutions ω of the Dirichlet problem*

$$f(x) = \begin{cases} F(D^2u, x) = 0 & in \ \ B_r \\ \omega = \omega_0 & in \ \ \partial B_r \end{cases}$$

satisfy he interior apriori estimate

$$||\omega||_{C^{1,1}(B_{r/2})} \leq Cr^{-2}||\omega||_{L^\infty(\partial B_r)} \tag{7}$$

Let $n < p < \infty$ and assume that $f \in L^p$ and for some $\theta = \theta(p)$ suficiently small

$$\sup_{B_1} \beta(x) \leq \theta(p) \tag{8}$$

Then $u|_{B_{1/2}}$ is in $W^{2,p}$ and

$$||u||_{W^{2,p}(B_{1/2})} \leq C\left(\sup_{\partial B_1}|u| + ||f||_{L^p}\right). \tag{9}$$

Theorem 3 ($C^{1,\alpha}$ Regularity)**.** *Assume that solutions ω to the equation*

$$F(D^2u, \omega) = 0 \tag{10}$$

in B_r satisfy the a priori estimate

$$||\omega||_{C^{1,\bar{\alpha}}(B_{r/2})} \leq Cr^{-(1+\bar{\alpha})}||\omega||_{L^\infty(B_r)} \tag{11}$$

Then for any $0 < \alpha < \bar{\alpha}$ there exists $\theta = \theta(\alpha)$ so that if

$$⨍_{B_r} \beta^n(x)dx \leq \theta \tag{12}$$

and

$$⨍_{B_r} |f(x)|^n dx \leq C_1 r^{(\alpha-1)n} \tag{13}$$

then any bounded solution u of

$$F(D^2u, x) = f(x) \tag{14}$$

in B_{r_0} is $C^{1,\alpha}$ at the origin. That is, there exist a linear function l such that for $r < r_0$

$$|u - l| \leq C_2 r^{1+\alpha} \tag{15}$$

and

$$||l||_{C^1} \leq C_3 \tag{16}$$

with

$$C_2, C_3 \leq C(\alpha) r_0^{-(1+\alpha)} \sup_{B_{r_0}} |u| + C_1^{1/n} \tag{17}$$

Theorem 4 ($C^{2,\alpha}$ Regularity)**.** *Assume the existence of $C^{2,\bar{\alpha}}$ interior a priori estimates for solutions of*

$$F(D^2\omega + M, 0) = 0 \tag{18}$$

for any M satisfying

$$F(M, 0) = F(0, 0) = 0. \tag{19}$$

Then if $0 < \alpha < \bar{\alpha}$,

$$⨍_{B_r} \beta^n dx \leq C r^{\alpha n}, \tag{20}$$

$$⨍_{B_r} |f(x)|^n dx \leq C r^{\alpha n} \tag{21}$$

and u is a solution of $F(D^2u, x) = f(x)$, then, u is $C^{2,\alpha}$ at the origin (in the same sense as above).

These regularity results are proven by exploring Alexander-Bakelman-Pucci maximum principle and Krylov-Safanov Harnack's Inequality.

8.4 Monge-Ampère Equation

We will discuss the solutions to the Monge-Ampère equation, $\det D_{ij}u = f$ and $0 < \lambda_1 \leq f \leq \lambda_2 < \infty$ on a convex set Ω. MA equation is perhaps the most famous example of non-uniformly elliptic PDE.

8.5 Geometric Properties, Alexandrov Solutions and Localization

Solutions to the MA quation are invariant under *affine transformations* with the proper renormalization; i.e. if u is a solution and $TX = AX + B$ is an *affine transformation*, then

$$w = \frac{1}{(\det T)^{2/n}} u(TX) \tag{22}$$

is also a solution. This means that one may produce new solutions by "stretching" the graph of u in some directions and "squeezing" it in other directions (in a way that keeps the Jacobian of ∇u fixed) and hence producing singular solutions.

This fact also tells us that the estimates on the solutions are inevitably dependent on the geometry of the domain of the definition.

Definition 5 (Generalized (Alexandrov) Solutions)**.** *Let ν be a Borel measure on Ω, an open and* ***convex*** *subset of $\mathbb{R}^n$. The convex function $u \in C(\Omega)$ is a generalized solution or Alexandrov solution to the MA equation*

$$\det D^2 u = \nu \tag{23}$$

if the MA measure Mu equals ν. Mu is defined as follows:

$$Mu(E) = |\partial u(E)| \tag{24}$$

Remark 6. *For the MA equation* $\det D^2 u = f$*, we take $\nu = f\mathcal{L}$ ($\mathcal{L}$: Lebesque measure.)*

Proposition 7. *if f is continuous then every Alexandrov solution u is also a viscosity solution.*

Lemma 8. *If $u, v \in C(\bar{\Omega})$, $u|_{\partial\Omega} = v_{\partial\Omega}$ and $v \geq u$ in Ω, then,*

$$\partial v(\Omega) \subset \partial u(\Omega) \tag{25}$$

Consequences of Lemma 8:

Theorem 9 (Alexandrov Maximum Principle). *$u : \Omega \to \mathbb{R}$ convex and $u|_{\partial u} = 0$ then,*

$$|u(x)|^n \leq C_n(diam\Omega)^{n-1}dist(x, \partial\Omega)|\partial\Omega| \ \ \forall x \in \Omega \tag{26}$$

Lemma 10 (Comparison Principle). *Let u, v convex functions on open bounded convex Ω and $u \geq v$ on $\partial\Omega$. If $\det D^2u \leq \det D^2v$ (in the MA measure sense) then,*

$$u \geq v \ \ in \ \ \Omega \tag{27}$$

One key tool in studying MA equation is John's Lemma:

Lemma 11 (John's Lemma). *For any open bounded convex set O, there exist an ellipsoid E such that*

$$E \subset S \subset nE \tag{28}$$

hence, there exist an invertible orientation preserving affine transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ such that $T(S)$ is normalized i.e. $B_1 \subset T(S) \subset B_n$

One immediate consequence is the following: Let u is a strictly convex solution of MA inequality on Ω then for any $x \in \Omega' \subset\subset \Omega$ and t sufficiently small; Then, if T normalizes the section $S(x,p,t)$, then the normalization, u^* of u given by:

$$u^*(y) := (\det T)^{2/n}\left(u(T^{-1}(y)) - u(x) - p.(T^{-1}(y) - x) - t\right) \tag{29}$$

solves the MA inequality on $T(S(x,p,t))$ with boundary condition $u^*\Big|_{\partial T(S(x,p,t))} = 0$.

Lemma 12. *Let Ω^* be a normalized open convex set i.e. $B_1 \subset \Omega^* \subset B_n$ and let u^* solve*

$$\lambda_1 \leq \det D^2u \leq \lambda_2 \quad u^*|_{\partial\Omega^*} = 0 \tag{30}$$

(we call u^* a normalized solution) then, there constants c_1, c_2 depending on λ_1, λ_2 such that

$$0 < c_1 \leq \left|\inf_{\Omega^*} u^*\right| \leq c_2 \tag{31}$$

Proof. Apply Lemma 10 to $\omega_1 = \lambda_1(|x|^2 - 1)/2$ and $\omega_2 = \lambda_2(|x|^2 - n^2)/2$ $\square$

Another important result is the localization Theorem which will be used in the proof of compactness theorem and strict convexity of solutions.

Theorem 13 (Localization [2])**.** *Let u be a solution of **MA inequality***

inside a convex set Ω, and let $l(x)$ is a supporting sloe to u. If the convex set

$$W = \{u(x) = l(x)\} \tag{32}$$

contains more than one point (hence not strictly convex) then it can not have an extremal point in Ω (i.e. this set has to exit the domain of the definition).

As a consequence of Lemmas 13, we get the compactness for normalized solutions. [1]

The proof of regularity results uses the properties of the sections of the solutions i.e. the sets

$$S(x,p,t) := \{y \in \Omega : u(y) \leq u(x)+p.(y-x)+t\} \;\; where, \;\; p \in \partial u(x) \;\; and \;\; t \geq 0. \tag{33}$$

the modulus of convexity is defined to be

$$\omega(x,u,t) := \sup_{p \in \partial u(x)} diam S(x,p,t) \tag{34}$$

and

$$\omega_{\Omega'} = \sup_{x \in \Omega'} \omega(x,u,t) \;\; \Omega' \subset\subset \Omega \tag{35}$$

8.6 $C^{1,\alpha}$ Regularity

It is enough to prove the $C^{1,\alpha}$ regularity of renormalized solutions u with $\inf_\Omega u = u(x_0)$ and $u|_{\partial\Omega} = 0$. Let $C_\beta \subset \mathbb{R}^{n-1}$ be the cone with vertex $(x_0, u(x_0))$ and base $\{u = (1-\beta)u(x_0)\}$. Suppose C_α is the graph of h_α.

Using the compactness result, one can find a universal δ for which

$$h_{1/2} \leq (1-\delta)h_1 \tag{36}$$

After renormalizing the level surface $\{u = 2^{-k}\}$ and iteration, we get:

$$h_{2^{-k}} \leq (1-\delta)^k h_1 \tag{37}$$

Since u is Lipschitz we have $h_1(x) \leq C|x-x_0|+u(x_0)$. Letting $2^{-\alpha'} = 1-\delta$ we get

$$h_{2^{-k}}(x) \leq C(2^{-k})^{\alpha}|x| \leq C2^{-k} \tag{38}$$

for $|x| \leq (2^{-k})^{(1-\alpha)}$

By the comparison Lemma 10, we have $u(x) \leq h_{2^{-k}}(x)$ as long as $h_{2^k}(x) \leq (1-2^{-k})u(x_0)$. This means that if for every x, we pick k that satisfies

$$(\frac{-C}{u(x_0)})2^{-(k+1)(1-\alpha')} \leq |x-x_0| \leq (\frac{-C}{u(x_0)})2^{-k(1-\alpha')} \tag{39}$$

then,

$$h_{2^k}(x) \leq (1-2^{-k})u(x_0) \tag{40}$$

And direct computation gives:

$$u(x) - u(x_0) \leq C|x-x_0|^{1+\alpha} \;\; where, \;\; \alpha = \frac{\alpha'}{1-\alpha'} \tag{41}$$

This shows that for all supporting planes l_{x_0}, we have:

$$\sup_{B(x_0,r)} |u(x) - l_{x_0}(x)| \leq Cr^{1+\alpha} \tag{42}$$

and this will imply that u is $C^{1,\alpha}$.

8.7 Sobolev Regularity

Theorem 14 (Caffarelli [1]). *Let u be a convex viscosity solution of the MA equation on a normalized convex set Ω and $u|_{\partial\Omega} = 0$ then,*

(I) *$\forall p < \infty, \exists \epsilon = \epsilon(p)$ s.t. if*

$$|f-1| \leq \epsilon \tag{43}$$

then,

$$u \in W^{2,p}\left(B_{1/2}\right) \tag{44}$$

and

$$||u||_{W^{2,p}\left(B_{1/2}\right)} \leq C(\epsilon) \tag{45}$$

(II) *If $f > 0$ and is continuous, then $u \in W^{2,p}\left(B_{1/2}\right)$ for any $p < \infty$ and*

$$||u||_{W^{2,p}\left(B_{1/2}\right)} \leq C(p,\sigma) \tag{46}$$

where σ is the modulus of continuity of f.

A consequence is the following theorem:

Theorem 15. $f \in C^{\alpha} \implies u \in C^{2,\alpha}$

Main Ideas of the Proof:

Lets consider a particular case: $1 \leq \det D_{ij}u \leq 1 + \epsilon(p)$ and we want to prove that $||u||_{W^{2,p}\left(B_{1/2}\right)} \leq C(p)$.

Step 1 Take the section $S_{\mu,L} = \{u - L \leq min(u - L) + \mu\}$ and normalize it by T_{μ}. Then approximate (using an approximation lemma as in [1]) the normalization of $u - L$ by solutions of $\det D_{ij}\omega = 1$. Notice that ω is $C^{2,\alpha}$

Step 2 Iterating this approximation at diadic levels $\mu = 2^{-k}$, one can show that

$$T_m u = D_{\mu}\tilde{T}_m u \tag{47}$$

where $D_m u = \left(\frac{1}{2\mu}\right)^{1/2} \mathrm{Id}$ is a dilation and and $\tilde{T}_m u$ is a transformation of norm

$$||\tilde{T}_{\mu}|| \ , \ ||\tilde{T}_m u^{-1}|| \leq \mu^{-\sigma} \tag{48}$$

with $\sigma = \sigma(\epsilon)$ is as small as we want.

So far we have a normalized solution u on $T_{\mu}(S_{\mu,L})$ with the following properties:

(a) $1 \leq \det D_{ij}u \leq 1+\epsilon$. **(b)** $\{u = 1\}$ is trapped between B_1 and B_n. **(c)** u is ϵ away from the $C^{2,\alpha}$ approximation function ω that solves $\det D_{ij}\omega = 1$ and $\{\omega = 0\} = \{u = 0\}$

Step 3

Lemma 16. *Let $\Gamma(u - \frac{1}{2}\omega)$ be the convex envelope of $u - \frac{1}{2}\omega$, then, the contact set $C = \{\Gamma(u - \frac{1}{2}\omega) = u - \frac{1}{2}\omega\}$ satisfies:*

$$\frac{|B_{1/2} \cap C|}{|B_{1/2}|} \geq 1 - C\epsilon^{1/2} \tag{49}$$

in other words, the contact points cover as large a portion of $B_{1/2}$ as we want.

Corollary 17. *At any contact point x_0, there exist a plane L_{x_0} such that in all of Ω*

$$L_{x_0}(x) \leq (u - \frac{1}{2}\omega)(x) \ \ and \ \ L_{x_0}(x_0) = (u - \frac{1}{2}\omega)(x_0) \tag{50}$$

which means that for any contact point x_0, u has a tangent paraboloid by below of the form

$$L_{x_0} + \frac{1}{N}|x - x_0|^2 \tag{51}$$

Remark 18. *If u has a tangent paraboloid by below $u \geq \frac{1}{\lambda}|x|^2$ then u has a tangent paraboloid by above $u \leq \lambda^{n-1}|x|^2$ because one can see that a paraboloid from below puts a uniform bound $||\tilde{T}_\mu|| \leq \lambda$ then since $\det \tilde{T}_m u = 1$, we also get a bound by below.*

Step 4 having controlled tangent paraboloids from above and below $\implies$ $W^{2,p}$ estimates.

References

[1] Caffarelli, L. A, *A priori estimates and the geometry of the Monge Ampère equation.* IAS/Park City Math. Ser. **2** (1996), no. 1, 5–63;

[2] Caffarelli, L. A, *A localization property of viscosity solutions to the Monge-Ampère equation and their strict convexity.* Ann. of Math. **131** (1990), no. 1, 129–134;

9 Partial differential equations and Monge-Kantorovich mass transport

after L.C. Evans [1]
A summary written by Tau Shean Lim

Abstract

We survey on Monge-Kantorovich mass transport with given cost function $c(x,y) = \frac{1}{2}|x-y|^2$ and $c(x,y) = |x-y|$. We also look at their applications on some PDE related topics.

9.1 Quick survey on Monge-Kantorovich problem

1. **Monge's optimal transport problem** - Given two probability measures $\mu^\pm$ on $\mathbb{R}^n$, define $\mathcal{A}$ be the class of functions

$$\mathcal{A} = \{\mathbf{s} : \mathbb{R}^n \to \mathbb{R}^n : \mathbf{s} \text{ is m-able, bijective, and } \mathbf{s}_\#(\mu^+) = \mu^-\}. \tag{1}$$

Define *cost density function* $c : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$. Corespondent to c, *total cost functional* of transfer plan $\mathbf{s}$ is given by by

$$I[\mathbf{s}] = \int_{\mathbb{R}^n} c(x, \mathbf{s}(x)) d\mu^+(x). \tag{2}$$

Optimal mass transfer problem is to determine and study the minimizer $\mathbf{s}^*$ of total cost functional I among class $\mathcal{A}$.

2. **Kantorovich reformulation** - Due to its high nonlinearity, Monge's problem remains difficulty. To ameliorate this difficulty, Kantorovich reformulates the problem by introducing its relaxation. Let $\mathcal{M}$ be the class of probability measures

$$\mathcal{M} = \{\mu\text{-Radon prob. meas. on } \mathbb{R}^n \times \mathbb{R}^n : \mathrm{proj}_x \mu = \mu^+, \mathrm{proj}_y \mu = \mu^-\}, \tag{3}$$

The new *relaxed cost functional* J on the space of Radon measure is given by

$$J[\mu] = \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x,y) d\mu(x,y) \tag{4}$$

The problem is to determine the minimizer μ^* of J among $\mathcal{M}$.

3. **Duality of Kantorovich problem** - One advantage of Kantovorich formulation is that both subject functional J and constrain $\mathcal{M}$ are linear. So it is a linear programming. More importantly, it provides a different way of viewing it - the *dual problem*. Let $\mathcal{L}$ be the collection of following functions

$$\mathcal{L} = \{(u, v) : u, v : \mathbb{C}^n \to \mathbb{R}^n, u(x) + v(y) \le c(x, y)\} \tag{5}$$

   The functional corresponds to dual problem is given by

$$K[u, v] = \int_{\mathbb{R}^n} u(x) d\mu^+(x) + \int_{\mathbb{R}^n} v(y) d\mu^-(y) \tag{6}$$

   The dual problem is to find the **maximizer** pair (u^*, v^*) which maximizes functional K. By duality principle in optimization theory, the optimality of primal and dual problem are equivalent, and

$$\min_{\mu \in \mathcal{M}} J[\mu] = \max_{(u,v) \in \mathcal{L}} K[u, v].$$

   If we impose certain condition to cost function c (convexity and coercivity), Kantorovich problem has a unique solution.

We will consider two special cases:

1. Uniform convex cost function $c(x, y) = \frac{1}{2}|x - y|^2$. In this space case, the optimal transport plan $\mathbf{s}^*$ exists, is unique, and $\mathbf{s}^* = \nabla \phi^*$ for some convex function ϕ^*.

2. Nonuniform convex cost function $c(x, y) = |x - y|$. The optimal transport plan exists, but no longer unique. The problem is harder due to the lack of uniform convexity.

9.2 Case for $c(x, y) = \frac{1}{2}|x - y|^2$

In this section, we assume $d\mu^\pm = f^\pm dx$ (or dy), where $f^\pm$ has compact support, and $\text{supp}(f^+) = X$, $\text{supp}(f^-) = Y$. We consider the dual problem of Kantovorich formulation. Let $(u, v) \in \mathcal{L}$ as in (5), that is,

$$u(x) + v(y) \le \frac{1}{2}|x - y|^2 \tag{7}$$

Corespondent to (u, v), we define the function pair (ϕ, ψ) by

$$\phi(x) = \frac{1}{2}|x|^2 - u(x), \quad \psi(y) = \frac{1}{2}|y|^2 - v(y) \tag{8}$$

Then (7) reads

$$\phi(x) + \psi(y) \geq x \cdot y, \quad (x, y) \in X \times Y \tag{9}$$

As it turns out, to minimize functional K over $\mathcal{L}$ is equivalent to **minimize** the following functional L over the constraint (9).

$$L[\phi, \psi] = \int_X \phi(x) f^+ dx + \int_Y \psi(y) f^- dy \tag{10}$$

Lemma 1. 1. *There exists minimizer* (ϕ^*, ψ^*) *of* L *subject to constrain* (9).

2. ϕ^* *and* ψ^* *are convex functions, and convex dual of each other.*

Sketch of Proof. For any pair of (ϕ, ψ) which satisfies (9), we may find another pair $(\hat{\phi}, \hat{\psi})$ that satisfies (9) and $\phi \leq \hat{\phi}$, $\psi \leq \hat{\psi}$. So $L[\phi, \psi] \leq L[\hat{\phi}, \hat{\psi}]$. Choose a maximizing sequence from dual pairs, which are uniform Lipschitz. This allows us to construct (ϕ^*, ψ^*) from uniform limit. □

To recover optimal transfer plan $\mathbf{s}^*$, we set

$$\mathbf{s}^* = \nabla \phi^*, \tag{11}$$

Since ϕ^* is convex according to lemma 1, so the derivative exists a.e.

Theorem 2. *Define* $\mathbf{s}^*$ *by* (11), *then*

1. $\mathbf{s}^* : X \mapsto Y$ *is essentially bijective.*

2. $\mathbf{s}^*_{\#}\mu^+ = \mu^-$, *hence* $\mathbf{s}^* \in \mathcal{A}$.

3. s^* *is the optimal solution of mass transfer problem.*

Sketch of Proof. Take $h : \mathbb{R}^n \to \mathbb{R}$ a continous function. To show 2, by replacing h into $-h$, it is enough to prove

$$\int_X h(\mathbf{s}^*(x)) f^+(x) dx \leq \int_Y h(y) f^-(y) dy \tag{12}$$

For $\tau \geq 0$, by taking $\psi_\tau(y) = \psi^*(y) + \tau h(y)$, $\phi_\tau(x)$ be Legendre transform of ψ_τ, (ϕ_τ, ψ_τ) satisfies (9). Then $i(\tau) = L[\phi_\tau, \psi_\tau] \geq L[\phi^*, \psi^*]$. $i(\tau)$ achieve minimum at $\tau = 0$. Hence,

$$\begin{aligned} 0 &\leq \frac{1}{\tau}(L[\phi_\tau, \psi_\tau] - L[\phi^*, \psi^*]) \\ &= \int_X \frac{\phi_\tau(x) - \phi^*(x)}{\tau} f^+(x)dx + \int_Y h(y) f^-(y)dy \end{aligned} \tag{13}$$

For each $x \in X$, let $y_\tau = y_{\tau,x} \in Y$ such that $\phi_\tau(x) = x \cdot y_\tau - \psi_\tau(y_\tau)$. Then

$$\phi_\tau(x) - \phi^*(x) = x \cdot y_\tau - \psi(y_\tau) - \tau h(y_\tau) - \phi^*(x) \leq -\tau h(y_\tau) \tag{14}$$

Lastly, let $\tau \to 0$, $y_\tau \to \nabla \phi^*(x) = \mathbf{s}^*(x)$ a.e. Monotone convergence, (13) and (14) gives (12). □

9.2.1 Application: Nonlinear diffusion equation

For $f^+, f^- \in \mathcal{M}$, define Wasserstein distance

$$d^2(f^+, f^-) = \inf \left\{ \frac{1}{2} \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 d\mu(x, y) \right\} \tag{15}$$

where the infimum is taken over all measure μ that $d\text{proj}_x \mu = f^+ dx$ and similar for projection on y. If μ achieve the minimality, then μ is the optimal solution of Kantorovich problem, and we know it is unique.

Now we define a sequence of $u_k \in \mathcal{M}$ (space of Radon probability measure with density) in following way. Choose $h > 0$. Let $u^0 = g \in \mathcal{M}$, and inductively define u^{k+1} by the minimizer following functional over $\mathcal{M}$:

$$N_k(v) = \frac{d^2(v, u^k)}{h} + \int_{\mathbb{R}^n} \beta(v) dx \tag{16}$$

where β is some convex, coercive real function. Also suppose

$$\int \beta(v) dx < \infty \tag{17}$$

Standard result from calculus of variation guarantees the existence and uniqueness of minimizer of functional N_k. So $\{u^k\}$ is well-defined.

Finally, we define u_h by "linearly gluing $\{u_k\}$ in step size h. More precisely, $u_h(kh, \cdot) = u^k(\cdot)$ for each $k \geq 0$, and $u_h(t, \cdot)$ be linear if $hk < t < h(k+1)$. The interesting question is what happen as $h \to 0$.

Theorem 3. *If $u_h \to u$ in $L^1_{LOC}(\mathbb{R}^n \times \mathbb{R}^+)$ as $h \to 0$, then u is weak solution of nonlinear diffusion equation:*

$$\begin{cases} u_t = \triangle\alpha(u) & (x,t) \in \mathbb{R}^n \times \mathbb{R}^+ \\ u = g & (x,t) \in \mathbb{R}^n \times \{t=0\} \end{cases} \tag{18}$$

where $\alpha(v) = \beta'(v)v - \beta(v)$

A interesting corollary is $\beta(v) = v\log v$. Then $\alpha(v) = v$, which in this case we obtain the solution of linear diffusion equation.

9.3 Case for $c(x,y) = |x-y|$

The case of nonuniformly convex function $c(x,y) = |x-y|$ is harder than the previous case. First of all, we again assume $d\mu^{\pm} = f^{\pm}dx$. Monge find that if $\mathbf{s}^*$ is the optimal transport plan, then there exists a function u^* such that

$$\nabla u^*(x) = -\frac{x - \mathbf{s}^*(x)}{|x - \mathbf{s}^*(x)|} \tag{19}$$

In another word, the direction of where the mass moving from x is gradient of some potential function.

To tackle the existence problem, we consider Kantorovich problem with $c(x,y) = |x-y|$. The proof of existence is similar as the previous case.

Lemma 4. *There exists (u^*, v^*) solves the dual problem. Moreover, we may take (u^*, v^*) such that $v^* = -u^*$, and*

$$`|u(x) - u(y)| \leq |x-y| \quad (x,y) \in \mathbb{R}^n \times \mathbb{R}^n \tag{20}$$

By setting $v^* = -u^*$, the optimization problem reduces to finding maximizer of

$$K[u] = \int_X u(f^+ - f^-)dx = \langle u, f^+ - f^- \rangle_{L^2} \tag{21}$$

over u in certain function space (L^2) such that $|\nabla u| \leq 1$ a.e. In the terminology of convex analysis, let $\mathbb{K} = \{v \in L^2(X) : |\nabla v| \leq 1 \text{ a.e}\}$, and $I_{\mathbb{K}}$ be the characteristic function of $\mathbb{K}$. That is,

$$I_{\mathbb{K}}[v] = \begin{cases} 0 & v \in \mathbb{K} \\ \infty & v \notin \mathbb{K} \end{cases} \tag{22}$$

Then u^* maximizes K iff $f^+ - f_- \in \partial I_{\mathbb{K}}[u^*]$.

The main difficulty is to construct optimal transfer map $\mathbf{s}$ out of u^*. There are several approach to deal with it, and we will follow a differential-equation-based method.

Theorem 5. *There exists* $\mathbf{s}^*$ *solves optimal transport problem.*

Roughly speaking, we first construct $a \in L^\infty$ that satisfies

$$-\mathrm{div}(a(x)\nabla u^*) = f^+ - f^- \tag{23}$$

a is constructed as the limit of solution of p-Laplacian equation. Then construct flow $z(t,x)$ (with $z(0,x) = x$) by ODE related to a, u^* and $f^\pm$. The optimal map $\mathbf{s}^*(x) = z(1,x)$

9.3.1 Application: Sandpile model

Monge transport problem has an application in modeling the evolution of sandpile. Let $f : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}$ be the source of the sand, where $f(x,t)$ record the rate of sand added at any position $x \in \mathbb{R}^n$ and any time $t \in [0,\infty)$. $u : [0,\infty) \times \mathbb{R}^n$ is the height of sand at given position and time. The equation that model the phenomenon is given by

$$f - u_t \in \partial I_{\mathbb{K}}[u] \tag{24}$$

where $I_{\mathbb{K}}[u]$ is defined as in (22). To interpret this, u indeed solves Kantorovich problem for $d\mu^+ = f dx$ and $d\mu^- = u_t dy$. In another word, the sandpile instantly rearrange itself with potential function u (in the way of Monge-Kantorovich transport) from $f dx$ into $u_t dy$. The constrain $I_{\mathbb{K}}$ gives that $|\nabla u| \leq 1$ a.e, which comes from the fact that the sandpile cannot stay in equilibrium if the slope at a point is larger by $\pi/4$.

References

[1] Evans, L. C, *Partial differential equations and Monge-Kantorovich mass transfer,* Current developments in mathematics, pp. 65–126;

Tau Shean Lim, UW-Madison
email: **taushean@math.wisc.edu**

10 Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality

after F.Otto and C. Villani [1]
A summary written by Javier Morales

Abstract

This paper discusses the relationship between the Wasserstein distance and the relative entropy.

10.1 Introduction

Let M be a smooth complete Riemannian manifold of dimension n, with geodesic distance

$$d(x,y) = \inf\left\{\sqrt{\int_0^1 |\dot{w}(t)|^2 dt}, w \in C^1((0,1);M), w(0) = x, w(1) = y\right\}$$

We define the Wasserstein distance, or transportation distance with quadratic cost, between two probability measure μ and ν on M, by

$$W(\mu,\nu) = \sqrt{\inf_{\pi \in \Pi(\mu,\nu)} \int_{M \times M} d(x,y)^2 d\pi(x,y)}$$

where $\Pi(\mu,\nu)$ denotes the set of probability measures on $M \times M$ with marginals μ and ν.

The Wasserstein distance metrizes the weak-* topology on $P_2(M)$, the set of probability measures on M with finite second moments, and is thus a natural way to measure distance between probability measures in a weak sense.

Let μ and ν be probability measures on M absolutely continuous w.r.t the standard volume measure dx. We define their relative entropy by

$$H(\mu \mid \nu) = \int_M \log \frac{d\mu}{d\nu} d\mu$$

We define their relative Fisher information by

$$I(\mu,\nu) = \int_M \left|\nabla \log \frac{d\mu}{d\nu}\right|^2 d\mu$$

The relative entropy has been classically used as a measure of discrimination between probability measures, it is not a metric since it is not symmetric nor it satisfies the triangular inequality but it is positive definite.

This paper studies the relationship between these two ways of distinguishing probability measures

A previous result in this direction is an inequality by Talagrand.
Let

$$d\gamma(x) = \frac{e^{-\frac{|x|^2}{2}}}{(2\pi)^{\frac{n}{2}}} dx$$

then for any probability measure μ on $\mathbb{R}^n$ absolutely continuous w.r.t dx

$$W(\mu,\gamma) \leqslant \sqrt{2H(\mu \mid \gamma)}$$

The authors generalize this inequality in a very wide class of probability measures: namely, all the probability measures ν (on a Riemannian manifold M) satisfying a logarithmic Sobolev inequality involving the relative Fisher information.

We will say that the probability measure ν satisfies a logarithmic Sobolev inequality with constant $\rho > 0$ (in short: LSI(ρ)) if for any probability measure μ absolutely continuous w.r.t ν,

$$H(\mu \mid \nu) \leq \frac{1}{2\rho} I(\mu,\nu)$$

We will say that the probability measure ν satisfies a Talagrand inequality with constant $\rho > 0$ (in short: T(ρ)) if for any probability measure μ absolutely continuous w.r.t ν,with finite moments of order 2

$$W(\mu,\nu) \leq \sqrt{\frac{2H(\mu,\nu)}{\rho}}$$

10.2 Main Results

The first theorem provides a bound for the relative entropy in terms of the Wasserstein distance

Theorem 1. *Let* $d\nu = e^{-\Psi}dx$ *be a probability measure with finite moments of order 2, such that* $\Psi \in C^2(M)$ *and* $D^2\Psi + Ric \geq -CI_n, C \in \mathbb{R}$*. If* ν *satisfies LSI(*ρ*) for some* $\rho > 0$ *then it also satisfies T(*ρ*)*

In the second theorem the authors recall a simple criterion that guarantees the hypothesis of the previous theorem. This is a result of Bakry and Emery

Theorem 2. *Let* $d\nu = e^{-\Psi}dx$ *be a probability measure on* M*, such that* $\Psi \in C^2(M)$ *and* $D^2\Psi + Ric \geq \rho I_n, \rho > 0$*. Then* ν *satisfies LSI(*ρ*).*

Finally, the third theorem provides a bound for the relative entropy in terms of the Wasserstein distance under suitable hypothesis

Theorem 3. *Let* $d\nu = e^{-\Psi}dx$ *be a probability measure on* M*, with finite moments of order 2 such that* $\Psi \in C^2(M)$ *,* $D^2\Psi + Ric \geq KI_n$*,* $K \in \mathbb{R}$ *(not necessarily positive). Then, for any probability measure* μ *on* M*, absolutely continuous w.r.t* ν*, we have*

$$H(\mu \mid \nu) \leq W(\mu,\nu)\sqrt{I(\mu \mid \nu)} - \frac{K}{2}W(\mu,\nu)^2$$

The proofs of these theorems are mainly based on partial differential equations. This last theorem is stated for $\mathbb{R}^n$ in the paper but the authors show, on a later work, that the condition holds as above by using the Hamilton-Jacobi equation on M.

10.3 Applications

Theorem 1 can be used to show that probability measures on Riemannian manifolds satisfying LSI(ρ) are concentrated.

Let $B \subset M$ be non empty and measurable . For any $t > 0$ let

$$B_t = \{x \in M; d(x, B) \geq t\}$$

Corollary 4. *Under the assumptions of Theorem1, for all measurable set* $B \subset M$*, and* $t \geq \sqrt{\frac{2}{\rho} \log \frac{1}{\nu(B)}}$ *we have*

$$\nu(B_t) \geq 1 - e^{-\frac{\rho}{2}\left(t - \sqrt{\frac{2}{\rho} \log \frac{1}{\nu(B)}}\right)^2}$$

From Theorem 1 and 2 we obtain

Corollary 5. *Let* $d\nu = e^{-\Psi} dx$ *be a probability measure on* M *with finite second moments of order 2, such that* $\Psi \in C^2(M)$ *and* $D^2\Psi + Ric \geq \rho I_n$*,*$\rho > 0$*. Then*

$$W(\mu, \nu) \leq \sqrt{\frac{2H(\mu, \nu)}{\rho}}$$

For all μ *absolutely continuous w.r.t* ν *and with finite moments of order 2*

Let ν denote the uniform measure on a Riemannian manifold M, with unit volume: $\nu(M) = 1$. Let us assume that ν satisfies the hypothesis of Theorem1 for some $\rho > 0$, and let A be any measurable subset of M, and $f = \frac{1_A}{\nu A}$, then we have

$$W(fd\nu, d\nu) \leq \sqrt{\frac{2\int f \log f d\nu}{\rho}} = \sqrt{\frac{2}{\rho} \log \frac{1}{\nu(A)}}$$

The authors use this inequality to give a simplified proof of a theorem by Ledoux, which establishes a partial converse to the statement (due to Rothaus) that compact manifolds always satisfy logarithmic Sobolev inequalities, the proof they give also has the advantage that it leads to much simpler numerical constants

Theorem 6. *(Ledoux). Let M be a smooth complete Riemannian manifold of dimension n, with uniform measure ν, $\nu(M) = 1$. Assume that ν satisfies LSI(ρ) some $\rho > 0$ and that $Ric \geq -RI_n$, $R > 0$. Then M has finite diameter D, with*

$$D \leq C\,\sqrt{n}\,max\left(\frac{1}{\sqrt{\rho}}, \frac{R}{\rho}\right),$$

where C is numerical

References

[1] Otto, F and Villani, C., *Generalization of an Inequality by Talagrand and Links with the Logarithmic Sobolev Inequality* Journal of Functional Analysis, **173** (2000), pp 361-400

Javier Morales, University of Texas at Austin
email: **jmorales@math.utexas.edu**

11 Shannon's monotonicity problem for free and classical entropy

after D. Shlyakhtenko and H. Schultz [4]
A summary written by Brent Nelson

Abstract

We give a unified proof, valid in both classical probability theory and Voiculescu's free probability theory, of the monotonicity of entropy (resp., free entropy).

11.1 Introduction

For a classical, real valued random variable X with distribution p, its entropy as defined as

$$H(X) = -\int_{\mathbb{R}} p(x) \log p(x)\ dx.$$

Amongst random variables with $E(X) = 0$ and $E(X^2) = 1$, entropy is maximized by the standard Gaussian random variable G with variance 1. Furthermore, if $X_1, X_2, \ldots$ are independent identically distributed random variables with $E(X_j) = 0$ and $E(X_j^2) = 1$ then the classical central limit theorem states that their central limit sums

$$Z_N = \frac{X_1 + \cdots + X_N}{\sqrt{N}},$$

converge in law to G. Moreover, the entropy of this sequence is monotone nondecreasing; a result due to Artstein, Ball, Barthe, and Naor [1].

Free probability theory (i.e. non-commutative probability theory) offers a complete parallel with the above classical results. Given a non-commutative random variable $X \in (A, \tau)$ with law $d\mu_X$, its free entropy is defined as

$$\chi(X) = \iint_{\mathbb{R}^2} \log|s-t|\ d\mu_X(s) d\mu_X(t) + \frac{3}{4} + \frac{1}{2}\log 2\pi.$$

Amongst non-commutative random variables with $\tau(X) = 0$ and $\tau(X^2) = 1$, free entropy is maximized by the random variable S with the semicircle law $d\mu_S(x) = \frac{1}{2\pi}\sqrt{4-t^2}\ dx$. Furthermore, if $X_1, X_2, ...$ are freely independent identically distributed random variables with $\tau(X_j) = 0$ and $\tau(X_j^2) = 1$ then Voiculescu's free central limit theorem states that their central limit sums (defined as in the classical case) converge in law to S.

Shlyakhtenko and Schultz in [4] showed that the entropy of this sequence is also monotone nondecreasing with a proof that is also valid in the classical case. In fact a more general result was proven replacing single random variables with p-tuples of random variables: let $X_j = (X_j^{(1)}, \dots, X_j^{(p)})$, $j = 1, 2, \dots$ be freely independent p-tuples of non-commutative random variables and let $Z_N = N^{-1/2}(X_1 + \cdots X_N)$ be their central limit sum. Then the free entropy of Z_N is a monotone function of N.

11.2 Preliminaries

11.2.1 Free probability

Let $\mathcal{H}$ be a Hilbert space and denote by $\mathcal{B}(\mathcal{H})$ the set of bounded operators on $\mathcal{H}$. Let $A \subset \mathcal{B}(\mathcal{H})$ be a $*$-subalgebra such that $1 \in A$. Let $\tau\colon A \to \mathbb{C}$ be a linear functional such that

1. $\tau(a^*a) \geq 0$ for all $a \in A$ (i.e. τ is *positive*);
2. $\tau(ab) = \tau(ba)$ for all $a, b \in A$ (i.e. τ is *tracial*); and
3. $\tau(1) = 1$.

Then (A, τ) is a *non-commutative probability space*. A *non-commutative random variable* is any element $a \in A$. In analogy with classical probability theory, $\tau(a)$ is the expectation of a and $\tau(a^2) - \tau(a)^2$ is the variance. We say that a is *centered* if $\tau(a) = 0$. The *law* of a non-commutative random variable a refers to the collection of its moments: $\{\tau(a^n)\colon n \in \mathbb{N}\}$. It can also be thought of as a linear functional on polynomials $\mu_a\colon \mathbb{C}[t] \to \mathbb{C}$ such that $\mu_A(t^n) = \tau(a^n)$ for monomial $t^n \in \mathbb{C}[t]$. In fact, if $a = a^*$ is self-adjoint, then it turns out there exists measure μ on $\mathbb{R}$ such that

$$\mu_a(t^n) = \int_{\mathbb{R}} t^n \, d\mu(t).$$

For example, recall that $d\mu(t) = \frac{1}{2\pi}\sqrt{4 - t^2}\, dt$ defines the semicircle distribution of radius 2. We say that a self-adjoint non-commutative random variable a has the semicircle law, or is *semicircular*, if

$$\tau(a^n) = \int_{\mathbb{R}} t^n \frac{1}{2\pi}\sqrt{4-t^2}\, dt = \begin{cases} C_k & n = 2k; \\ 0 & n = 2k+1, \end{cases}$$

where $\{C_k\}$ are the Catalan numbers

$$C_k = \frac{1}{k+1}\binom{2k}{k}.$$

Given several random variables $X_1, \ldots, X_n$ one can also consider their *joint law*, which can be thought of as a linear functional on non-commutative polynomials $\mu\colon \mathbb{C}\langle t_1, \ldots, t_n\rangle \to \mathbb{C}$ such that

$$\mu(p(t_1, \ldots, t_n)) = \tau(p(X_1, \ldots, X_n))$$

for $p \in \mathbb{C}\langle t_1, \ldots, t_n\rangle$. In this situation, one no longer has a single moment of each degree. For example, the second moments of $X_1, \ldots, X_n$ would be the values $\tau(X_i X_j)$ for $i, j \in \{1, \ldots, n\}$.

11.2.2 Free independence

The non-commutative replacement for the classical notation of independence is *free independence*. Let $F_1, F_2 \subset A$ be two families of non-commutative random variables. Then we say F_1 and F_2 are *freely independent* if

$$\tau(a_1 \cdots a_n) = 0$$

whenever $a_j \in \mathbf{Alg}(1, F_{i(j)})$, where $i(1) \neq i(2)$, $i(2) \neq i(3)$, etc., and $\tau(a_1) = \cdots = \tau(a_n) = 0$.

Given $X_1, X_2 \in A$ which are freely independent, one is able to compute moments of their joint law in terms of their individual laws by *centering* each varible: $\tilde{X}_j := X_j - \tau(X_j)$. Thus $\tau(\tilde{X}_j) = 0$ and satisfies the hypothesis in the above definition. So computing $\tau(X_1X_2)$ proceeds as follows:

$$\begin{aligned}
\tau(X_1X_2) &= \tau((\tilde{X}_1 + \tau(X_1))(\tilde{X}_2 + \tau(X_2))) \\
&= \tau(\tilde{X}_1\tilde{X}_2) + \tau(\tilde{X}_1)\tau(X_2) + \tau(X_1)\tau(\tilde{X}_2) + \tau(X_1)\tau(X_2) \\
&= \tau(X_1)\tau(X_2),
\end{aligned}$$

where the first term vanishes due to free independence. While this seems to agree with the classical notion of independence wherein $E(X_1X_2) = E(X_1)E(X_2)$, this is not the case for longer products: $\tau(X_1X_2X_1X_2) = \tau(X_1^2)\tau(X_2)^2 + \tau(X_1)^2\tau(X_2^2) - \tau(X_1)^2\tau(X_2)^2$. Regardless, this is precisely the notion needed in the free central limit theorem.

11.2.3 $L^2(A,\tau)$ and Hilbert space tensor products

Using a standard construction in Operator Algebras (the Gelfand-Naimark-Segal construction), we can produce a Hilbert space from (A,τ). For $a,b \in A$ consider

$$\langle a,b\rangle_{L^2(A,\tau)} := \tau(a^*b).$$

As A is a complex vector space, this defines a sesquilinear form that is complex linear in the second entry. It turns out that $N = \{a \in A\colon \tau(a^*a) = 0\}$ forms a vector subspace and hence we can consider the vector space A/N, for which $\langle\cdot,\cdot\rangle_{L^2(A,\tau)}$ defines an inner product. Thus $\|\bar{a}\|_{L^2(A,\tau)} := \tau(a^*a)^{\frac{1}{2}}$, $\bar{a} \in A/N$ defines a norm. The completion of A/N with respect to this norm is a Hilbert space which we denote $L^2(A,\tau)$. Moreover, every element $a \in A$ defines an element of $\mathcal{B}(L^2(A,\tau))$: $a \cdot \bar{b} = \overline{ab}$ for $\bar{b} \in A/N$. Since

$$\|\overline{ab}\|^2_{L^2(A,\tau)} = \tau(b^*a^*ab) \leq \|a\|^2\tau(b^*b) = \|a\|^2\|\bar{b}\|_{L^2(A,\tau)}$$

this can be extended to a bounded operator on all of $L^2(A,\tau)$.

11.2.4 Fock space example

Let $\mathcal{H}$ be a Hilbert space, and for $k \in \mathbb{N}$ let $\mathcal{H}^{\otimes k}$ denote the k-fold Hilbert space tensor product of $\mathcal{H}$. Then the *full Fock space* is defined as

$$\mathcal{F}(\mathcal{H}) = \mathbb{C}\Omega \oplus \bigoplus_{k=1}^{\infty} \mathcal{H}^{\otimes k}.$$

It is linearly spanned by Ω (the vacuum vector) and elements of the form $\xi_1 \otimes \cdots \otimes \xi_k \in \mathcal{H}^{\otimes k}$, and has an inner product defined by

$$\langle \xi_1 \otimes \cdots \otimes \xi_k, \eta_1 \otimes \cdots \otimes \eta_l \rangle = \delta_{k=l} \prod_{j=1}^{k} \langle \xi_j, \eta_j \rangle .$$

Given $\xi \in \mathcal{H}$ we can define the *left creation operator* $l(\xi) \in \mathcal{B}(\mathcal{F}(\mathcal{H}))$ by

$$l(\xi)\Omega = \xi \qquad \text{and} \qquad l(\xi)\xi_1 \otimes \cdots \otimes \xi_k = \xi \otimes \xi_1 \otimes \cdots \otimes \xi_k.$$

Its adjoint (*the left annihilation operator*) is given by

$$l(\xi)^*\Omega = 0 \qquad \text{and} \qquad l(\xi)^*\xi_1 \otimes \cdots \otimes \xi_k = \langle \xi, \xi_1 \rangle \xi_2 \otimes \cdots \otimes \xi_k.$$

Denote $c(\xi) = l(\xi) + l(\xi)^*$. The map $a \mapsto \langle \Omega, a\Omega \rangle$ on $\mathcal{B}(\mathcal{F}(\mathcal{H}))$ defines a positive linear functional τ. It turns out that $c(\xi)$ has the semicircle law with respect to τ. Moreover, if $\eta \in \mathcal{H}$ is another vector orthogonal to to ξ then $c(\eta)$ and $c(\xi)$ are freely independent.

11.2.5 Conjugate variables, free Fisher information, and free entropy

Let $X_1, \ldots, X_n$ be some non-commutative random variables in an operator algebra A equipped with a positive tracial linear functional $\tau \colon A \to \mathbb{C}$. Let $L^2(A, \tau)$ be as above and consider the Hilbert space tensor product of it with itself: $L^2(A,\tau)\bar{\otimes}L^2(A,\tau)$. For each $j = 1, \ldots, n$ we define a derivation $\partial_j \colon \mathbf{Alg}(1, X_1, \ldots, X_n) \to L^2(A,\tau)\bar{\otimes}L^2(A,\tau)$ by $\partial_j X_k = \delta_{j=k} 1 \otimes 1$ and the Leibniz rule:

$$\partial_j(ab) = \partial_j(a) \cdot b + a \cdot \partial_j(b).$$

When $1 \otimes 1$ is in the domain of the adjoint map $\partial_j^* \colon L^2(A,\tau)\bar{\otimes}L^2(A,\tau) \to L^2(A,\tau)$ we define the *conjugate variables* by

$$\xi_j = J(X_j \colon X_1, \ldots, \hat{X}_j, \ldots, X_n) = \partial_j^*(1 \otimes 1), \qquad \text{for } j = 1, \ldots, n.$$

This implies for $\eta \in L^2(A,\tau)$ that $\langle \eta, \xi_j \rangle_{L^2(A,\tau)} = \langle \partial_j(\eta), 1 \otimes 1 \rangle_{L^2(A,\tau)\bar{\otimes}L^2(A,\tau)}$.

The free Fisher information is defined as

$$\Phi^*(X_1, \ldots, X_n) := \sum_j \|\xi_j\|^2_{L^2(A,\tau)},$$

and the free entropy of an n-tuple $(X_1, \ldots, X_n)$ is defined as

$$\chi^*(X_1, \ldots, X_n) = \frac{1}{2} \int_0^\infty \left[\frac{n}{1+t} - \Phi^*(X_1^t, \ldots, X_n^t) \right] dt + \frac{n}{2} \log 2\pi e, \qquad (1)$$

where $X_j^t = X_j + \sqrt{t} S_j$ and $S_1, \ldots, S_n$ are freely iid centered semicircular random variables of variance 1, freely independent from $X_1, \ldots, X_n$.

While (1) appears to be quite different from $\chi(X)$, these two quantities agree for single random variables: $\chi(X) = \chi^*(X)$.

With this definition in hand we are now prepared to state the main theorem.

11.3 Montonicity of free entropy

Theorem 1. *Let $X_j = (X_j^{(1)}, \dots, X_j^{(p)})$ be a sequence of p-tuples of random variables, so that $\{(X_j^{(1)}, \dots, X_j^{(p)}) \colon j = 1, 2, \dots\}$ are freely independent and are identically distributed and have finite second moments. Let $Z_N = N^{-1/2}(X_1 + \cdots + X_N)$. Then the function $N \mapsto \chi^*(Z_N^{(1)}, \dots, Z_N^{(p)})$ is monotone nondecreasing.*

The proof proceeds by first establishing that the Fisher information is monotone nonincreasing, which will be shown suffices given the formula (1). This is done by by exploiting the free independence of the random variables to express τ as a composition of orthogonal projections. This then allows us to use Lemma 5 from the classical proof [1], which provides the necessary bound.

Theorem 2. *Let $X_j = (X_j^{(1)}, \dots, X_j^{(p)})$ be a sequence of p-tuples of random variables, so that $\{X_j \colon j = 1, 2 \dots\}$ are freely independent and identically distributed and have finite second moments. Define $Z_N = (X_1 + \cdots + X_N)/\sqrt{N}$. Then the function $N \to \Phi^*(Z_N^{(1)}, \dots, Z_N^{(p)})$ is monotone nonincreasing.*

Proof of Theorem 1. Let X_j and Z_N be as in the statement of Theorem 1. Let $\{S_j^{(k)}\}_{j,k}$ be freely iid centered semicircular variables of variance 1, which are freely independent from $\{X_j^{(k)}\}_{j,k}$. Define $X_j^{(k,t)} = X_j^{(k)} + \sqrt{t} S_j^{(k)}$ and $Z_N^{(k,t)} = N^{-1/2}(X_1^{(k,t)} + \cdots + X_N^{(k,t)})$. Applying Theorem 2 yields

$$\Phi^* \left(Z_N^{(1,t)}, \dots, Z_N^{(p,t)} \right) \geq \Phi^* \left(Z_{N+1}^{(1,t)}, \dots, Z_{N+1}^{(p,t)} \right).$$

Note that $Z_N^{(k,t)} = Z_N^{(k)} + \sqrt{t} S^{(N,k)}$ where for each fixed N, $S^{(N,k)} = N^{-1/2}(S_1^{(k)} + \cdots + S_N^{(k)})$, $k = 1, \dots, p$ is a family of centered freely iid semicircular variables, freely independent from $\{Z_N^{(k)}\}_k$ and having variance 1. So by the definition of χ^* above (having replaced S_k with $S^{(N,k)}$) we obtain

$$\begin{aligned} \chi^* \left(Z_N^{(1)}, \dots, Z_N^{(p)} \right) &= \frac{1}{2} \int_0^\infty \left[\frac{p}{1+t} - \Phi^* \left(Z_N^{(1,t)}, \dots, Z_N^{(p,t)} \right) \right] \, dt + c_p \\ &\leq \frac{1}{2} \int_0^\infty \left[\frac{p}{1+t} - \Phi^* \left(Z_{N+1}^{(1,t)}, \dots, Z_{N+1}^{(p,t)} \right) \right] \, dt + c_p \\ &= \chi^* \left(Z_{N+1}^{(1)}, \dots, Z_{N+1}^{(p)} \right), \end{aligned}$$

where $c_p = \frac{p}{2} \log 2\pi e$. □

References

[1] Artstein, S., Ball, K., Barthe, F., and Naor, A., *Solution of Shannon's Problem on the monotonicity of entropy.* Journal Amer. Math. Soc. **17** (2004), 975–982;

[2] Shlyakhtenko, D., *A free analogue of Shannon's problem on monotonicity of entropy.* Adv. Math. **208** (2007), no. 2, 824–833;

[3] Shlyakhtenko, D., *Notes on free probability theory.* Preprint, arXiv.org:0504063v1, 2005;

[4] Shlyakhtenko, D. and Schultz, H., *Shannon's monotonicity problem for free and classical entropy.* Proc. Nat. Acad. Sci. **104** No. 39 15254–15258 (2007);

[5] Voiculescu, D., *The analogues of entropy and of Fisher's information measure in free probability, V.* Invent. Math. **132** (1998), 189–227.

Brent Nelson, UCLA
email: **bnelson6@math.ucla.edu**

12 An isoperimetric inequality for uniformly log-concave measures and uniformly convex bodies

after E. Milman and S. Sodin [8]
A summary written by Diogo Oliveira e Silva

Abstract

We study the problem of transferring isoperimetric estimates for log-concave measures to normalized volume measures on convex bodies, making a conscious effort to highlight the parts of the theory which are of optimal transport nature.

12.1 Introduction to isoperimetric inequalities

12.1.1 The Euclidean case

Among all simple, closed plane curves of a given length L, the circle of circumference L encloses maximum area. This property is most succinctly expressed in terms of the isoperimetric inequality

$$L^2 \geq 4\pi A, \tag{1}$$

where A is the area enclosed by a curve γ of length L, and where equality holds if and only if γ is a circle. Inequality (1) can be proved in many different ways, but an especially elegant one is based on an ingenious combination of Wirtinger's inequality and Green's theorem on the plane, see [10]. In two dimensions the result for convex domains immediately implies the general result, but matters are quite different in higher dimensions. However, the following result still holds:

Theorem 1 (Isoperimetric inequality in $\mathbb{R}^n$)**.** *Let $A \subset \mathbb{R}^n$ be a bounded, open subset, and let $B \subset \mathbb{R}^n$ be a ball with the same volume as A. Then*

$$m(\partial A) \geq m(\partial B).$$

Since the theorem is stated somewhat informally, a few comments are in order. (i) By "volume" we mean the usual n-dimensional Lebesgue measure, henceforth denoted mes_n, whereas the boundary measure m should be interpreted as the $(n-1)$-dimensional Hausdroff measure. (ii) This is by no means the most general statement one can aspire to. For instance, if we merely assume that a given set $A \subset \mathbb{R}^n$ has finite perimeter on a bounded region $U \subset \mathbb{R}^n$ i.e. $1_A \in BV(U)$, then the following inequality still holds:

$$n\omega_n^{1/n}\mathrm{mes}_n(A)^{(n-1)/n} \leq \pi(\partial A),$$

where ω_n is the volume of the unit ball in $\mathbb{R}^n$ and $\pi(\cdot)$ denotes the distributional perimeter measure, see [5]. (iii) It is by now a well-understood fact that isoperimetric inequalities are equivalent, in a rather strong sense, to the ubiquitous Sobolev inequalities from the PDE world, of which the aforementioned Writinger's inequality is a special case.

Perhaps surprisingly, Theorem 1 is an immediate consequence of the Brunn-Minkowski inequality. Let us record here one of its more elementary versions: if A, B are nonempty, bounded open subsets of $\mathbb{R}^n$, then

$$\mathrm{mes}_n(A+B)^{1/n} \geq \mathrm{mes}_n(A)^{1/n} + \mathrm{mes}_n(B)^{1/n}. \tag{2}$$

This inequality, in turn, is an immediate consequence of another, closely related inequality of geometric nature which was originally established by Prékopa and Leindler. Several other proofs of the Brunn-Minkowski inequality are known, but of special interest to the reader might be one of optimal transport nature, see [12].

12.1.2 Other variants

Isoperimetric inequalities are known in a variety of other settings. Take, for instance, a simple closed curve on a 2-dimensional sphere of radius 1. The spherical isoperimetric inequality states that

$$L^2 \geq A(4\pi - A),$$

and that equality holds if and only if the curve is a circle. This inequality was proved in 1919 by Lévy, who also extended it to higher dimensions and more general surfaces. Even more generally, the isoperimetric problem can be formulated using the notion of *isoperimetric profile* of a metric measure space (X, d, μ). Isoperimetric profiles have been studied for Caley graphs and for special classes of Riemannian manifolds, where usually only regions with regular boundary are considered. See [10] and especially the multiple references therein.

Far reaching generalizations are known to exist in rather different directions. For the sake of presenting a concrete example, let us take a normed space $V = (\mathbb{R}^n, \|\cdot\|)$, and let μ be a probability measure on V with density $f = \exp(-g)$, for some function $g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$. If g is convex, the function f and the measure μ are called *log-concave*. Log-concave functions and measures have been extensively studied. Two of the most prominent examples include Gaussians and characteristic functions of convex bodies. It is easy to show that the class of log-concave functions is closed under taking products and, using the Prékopa-Leindler inequality, one can prove the same for convolutions. For more examples and many other properties, we refer the reader to [1, 7]. In this summary, as in [8], we restrict our attention to a more restricted class of measures. To define it, let $\delta : \mathbb{R}_+ \to \mathbb{R}_+ \cup \{\infty\}$ and consider the following condition:

$$\frac{g(x) + g(y)}{2} - g\Big(\frac{x+y}{2}\Big) \geq \delta(\|x - y\|). \tag{3}$$

Assume that the measure μ satisfies (3) with respect to the Euclidean norm $|\cdot|$ on $\mathbb{R}^n$ and $\delta(t) = t^2/8$. Then the following inequality holds:[5]

$$\mu^+_{|\cdot|}(A) \geq \phi(\Phi^{-1}(\mu(A)^*)), \tag{4}$$

where $\mu^+_{|\cdot|}(\cdot)$ denotes the so-called Minkowski boundary measure associated to μ, and $\mu(A)^* = \min\{\mu(A), 1 - \mu(A)\}$. Inequality (4) was proved by Bakry and Ledoux [2] and is the subject of A. Reznikov's summary in this volume. It extends the so-called *Gaussian isoperimetric inequality* proved by Sudakov and Tsirelson [11], and independently by Borell [4], which in particular states that, for fixed Gaussian volume, half-spaces have maximal Gaussian surface. It also provides the motivation for one of the main results of [8] which is the subject of the next section.

12.2 The main isoperimetric result

The main result will focus on measures μ satisfying (3) with respect to a function δ which satisfies

[5] As usual, $\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ denotes the density function of the standard Gaussian, and $\Phi(t) = \int_{-\infty}^{t} \phi(s) ds$ is the corresponding cumulative distribution function.

$$\begin{cases} \delta(t) > 0, & t > 0 \\ t \mapsto \delta(t)/t & \text{is nondecreasing.} \end{cases} \tag{5}$$

Before we state it, let us recall the notion of *Minkowski boundary measure* associated to the measure μ and the norm $\|\cdot\|$. It is defined by

$$\mu^+_{\|\cdot\|}(A) := \liminf_{\epsilon \to 0} \frac{\mu(A_{\epsilon,\|\cdot\|}) - \mu(A)}{\epsilon}, \quad A \subset \mathbb{R}^n,$$

where $A_{\epsilon,\|\cdot\|}$ denotes the ϵ-neighborhood of the set A with respect to the norm $\|\cdot\|$. Recall that $\mu(A)^* = \min\{\mu(A), 1-\mu(A)\}$.

Theorem 2. *Suppose μ satisfies conditions* (3) *and* (5). *Then*

$$\mu^+_{\|\cdot\|}(A) \geq C_\delta \mu(A)^* \gamma\Big(\log \frac{1}{\mu(A)^*}\Big) \quad \text{for every } A \subset \mathbb{R}^n, \tag{6}$$

where $\gamma(t) = t/\delta^{-1}(t/2)$.

A few remarks may help to further orient the reader. (i) The constant C_δ depends only on the function δ and can be explicitly calculated, see [8, Theorem 1.1]. (ii) We lose no generality in restricting attention to sets $A \subset \mathbb{R}^n$ for which $\mu(A) \leq 1/2$. (iii) Specializing to the case $\delta(t) = \alpha t^p$ (for some $\alpha > 0$ and $p \geq 2$), inequality (6) implies the existence of a universal constant $c > 0$ for which

$$\mu^+_{\|\cdot\|}(A) \geq c\alpha^{1/p} \mu(A)^* \log^{1-1/p}\Big(\frac{1}{\mu(A)^*}\Big).$$

Theorem 2 can then be easily seen to extend the aforementioned result of Bakry and Ledoux (up to constants). (iv) The proof of Theorem 2 consists of a localization lemma in terms of the so-called *μ-needles* which reduces matters to the much more tractable one-dimensional situation.

12.3 Applications to convex bodies

This section could have equally well been titled "How to transfer isoperimetric inequalities from log-concave measures to convex bodies". To understand why this is not a trivial problem, let $V = (\mathbb{R}^n, \|\cdot\|)$ be a normed space as before. The normalized volume measure $\lambda = \lambda_V$ on the unit ball of V *never* satisfies condition (3), no matter what function $\delta > 0$ we choose. Therefore, in order to prove an inequality for λ similar to (6), one cannot just appeal to the results from the previous section. We remedy this by appealing to the following general principle:[6]

Lipschitz maps preserve isoperimetric inequalities.

The strategy, very much in the spirit of optimal transport, will be to construct an auxiliary measure μ on V which does satisfy (3) and a map $T : (V, \mu) \to (V, \lambda)$ which *transports* μ *to* λ in the sense that $T_*\mu = \lambda$. After appropriate Lipschitz bounds have been established for T, the result will follow from Theorem 2 together with the general principle mentioned above. All of this requires careful justification.

As a warm-up model, let us consider the case of p-uniformly convex bodies ($p \geq 2$). Let μ be the probability measure with density given by

$$f(x) = \frac{\exp(-\|x\|^p)}{\Gamma(1 + n/p) \cdot \mathrm{mes}_n(\{\|x\| \leq 1\})}$$

with respect to Lebesgue measure mes_n. Bobkov and Ledoux [3] proved the existence of a map $S : V \to V$ which transports μ to λ in the sense described above, and for which

$$\|S\|_{\mathrm{Lip}} \leq \frac{C}{\Gamma(1 + n/p)^{1/n}}.$$

This can be combined with (6) to yield an isoperimetric inequality for λ.

We would like to generalize this discussion to arbitrary uniformly convex spaces $V = (\mathbb{R}^n, \|\cdot\|)$. Recall that a normed space V is said to be *uniformly convex* if its modulus of convexity[7] satisfies $\delta_V(\epsilon) > 0$ for all $\epsilon > 0$. An important property of the modulus of convexity, first observed in the work of Figiel and Pisier [6], is the following inequality:

$$\frac{\|x\|^2 + \|y\|^2}{2} - \left\|\frac{x+y}{2}\right\|^2 \geq c\delta_V\Big(\frac{\|x-y\|}{4}\Big), \tag{7}$$

[6] One way to turn this principle into a rigorous statement is via the notion of "isoperimetric profile" which we already mentioned (but did not define) in the introduction.

[7] By definition, this is the function $\delta_V : [0, 2] \to [0, 1]$ defined as

$$\delta_V(\epsilon) = \inf\Big\{1 - \left\|\frac{x+y}{2}\right\| : \|x\|, \|y\| \leq 1, \|x-y\| \geq \epsilon\Big\}.$$

valid for all $x, y \in \mathbb{R}^n$ for which $\|x\|^2 + \|y\|^2 \leq 2$. Now, choose μ to be a probability measure on $\mathbb{R}^n$ with density

$$f(x) = \frac{1}{C} \exp\Big(- \frac{n}{c} \|4x\|^2 \Big) \mathbf{1}_{\{\|x\| \leq 1/4\}}(x)$$

with respect to Lebesgue measure mes_n, where $C > 0$ is a scaling factor. Inequality (7) clearly implies that μ is uniformly log-concave in the sense of (3), and so we can apply Theorem 2 to deduce an isoperimetric inequality for μ. To transfer this inequality to the measure λ, we need to extend the result of Bobkov and Ledoux mentioned in the previous paragraph.

Let $d\mu = f d\mathrm{mes}_n$ be an *even* log-concave probability measure. Ball [1] showed that

$$K_f = \Big\{ x \in \mathbb{R}^n : n \int_0^\infty f(rx) r^{n-1} dr \geq 1 \Big\}$$

defines a symmetric convex body i.e. the unit ball of a certain norm $\|\cdot\|_{K_f}$. Moreover, one can easily see that there exists a canonical radial map T_f transporting μ to the restriction λ of the Lebesgue measure on K_f. The following theorem contains the Bobkov-Ledoux result as a particular case:

Theorem 3. *Let $d\mu = f d\mathrm{mes}_n$ be an even log-concave probability measure on $\mathbb{R}^n$, let λ denote the restriction of Lebesgue measure to K_f, and let $T = T_f$ denote the unique radial map such that $T_* \mu = \lambda$. Then, as a map $T : V \to V$ where $V = (\mathbb{R}^n, \|\cdot\|)$, we have that $\|T\|_{\mathrm{Lip}} \leq C f(0)^{1/n}$, for some universal constant $C > 0$.*

As before, Theorems 2 and 3 conspire together to yield an isoperimetric inequality for the uniform measure $\lambda = \lambda_V$ on the unit ball of the uniformly convex space $V = (\mathbb{R}^n, \|\cdot\|)$. To wit,

$$\lambda^+_{\|\cdot\|}(A) \gtrsim_{n,\delta} \frac{\lambda(A)^* \log \frac{1}{\lambda(A)^*}}{\delta^{-1}\big(\frac{1}{2n} \log \frac{1}{\lambda(A)^*}\big)}.$$

The proof of Theorem 3 involves a neat approximation argument, together with several interesting facts about the geometry of log-concave functions. Details will be provided upon arrival at the summer school.

References

[1] Ball, K., *Logarithmically concave functions and sections of convex sets in $\mathbb{R}^n$.* Studia Math. **88** (1988), no. 1, 69–84.

[2] Bakry, D. and Ledoux, M., *Lévy-Gromov's isoperimetric inequality for an infinite-dimensional diffusion generator.* Invent. Math. **123** (1996), no. 2, 259–281.

[3] Bobkov, S. G. and Ledoux, M., *From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities.* Geom. Funct. Anal. **10** (2000), no. 5, 1028–1052.

[4] Borell, Ch., *The Brunn-Minkowski inequality in Gauss space.* Invent. Math. **30** (1975), no. 2, 207–216.

[5] Evans, L. C. and Gariepy, R. F., *Measure Theory and Fine Properties of Functions.* Studies in Advanced Mathematics, CRC Press, 1992.

[6] Figiel, T. and Pisier, G., *Séries aléatoires dans les espaces uniformément convexes ou uniformément lisses.* C. R. Acad. Sci. Paris Sér. A **279** (1974), 611–614.

[7] Klartag, B. and Milman, V. D., *Geometry of log-concave functions and measures.* Geom. Dedicata **112** (2005), pp. 169–182.

[8] Milman, E. and Sodin, S., *An isoperimetric inequality for uniformly log-concave measures and uniformly convex bodies.* J. Funct. Anal. **254** (2008), no. 5, pp. 1235–1268.

[9] Oliveira e Silva, D., *O que é um círculo?* Números, Cirurgias e Nós de Gravata, IST Press, 2012, pp. 217–229.

[10] Osserman, R., *The isoperimetric inequality.* Bull. Amer. Math. Soc. **84** (1978), no. 6, pp. 1182–1238.

[11] Sudakov, V. N. and Tsirelson, B. S., *Extremal properties of half-spaces for spherically invariant measures. Problems in the theory of probability distributions, II.* Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **41** (1974), 14–24, 165.

[12] Villani, C., *Topics in optimal transportation.* Graduate Studies in Mathematics, 58. American Mathematical Society, 2003.

Diogo Oliveira e Silva, HCM Bonn
email: **dosilva@math.uni-bonn.de**

13 Lévy-Gromov's isoperimetric inequality for an infinite dimensional diffusion generator

after D. Bakry and M. Ledoux, [1]
A summary written by Alexander Reznikov

Abstract

Using a semigroup approach, authors establish a certain inequality for a diffusion generator of infinite dimension and positive curvature. The definitions of "dimension" and "curvature" are pure algebraic, but with a strong geometric meaning. Also, the Sobolev logarithmic inequality is discussed.

13.1 Introduction and main results

Suppose $\Phi(r) = \int\limits_{-\infty}^{r} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$, and $\phi(r) = \Phi'(r)$. We denote

$$\mathcal{U} = \phi \circ \Phi^{-1}.$$

The main (and almost the only important) property of the function $\mathcal{U}$ is the ODE it satisfies:

$$\mathcal{U}'' = -\frac{1}{\mathcal{U}}. \tag{1}$$

We start with proving a "baby" version of the main theorem of the paper. Suppose γ is the Gaussian measure on $\mathbb{R}^n$ (i.e., the measure with density $\frac{1}{(2\pi)^{\frac{k}{2}}} e^{-\frac{|x|^2}{2}}$). Then the following inequality holds.

Theorem 1. *Using the above notation, it is true that*

$$\mathcal{U}(\int f d\gamma) \leq \int \sqrt{\mathcal{U}(f)^2 + |\nabla f|^2} d\gamma.$$

This inequality is established using the following semigroup:

$$P_t(f)(x) = \int f(e^{-t}x + (1 - e^{-2t})^{\frac{1}{2}} y) d\gamma(y).$$

We are ready consider the general case.

Let $(E, \mathcal{E}, \mu)$ be a measure space, and $(P_t)_{t\geq 0}$ be a semigroup, continuous in $L^2(\mu)$. Let L be a generator associated to P_t. This means that, basically,

$$\frac{d}{dt}P_t f = LP_t f = P_t Lf.$$

In other words, vaguely speaking, $P_t = e^{tL}$.

We assume that there is an algebra of functions $\mathcal{A}$, which is nice enough. The exact meaning of the word "nice" will be described later. Introduce a bilinear operator

$$\Gamma(f, g) = \frac{1}{2} \cdot (L(fg) - fLg - gLf),$$

and the iterated operator

$$\Gamma_2(f, g) = \frac{1}{2} \cdot (L\Gamma(f, g) - \Gamma(f, Lg) - \Gamma(g, Lf)).$$

We are ready to define the dimension and curvature of the operator L.

Definition 2. *We say that the operator L is of dimension n and curvature R, if for any function $f \in \mathcal{A}$ the following inequality holds:*

$$\Gamma_2(f, f) \geq R\Gamma(f, f) + \frac{1}{n}(Lf)^2.$$

We say that the operator L is of infinite dimension and curvature R, if

$$\Gamma_2(f, f) \geq R\Gamma(f, f).$$

To state the main theorem we need one more definition.

Definition 3. *The operator L is called a diffusion operator if for any C^∞ function Ψ on $\mathbb{R}^k$, and every finite family $F = (f_1, \ldots, f_k)$ in $\mathcal{A}^k$, the following holds:*

$$L\Psi(F) = \nabla\Psi(F) \cdot LF + \nabla\nabla\Psi(F) \cdot \Gamma(F, F).$$

Remark 4. *This definition essentially says that L is a second order differential operator with no constant term.*

For the sake of simplicity we denote $\Gamma(f) = \Gamma(f, f)$, and $\Gamma_2(f) = \Gamma_2(f, f)$. We are ready to state the main theorem of the paper.

Theorem 5. *Let L be a diffusion operator, that generates the semigroup P_t. Let the algebra $\mathcal{A}$ be dense in the L^2-domain of L, and stable under L, P_t and action of C^∞ functions which are zero at zero. Suppose that L is of infinite dimension and curvature R. Then, for any f from $\mathcal{A}$ with values in $[0,1]$, every $\alpha \geq 0$, every $t \geq 0$ the following inequality holds:*

$$\sqrt{\mathcal{U}(P_t f)^2 + \alpha\Gamma(P_t f)} \leq P_t\left(\sqrt{\mathcal{U}(f)^2 + c_\alpha(t)\Gamma(f)}\right),$$

where

$$c_\alpha(t) = \frac{1 - e^{-2Rt}}{R} + \alpha e^{-2Rt}.$$

13.1.1 Corollaries

It is interesting to play with parameters α and t in the previous theorem. First, send $\alpha \to \infty$. Then we get the following.

Corollary 6. *For any $t \geq 0$,*

$$\sqrt{\Gamma(P_t f)} \leq e^{-Rt} P_t(\sqrt{\Gamma(f)}).$$

Sadly, we will need to use this to prove the main theorem. Thus, it requires another proof.

Next, we put $\alpha = \frac{1}{R}$. Then we have $c_{\frac{1}{R}}(t) = \frac{1}{R}$ for every t.

Corollary 7. *For any $t \geq 0$,*

$$\sqrt{R\mathcal{U}(P_t f)^2 + \Gamma(P_t f)} \leq P_t\left(\sqrt{R\mathcal{U}(f)^2 + \Gamma(f)}\right).$$

In the next corollary we assume that μ is a probability measure and send $t \to \infty$. We make an additional assumption that $P_\infty f = \int f d\mu$. Then we get the following.

Corollary 8. *Under assumptions above, it holds that*

$$\sqrt{R}\mathcal{U}(\int f d\mu) \leq \int \left(\sqrt{R\mathcal{U}(f)^2 + \Gamma(f)}\right) d\mu.$$

Finally, we set $\alpha = 0$. Then $c_0(t) = \frac{1-e^{-2Rt}}{R}$, and what we get is the following.

Corollary 9. *For any $t \geq 0$,*

$$\mathcal{U}(P_t f) \leq P_t\left(\sqrt{\mathcal{U}(f)^2 + c_0(t)\Gamma(f)}\right).$$

13.2 Remark on the geometry

We feel obligated to state the geometrical meaning of results above. Suppose $d(x,y) = \text{esssup}|f(x) - f(y)|$ — a pseudo-metric on the space E. The supremum here is taken over all functions $f \in \mathcal{A}$.

We assume further that $\Gamma(f)$ is almost the square of modulus of the gradient of f. More precisely, we assume

$$\sqrt{\Gamma(f)} = \limsup_{d(x,y)\to 0} \frac{|f(x) - f(y)|}{d(x,y)}.$$

Under these assumptions, when f approximates the indicator function of a closed set A, the $\int \sqrt{\Gamma(f)}d\mu$ approaches the

$$\mu_s(\partial A) = \liminf_{r\to 0} \frac{1}{r}(\mu(A_r) - \mu(A)),$$

where A_r is the r-neighborhood of A. Now we use the Corollary 8. Since $\mathcal{U}(1) = 0$, the term $\mathcal{U}(f)$ disappears. Thus, with $R = 1$, we can read the Corollary 8 as follows:

$$\mathcal{U}(\mu(A)) \leq \mu_s(\partial A).$$

This extends the known inequality

$$\mathcal{U}(\gamma(A)) \leq \gamma_s(\partial A)$$

for the Gaussian measure γ and any Borel set $A \in \mathbb{R}^k$.

13.3 Further extensions

Two more results deserve to be stated here. First one is a "multidimensional" (or, as authors prefer to say, "tensorised") version of the main theorem. We assume that there are two spaces $E^{1,2}$, two semigroups $P^{1,2}$ and, thus, two operators $\Gamma^{1,2}$. We suppose that each of them satisfy an inequality from the main theorem, i.e.

$$\sqrt{\mathcal{U}(P^i f)^2 + \alpha^i \Gamma^i(P^i f)} \leq P^i \left(\sqrt{\mathcal{U}(f)^2 + \beta^i \Gamma^i(f)} \right),\ i = 1, 2.$$

Then the following holds.

Theorem 10. *Suppose $f : E^1 \times E^2 \to [0,1]$ is in the domain of $\Gamma^1 \times \Gamma^2$. Then*

$$\sqrt{\mathcal{U}(P^1P^2f)^2 + \alpha^1\Gamma^1(P^1P^2f) + \alpha^2\Gamma^2(P^1P^2f)} \leq \\ P^1P^2\left(\sqrt{\mathcal{U}(f)^2 + \beta^1\Gamma^1(f) + \beta^2\Gamma^2(f)}\right),$$

The next theorem is a version of the logarithmic Sobolev inequality. In fact, let

$$E(f) = \int f\log(f)d\mu - \int fd\mu \cdot \log(\int fd\mu).$$

The following theorem holds.

Theorem 11. *Suppose that the operator Γ, defined as usual, satisfies*

$$\mathcal{U}(\int fd\mu) \leq \int \sqrt{\mathcal{U}(f)^2 + \Gamma(f)}d\mu$$

for every function $f \in \mathcal{A}$ with values in $[0,1]$. Then Γ satisfies the logarithmic Sobolev inequality:

$$E(f^2) \leqslant 2\int \Gamma(f)d\mu$$

for every $f \in \mathcal{A}$.

References

[1] Bakry D. and Ledoux M. *Lévy-Gromov's isoperimetric inequality for an infinite dimensional diffusion generator.* Inventiones mathematicae. **123.2** (1996), 259–281;

Alexander Reznikov, MSU
email: **rezniko2@msu.edu**

14 Resolution of Shannon's problem on the monotonicity of entropy

after S. Artstein, K. Ball, F. Barthe, A. Naor [1]
A summary written by Ed Scerbo

Abstract

Let $X_1, X_2, \ldots,$ be a sequence of iid random variables, and let Y_n be defined by $Y_n = \frac{X_1 + \ldots X_n}{\sqrt{n}}$. We prove that the entropy of Y_n, $H(Y_n)$, is nondecreasing in n by proving a more general statement for sums of non-identically distrbuted random variables.

14.1 Introduction

Remark 1. *Throughout this summary, we adhere to the following conventions:*

- log *shall always denote natural log,*
- $0 \log 0$ *is defined to be* 0.

In the 1940s, Shannon [3] introduced the notion of entropy in the information theoretic sense:

Definition 2. *Let X be a random variable. If X has a density f, the entropy of X, $H(X)$, is $-\int f \log f$, assuming the integral exists. If X does not have a density, we set $H(X) = -\infty$.*

To avoid questions of the existence of $H(X)$, we point out that the integral defining $H(X)$ exists and $H(X) < \infty$ if X has finite variance. Thus, henceforth we shall restrict our discussion to random variables with finite variance.

Some basic properties of H follow:

- $H(X)$ depends only on the distribution of X: $H(X) = H(Y)$ if $X \stackrel{d}{=} Y$.
- $H(X)$ is translation-invariant: $H(X + c) = H(X)$.
- $H(X)$ is not scale-invariant: $H(cX) = H(X) + \log |c|$.

- Among all random variables of a given (finite) variance, H is maximized by Gaussians: If X is a random variable with variance $\sigma^2 < \infty$ and G is a Gaussian with the same variance, then

$$H(X) \leq H(G) = \frac{1}{2}\log(2\pi e\sigma^2).$$

- If X, Y are independent random variables, then $H(X+Y) \geq H(X)$.

Let $X_1, X_2, \ldots$ be a sequence of iid random variables with variance $\sigma^2 < \infty$ and mean μ, and set $Y_n = \frac{X_1+\ldots+X_n}{\sqrt{n}}$. The central limit theorem states that $Y_n - \sqrt{n}\mu$ converges in distribution to a centered Gaussian G with variance σ^2. In light of this, translation-invariance of H, and the fact that H is maximized by Gaussians, it is reasonable to wonder about the behavior of $H(Y_n)$ as a function of n: Does it limit to $H(G)$? Does it do so monotonically?

If each Y_n has a distribution with atomic part, then clearly $-\infty = H(Y_n) \nrightarrow H(G)$. However, the following theorem of Shannon [3] hints that the monotonicity of $H(Y_n)$ might still be salvageable:

Proposition 3. *$H(Y_2) \geq H(Y_1)$, and consequently $H(Y_{2^k}) \geq H(Y_{2^{k-1}})$, for all $k \in \mathbb{N}$.*

(In fact, the first rigorous proof of the above Proposition was proved by Stam [4].) Thus, it was conjectured that the entire sequence $H(Y_n)$ was nondecreasing in n. However, it remained unknown even whether $H(Y_3) \geq H(Y_2)$. This conjecture was finally resolved in the affirmative in 2004 [1]. Before we begin our proof, we require the notion of Fisher information.

14.1.1 Fisher information and connection with entropy

Definition 4. *If X is a random variable with a density f that is everywhere differentiable and strictly positive, the Fisher information of X, $J(X)$, is $\int \frac{(f')^2}{f}$. If X does not have a density, we set $J(X) = \infty$.*

Remark 5. *$J(X)$ can be defined for a more general class of random variables, but we will have no need for greater generality, so we define $J(X)$ as above.*

$J(X)$ enjoys a list of properties which are rather complementary to those of $H(X)$ given above; we, however, will need but two:

- $J(X)$ depends only on the distribution of X: $J(X) = J(Y)$ if $X \stackrel{d}{=} Y$.
- $J(X)$ is not scale-invariant: $J(cX) = \frac{1}{c^2} J(X)$.

The connection between H and J we require is de Bruijn's identity (cf. Barron [2]):

Theorem 6 (de Bruijn's identity)**.** *Let X be a random variable with finite variance, and let G be an independent standard Gaussian. Then for all $t > 0$,*

$$\frac{d}{dt} H(X + \sqrt{t} G) = \frac{1}{2} J(X + \sqrt{t} G).$$

Combining the properties of H and J above with Theorem 6 and an application of the chain rule, we arrive at

Corollary 7. *Let X, G be as above. Then for all $t > 0$,*

$$\frac{d}{dt} H\left(\sqrt{e^{-2t}} X + \sqrt{1 - e^{-2t}} G\right) = J\left(\sqrt{e^{-2t}} X + \sqrt{1 - e^{-2t}} G\right) - 1.$$

Combining this with the Fundamental Theorem of Calculus and an ad hoc argument to ensure convergence at the endpoints, we obtain the principal identity we require:

Corollary 8. *Let X, G be as above. Then*

$$H(X) = \frac{1}{2} \log(2\pi e) - \int_0^\infty \left[J\left(\sqrt{e^{-2t}} X + \sqrt{1 - e^{-2t}} G\right) - 1 \right] dt. \qquad (1)$$

Remark 9. *The above identities are meaningful because for X, G as above and $a, b > 0$, $aX + bG$ has a density that is everywhere differentiable and strictly postive. In other words, all the Fisher informations in the above formulas are defined.*

We are now ready to state and prove our main theorems.

14.2 Main results

As stated above, the conjecture on monotonicity of $H(Y_n)$ has been resolved affirmatively:

Theorem 10 (Artstein-Ball-Barthe-Naor, [1])**.** *Let $X_1, X_2, \ldots$ be a sequence of iid random variables with finite variance. Then*

$$H\left(\frac{X_1 + \ldots + X_n}{\sqrt{n}}\right) \le H\left(\frac{X_1 + \ldots + X_{n+1}}{\sqrt{n+1}}\right).$$

Just as there are versions of the central limit theorem for random variables that are not identically distributed, we have the following version of Theorem 10 for non-identically distributed random variables:

Theorem 11 (Artstein-Ball-Barthe-Naor, [1])**.** *Let $X_1, \ldots, X_{n+1}$ be independent random variables, each with finite variance, and let $(a_1, \ldots, a_{n+1}) \in S^n$ be a unit vector with no $a_j = 1$. Then*

$$H\left(\sum_{i=1}^{n+1} a_i X_i\right) \ge \sum_{j=1}^{n+1} \frac{1-a_j^2}{n} \cdot H\left(\frac{1}{\sqrt{1-a_j^2}} \sum_{i \neq j} a_i X_i\right).$$

Note that Theorem 10 is a direct consequence of Theorem 11, as seen by setting each $a_j = \frac{1}{\sqrt{n+1}}$ and using the fact that $H(X) = H(Y)$ if $X \stackrel{d}{=} Y$.

We begin the proof of Theorem 11 with a sequence of lemmas:

Lemma 12. *Let $X_1, \ldots, X_n$ be independent random variables, and suppose they have a strictly positive joint density w with mild regularity and decay properties (precise sufficient conditions are given in [1]). Let $a = (a_1, \ldots, a_n) \in S^{n-1}$. Then for any vector field p on $\mathbb{R}^n$ satisfying its own mild smoothness and decay properties (again, see [1]) and the additional property that for each $x \in \mathbb{R}^n$, $\langle p(x), a\rangle = 1$, we have*

$$J\left(\sum_{i=1}^{n} a_i X_i\right) \le \int_{\mathbb{R}^n} \left[\frac{div(pw)}{w}\right]^2 w.$$

Moreover, equality holds for some such p.

We state the above Lemma for dimension n, although in our application we shall use it variously in dimensions n and $n+1$.

Lemma 13. *Let $X_1, \ldots, X_{n+1}$ be independent random variables with finite variances, and suppose each X_i has density f_i. Let w denote the joint density, $w(x_1, \ldots x_{n+1}) = \prod_{j=1}^{n+1} f_j(x_j)$, and suppose w is strictly positive and* *satisfies the smoothness and decay properties of Lemma 12. Then for any $a = (a_1, \ldots, a_n) \in S^n$ such that no $a_j = 1$, we have*

$$J\left(\sum_{i=1}^{n+1} a_i X_i\right) \leq \sum_{j=1}^{n+1} \frac{1-a_j^2}{n} \cdot J\left(\frac{1}{\sqrt{1-a_j^2}} \sum_{i \neq j} a_i X_i\right).$$

Proof. For each $j = 1, \ldots, n+1$, set

$$\hat{a}_j = \frac{1}{\sqrt{1-a_j^2}}(a_1, \ldots, a_{j-1}, 0, a_{j+1}, \ldots, a_{n+1}).$$

(Note that while a_j is the j^{th} component of the unit vector a, $\hat{a}_j$ is itself a unit vector in $\mathbb{R}^{n+1}$.) Also, let X denote the random vector $(X_1, \ldots, X_{n+1})$. Let $p^j : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ be a vector field which realizes $J(\langle \hat{a}_j, X\rangle)$ as in Lemma 12. That is, $\langle p^j, \hat{a}_j \rangle \equiv 1$, and

$$J\left(\frac{1}{\sqrt{1-a_j^2}} \sum_{i \neq j} a_i X_i\right) = \int_{\mathbb{R}^{n+1}} \left[\frac{\operatorname{div}(wp^j)}{w}\right]^2 w.$$

We may choose p^j to not depend on the j^{th} coordinate x_j and to have j^{th} component identically 0: Think of $\hat{a}_j$ as actually a vector in $S^{n-1} \subset \mathbb{R}^n$ by the obvious projection, and apply Lemma 12 to this new $\hat{a}_j$ and the random variables $(X_i)_{i \neq j}$ to get a vector field on $\mathbb{R}^n$. (It is simple to check that the joint density of $(X_i)_{i \neq j}$ inherits the required regularity and decay from w, so Lemma 12 applies.) We may then artificially reinsert the missing j^{th} coordinate to construct p^j.

Define the vector field $p : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ by $p = \sum_{j=1}^{n+1} \frac{1}{n}\sqrt{1-a_j^2} p^j$. p is a vector field on $\mathbb{R}^{n+1}$ satisfying the same mild smoothness and decay properties as the constituent p^j's, and $\langle p, a \rangle \equiv 1$. (This last holds because of how we constructed the p^j's.) Thus, by Lemma 12,

$$
\begin{aligned}
J\left(\sum_{i=1}^{n+1} a_i X_i\right) &\le \int_{\mathbb{R}^{n+1}} \left[\frac{\operatorname{div}(wp)}{w}\right]^2 w \\
&= \int_{\mathbb{R}^{n+1}} \left[\sum_{j=1}^{n+1} \frac{\sqrt{1-a_j^2}}{n} \cdot \frac{\operatorname{div}(wp^j)}{w}\right]^2 w \\
&\le \sum_{j=1}^{n+1} \frac{1-a_j^2}{n} \cdot \int_{\mathbb{R}^{n+1}} \left[\frac{\operatorname{div}(wp^j)}{w}\right]^2 w \\
&= \sum_{j=1}^{n+1} \frac{1-a_j^2}{n} \cdot J\left(\frac{1}{\sqrt{1-a_j^2}} \sum_{i \neq j} a_i X_i\right),
\end{aligned}
$$

where the second inequality holds by Lemma 14: Take $m = n+1$, $H = L^2(w) = L^2(\mathbb{R}^{n+1}, w\,dx)$,

$$
T_j \phi(x) = \int_{\mathbb{R}} \phi(x_1, \ldots, x_{j-1}, u, x_{j+1}, \ldots, x_{n+1}) f_j(u) du,
$$

and

$$
y_j = \frac{\sqrt{1-a_j^2}}{n} \cdot \frac{\operatorname{div}(wp^j)}{w}.
$$

□

Lemma 14. *Let $T_1, \ldots, T_m$ be commuting orthogonal projections on a Hilbert space H, and suppose we have $y_1, \ldots, y_m \in H$ such that for each $j = 1, \ldots, m$, $T_1 \cdots T_m y_j = 0$. Then*

$$
||T_1 y_1 + \ldots T_m y_m||^2 \le (m-1)\left(||y_1||^2 + \ldots + ||y_m||^2\right).
$$

Thus we are able to improve upon the trivial bound yielded by Cauchy-Schwarz by using specfic properties satisfied by the objects in question.

The proof of Theorem 11 follows:

Proof of Theorem 11. We may without loss of generality assume that each X_i has a compactly supported density. Let $G_1, \ldots, G_{n+1}$ be standard Gaussians such that $X_1, \ldots, X_{n+1}, G_1, \ldots, G_{n+1}$ are independent. Let G be another standard Gaussian that is independent of $\sum_{i=1}^{n+1} a_i X_i$, and set $X_i^{(t)} =$

$\sqrt{e^{-2t}}X_i + \sqrt{1-e^{-2t}}G_i$ for $t \geq 0$. Using (1) together with the facts that $\sum_{i=1}^{n+1} a_i G_i \stackrel{d}{=} G$ and that J depends only on the distribution of its input, we see that

$$H\left(\sum_{i=1}^{n+1} a_i X_i\right) = \frac{1}{2}\log(2\pi e) - \int_0^\infty \left[J\left(\sum_{i=1}^{n+1} a_i X_i^{(t)}\right) - 1\right] dt,$$

and similarly for $H\left(\frac{1}{\sqrt{1-a_j^2}}\sum_{i\neq j} a_i X_i\right)$. Thus, it suffices to show that for any $t > 0$,

$$J\left(\sum_{i=1}^{n+1} a_i X_i^{(t)}\right) \leq \sum_{j=1}^{n+1} \frac{1-a_j^2}{n} \cdot J\left(\frac{1}{\sqrt{1-a_j^2}}\sum_{i\neq j} a_i X_i^{(t)}\right).$$

Since each X_i has compactly supported density, the joint density $w^{(t)}$ of $X_1^{(t)}, \ldots, X_{n+1}^{(t)}$ has enough smoothness and decay to apply Lemma 12, completing the proof. □

References

[1] S. Artstein, K. Ball, F. Barthe, A. Naor *Solution of Shannon's problem on the monotonicity of entropy.* J. Amer. Math. Soc. **17** (2004), 975-982;

[2] A. R. Barron *Entropy and the central limit theorem.* Ann. Probab. **14** (1986), 336-342;

[3] C. E. Shannon and W. Weaver *The mathematical theory of communication* University of Illinois Press, Urbana, IL, (1949);

[4] A.J. Stam *Some inequalities satisfied by the quantities of information of Fisher and Shannon* Inform. and Control **2** (1959), 101-112.

Ed Scerbo, University of California, Berkeley
email: **efscerbo@berkeley.edu**

15 Wasserstein Gradient Flows and Evolution Equations

after R. Jordan, D. Kinderlehrer, and F. Otto [1]
A summary written by Chris D. White

Abstract

We begin by discussing the abstract notion of gradient flows in the space of probability measures equipped with the Wasserstein distance via a time-discretization scheme. We then use this notion to show that a large class of evolution equations can be viewed as gradient flows with respect to the Wasserstein distance for certain convex energy functionals. This allows us to make precise sense of the folk knowledge that diffusion maximizes entropy.

15.1 Time Discretized Gradient Flows

The Wasserstein distance of order two between two probability measures on $\mathbb{R}^n$ is defined by

$$d(\mu_1, \mu_2)^2 = \inf_{p \in \mathcal{P}(\mu_1, \mu_2)} \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 \; p(dxdy)$$

where $\mathcal{P}(\mu_1, \mu_2)$ is the set of all probability measures whose first marginal is μ_1 and whose second marginal is μ_2. On the space of probability measures with finite second moment, the Wasserstein distance is in fact a metric and this space is complete with respect to this metric. In order to discuss the notion of *gradient flow* on such a space with no inner product, we need to first introduce some concepts.

Let $F : \mathcal{H} \to \mathbb{R}$ be a strictly convex functional on a Hilbert space $(\mathcal{H}, \langle \cdot \rangle)$. We know from basic theory that the gradient flow

$$\begin{cases} \frac{\partial \rho}{\partial t} = -\nabla F(\rho(t)) \\ \\ \rho(0) = \rho^0 \end{cases} \tag{1}$$

for F will always converge to a unique stationary state, F^*, independently of the initial condition; moreover this stationary state is the unique minimizer of F. A simple time-discretization of (1), called the *backward Euler method* in the ODE literature, proceeds as follows. Let $\tilde{\rho}^0 := \rho^0$, and iteratively define $\tilde{\rho}^{k+1}$ implicitly via

$$\frac{\tilde{\rho}^{k+1} - \tilde{\rho}^k}{h} = -\nabla F(\tilde{\rho}^{k+1}). \tag{2}$$

Proposition 1. *The solution $\tilde{\rho}^{k+1}$ to equation (2) is given by*

$$\arg\min_{u \in \mathcal{H}} F(u) + \frac{1}{2h} d(u, \tilde{\rho}^k)^2 \tag{3}$$

Proof. u minimizes (3) if and only if

$$0 = \nabla F(u) + \frac{1}{h}(u - \tilde{\rho}^k)$$

which is precisely the solution to (2). $\square$

Remark 2. *The operator defined by* $\operatorname{prox}_F(\tilde{\rho}) := \arg\min_{u \in \mathcal{H}} F(u) + \frac{1}{2h} d(u, \tilde{\rho}^k)^2$ *is called the proximal operator of F in the convex optimization literature.*

Proposition (1) allows us to consider the iterative scheme (2) in the more general setting of metric spaces (without referencing gradients). For the purposes of this paper, then, if the time discretized solutions to (3) converge as $h \downarrow 0$ we will call the resulting limit the *gradient flow* for F with respect to the metric d.

15.2 Fokker-Planck equations

Now we consider the general class of Fokker-Planck equations

$$\frac{\partial \rho}{\partial t} = \operatorname{div}(\nabla \Psi(x) \rho) + \beta^{-1} \Delta \rho, \quad \rho(x, 0) = \rho^0(x) \tag{4}$$

where we assume the potential satisfies

$$\begin{aligned} \Psi \in C^\infty(\mathbb{R}^n) \\ \Psi(x) \geq 0 \\ |\nabla \Psi(x)| \leq C(\Psi(x) + 1). \end{aligned} \tag{5}$$

It is well known that equation (4) has a unique stationary distribution given by

$$\rho_s(x) \propto \exp(-\beta \psi(x))$$

which is easily seen to be the unique minimizer of the following free energy functional over the space of probability densities on $\mathbb{R}^n$:

$$F(\rho) = \int_{\mathbb{R}^n} \Psi\rho \, dx + \beta^{-1} \int_{\mathbb{R}^n} \rho \log \rho \, dx.$$

However, it is not immediately obvious by what path the Fokker-Plank equation minimizes F. Surprisingly, our main result states that (4) follows the gradient flow of F with respect to the Wasserstein metric.

Following the discussion in Section §15.1, we consider the semi-discrete scheme (3) for approximating the Fokker-Planck dynamics; namely for a given time step $h > 0$ we solve

$$\rho_h^{(k+1)} = \arg\min_{\rho \in K} F(\rho) + \frac{1}{2h} d(\rho, \rho_h^{(k)})^2 \tag{6}$$

where d is the Wasserstein metric with respect to quadratic cost. Define the admissible class of probability densities on $\mathbb{R}^n$ to be

$$K := \left\{ \rho : \mathbb{R}^n \to [0, \infty) \text{ measurable } : \int_{\mathbb{R}^n} \rho(x) dx = 1, \int_{\mathbb{R}^n} |x|^2 \rho(x) dx < \infty \right\}.$$

We can now state the main result.

Theorem 3. *[1]*

Let $\rho^0 \in K$ satisfy $F(\rho^0) < \infty$, and for given $h > 0$, let $\{\rho_h^{(k)}\}_{k \in \mathbb{N}}$ be the solution of (6). Define the interpolation $\rho_h : (0, \infty) \times \mathbb{R} \to [0, \infty)$ by

$$\rho_h(t) = \rho_h^{(k)} \quad \textit{for } t \in [k, (k+1)/h) \textit{ and } k \in \mathbb{N} \cup \{0\}.$$

Then as $h \downarrow 0$,

$$\rho_h(t) \rightharpoonup \rho(t) \quad \textit{weakly in } L^1(\mathbb{R}^n) \quad \textit{for all } t \in (0, \infty),$$

where $\rho \in C^\infty((0, \infty) \times \mathbb{R}^n)$ is the unique solution of

$$\frac{\partial \rho}{\partial t} = \operatorname{div}(\rho \nabla \Psi) + \beta^{-1} \Delta \rho,$$

with initial condition

$$\rho(t) \rightharpoonup \rho^0 \quad \textit{strongly in } L^1(\mathbb{R}^n) \quad \textit{for } t \downarrow 0$$

and

$$M(\rho), E(\rho) \in L^\infty((0, T)) \quad \textit{for all } T < \infty.$$

Proof (Sketch). We first must verify that the functional defined by (6) is well-defined. As the functional is strictly convex, it has at most one minimizer. Appealing to the entropy bound

$$\int_{\mathbb{R}^n} \rho \log \rho \; dx \geq -C\Big(\int_{\mathbb{R}^n} |x|^2 \rho(x) dx + 1\Big)^\alpha$$

and the inequality

$$\int_{\mathbb{R}^n} |x|^2 \rho_1(x) dx \leq 2 \int_{\mathbb{R}^n} |x|^2 \rho_0(x) dx + 2d(\rho_0, \rho_1)^2 \quad \text{for all } \rho_0, \rho_1 \in K$$

one can show that (6) is in fact bounded below, and can then construct a minimizing sequence which can be shown to converge.

After establishing uniform bounds on the second moments of the sequence, the negative entropy, and the expectation $\mathbb{E}_{\rho_h^{(N)}}(\Psi)$ one can show that (after possibly passing to a subsequence),

$$\rho_h(t) \rightharpoonup \rho \in K \quad \text{weakly in } L^1((0,T) \times \mathbb{R}^n) \quad \text{for all } T < \infty.$$

One of the more technical aspects of the proof consists in establishing the bound

$$\Big| \int_{\mathbb{R}^n} \{ \frac{1}{h}(\rho^{(k)} - \rho^{(k-1)})\xi + (\nabla \Psi \cdot \nabla \xi - \Delta \xi)\rho^{(k)} \} \; dy \Big|$$
$$\leq \frac{1}{2} \sup_{\mathbb{R}^n} |\nabla^2 \xi| \frac{1}{h} d(\rho^{(k-1)}, \rho^{(k)})^2$$

for all $\xi \in C_0^\infty(\mathbb{R}^n)$, from which it follows that the limiting ρ satisfies the equation (4) in a weak sense. We then proceed via classical regularity theory arguments to establish that ρ is in fact a smooth solution. □

Remark 4. *The interpolation scheme used in Theorem 3 is sometimes called a Minimum Movement Curve; generalities about such curves can be found in [2].*

Remark 5. *An interesting consequence of Theorem 3 comes from taking $\Psi \equiv 0$ and $\beta = 1$, in which case (4) reduces to the standard heat equation. The result then states that the heat equation follows the Wasserstein gradient flow of negative entropy.*

A very similar set of arguments can be used to show the following similar result.

Theorem 6. *Let $\beta : \mathbb{R}^n \to \mathbb{R}$ be a convex function with super linear growth, and suppose*

$$\int_{\mathbb{R}^n} \beta(u^0)\ dx < \infty$$

for some u^0. Define u^{k+1} by the rule (6) with $F(u) := \int_{\mathbb{R}^n} \beta(u)\ dx$. Then using the same interpolation scheme as in Theorem 3 we have

$$u_h \to u \qquad \textit{strongly in } L^1_{loc}(\mathbb{R}^n \times (0,\infty)),$$

where u is a weak solution of the diffusion problem

$$\frac{\partial u}{\partial t} = \Delta\alpha(u), \quad u(0) = u^0$$

with $\alpha(z) = \beta'(z)z - \beta(z)$.

References

[1] Jordan, R., Kinderlehrer D., and Otto, F. *The Variational Formulation of the Fokker–Planck Equation* SIAM journal on mathematical analysis **29.1** (1998), 1-17;

[2] Ambrosio, L., Gigli, N., and Savarae, G., *Gradient Flows.* (2005) Springer.

Chris D. White, UT-Austin
email: **cwhite@math.utexas.edu**

16 On Sobolev Regularity of Mass Transport and Transportation Inequalities

after Alexander V. Kolesnikov [1]
A summary written by Shuangjian Zhang

Abstract

This paper provided Sobolev a priori estimates for optimal transportation $T = \nabla\Phi$ between probability measure $\mu = e^{-V}dx$ and $\nu = e^{-W}dx$ on $\mathbb{R}^d$. With uniform convexity of potential W, it showed that $\int \|D^2\Phi\|_{HS}^2 d\nu$ is bounded above by $\int |\nabla V|^2 d\mu$. Besides, similar estimate for the $L^p(\mu)-$norms and some L^p-generalizations of the well-known Caffarelli contraction theorem were also shown in this paper. At last, it presented some operator norm estimates for $D^2\Phi$.

16.1 Introduction

Let $\mu = e^{-V}dx$ and $\nu = e^{-W}dx$ be probability measures on $\mathbb{R}^d$ and let $T = \nabla\Phi$ be the optimal transportation mapping such that $T_{\#}\mu = \nu$. Assume that W is uniformly convex ($D^2W \geq K \cdot \mathrm{Id}, where K > 0$). It will be shown that

$$\mathcal{I}_\mu := \int |\nabla V|^2 d\mu \geq K \int \|D^2\Phi\|_{HS}^2 d\mu$$

More generally it will be shown that for every unit $e \in \mathbb{R}^d$ and $p \geq 1$

$$\frac{p+1}{2}\|V_e^2\|_{L^p(\mu)} \geq K\|\Phi_{ee}^2\|_{L^p(\mu)}$$

These results can be considered as (global, dimension-free) Sobolev a priori estimates for the following Monge-Ampère equation

$$e^{-V} = e^{-W(\nabla\Phi)} det D^2\Phi$$

In subsection 3, it will be provided that

$$\int (V(x+e) - V(x)) d\mu \geq \frac{K}{2}\int |\nabla\Phi(x+e) - \nabla\Phi(x)|^2 d\mu$$

In Subsection 4, let $\mu = g \cdot \gamma$(with smooth g) and $\nu = \gamma$, then

$$I_\gamma g = 2Ent_\gamma g - 2\int \log \det{}_2(D^2\Phi - \mathrm{Id}) g d\gamma$$
$$+\int \|D^2\Phi - \mathrm{Id}\,\|_{HS}^2 g d\gamma + \sum_{k=1}^{d} \int \mathrm{Tr}[(D^2\Phi)^{-1} D^2\Phi_{x_k}]^2 g d\gamma$$

where $I_\gamma g = \int \frac{|\nabla g|^2}{g} d\gamma$ (relative information), $\det_2(D^2\Phi - \mathrm{Id}) = \det D^2\Phi \cdot \exp(d - \Delta\Phi)$ (the Fredholm-Carleman determinant of $D^2\Phi - \mathrm{Id}$). In particular, this identity implies the following stronger version of log-Sobolev inequality

$$I_\gamma g \geq 2Ent_\gamma g - \int \log \det{}_2(D^2\Phi)^2 g d\gamma$$

and

$$I_\gamma g \geq \int \|D^2\Phi - \mathrm{Id}\,\|_{HS}^2 g d\gamma$$

In addition, the paper prove some dimension-free results for general log-concave reference measure.

In subsection 5, there are several L^p-generalizations of the main result. For every fixed unit vector e and $p \geq 1$ one has

$$K\|\Phi_{ee}^2\|_{L^p(\mu)} \leq \|(V_{ee})_+\|_{L^p(\mu)}$$
$$K\|\Phi_{ee}^2\|_{L^p(\mu)} \leq \frac{p+1}{2}\|V_e^2\|_{L^p(\mu)}$$

Note that the contraction theorem follows from these estimates and this is exactly the case when $p = \infty$.

In Subsection 6, it was proven that

$$K(\int \|D^2\Phi\|^{2p} d\mu)^{\frac{1}{p}} \leq (\int \|(D^2V)_+\|^p d\mu)^{\frac{1}{p}}$$

16.2 Main Result

Theorem 1. *Assume that $\mathcal{I}_\mu < \infty$, μ admits the finite second moment, and W satisfies $D^2W \geq K \cdot \mathrm{Id}$ for some $K > 0$. Then $\Phi \in W^{2,2}(\mu)$ and*

$$\mathcal{I}_\mu \geq K \int \|D^2\Phi\|_{HS}^2 d\mu$$

Remark 2. *Note that some global bounds on the third derivatives of* Φ *are also available. Indeed if* Φ *is sufficiently smooth and*

$$\int V_{x_i}^2 d\mu = \int < D^2\Phi \cdot D^2W(\nabla\Phi) \cdot D^2\Phi \cdot e_i, e_i > d\mu + \int \mathrm{Tr}([(D^2\Phi)^{-1} D^2\Phi_{x_i}]^2) d\mu$$

holds, then

$$\int |\nabla V|^2 d\mu \geq 2\sqrt{K} \int [\sum_{i=1}^{d} \|D^2\Phi_{x_i}\|_{HS}^2]^{\frac{1}{2}} d\mu$$

Remark 3. *Some results of triangular mappings were also provided. For* $T = (T_1(x_1), T_2(x_1, x_2), \cdots, T_d(x_1, \cdots, x_d))$*, where every* T_i *is increasing in* x_i*, one has*

$$\int V_{x_i}^2 d\mu = \int < D^2W(T) \cdot \partial_{x_i}T, \partial_{x_i}T > d\mu + \sum_{k=i}^{d} \int (\frac{\partial_{x_i x_k} T_k}{\partial_{x_k} T_k})^2 d\mu$$

$$\mathcal{I}_\mu = \int \mathrm{Tr}[DT \cdot D^2W(T) \cdot (DT)^*] d\mu + \sum_{k=1}^{d} \int |\nabla \ln \partial_{x_k} T_k|^2 d\mu$$

16.3 Transportation Inequalities

Remark 4. *The Talagrand inequality*

$$\int \rho \log \rho d\nu \geq \frac{K}{2} \int |T(x) - x|^2 \rho d\nu$$

holds for any reasonable transportation mapping T *sending* $\rho \cdot \nu$ *onto* ν *and satisfying*

$$div(T^{-1}) - d - \log\det D(T^{-1}) \geq 0$$

Let $f \cdot \nu$, $g \cdot \nu$ *be probability measures,* $\nu = e^{-W}dx$ *with* $D^2W \geq K \cdot \mathrm{Id}, K > 0$*. Let* $T_f(T_g)$ *be the optimal transportation mapping pushing forward* $f \cdot \nu(g \cdot \nu)$ *onto* ν*. With setting*

$$\rho = \frac{f}{g} \circ (T_g^{-1}), T = T_f \circ T_g^{-1},$$

one can get the following inequality

$$\int f \log \frac{f}{g} d\nu \geq \frac{K}{2} \int |T_f - T_g|^2 f d\nu$$

from Talagrand inequality. Let $f(x) = e^{-V(x)+W(x)}$ *and* $g(x) = e^{-V(x+e)+W(x)}$ *(here* e *is a fixed vector), then* $T_f = \nabla\Phi$ *and* $T_g = \nabla\Phi(x+e)$*. Thus, one can obtain*

$$\int (V(x+e) - V(x))d\mu \geq \frac{K}{2}\int |\nabla\Phi(x+e) - \nabla\Phi(x)|^2 d\mu$$

Lemma 5. *Let* $\phi : A \to \mathbb{R}$*,* $\psi : B \to \mathbb{R}$ *be convex functions on convex sets* A, B*. Assume that* $\nabla\psi(B) \subset A$*. Then*

$$div(\nabla\phi \circ \nabla\psi) \geq \mathrm{Tr}[D_a^2\phi(\nabla\psi) \cdot D_a^2(\psi)]dx \geq 0$$

where div is the distributional derivative.

Theorem 6. *Assume that* W *is* K*-uniformly convex. Then for every* $e \in \mathbb{R}^d$

$$\int (V(x+e) - V(x))d\mu \geq \frac{K}{2}\int |\nabla\Phi(x+e) - \nabla\Phi(x)|^2 d\mu$$

Proposition 7. *The inequality*

$$\int (V(x+e) - V(x))d\mu \geq \frac{K}{2}\int |\nabla\Phi(x+e) - \nabla\Phi(x)|^2 d\mu$$

implies inequality

$$\int |\nabla V|^2 d\mu \geq K\int \|D^2\Phi\|_{HS}^2 d\mu$$

16.4 Dimension-free Inequalities

16.4.1 Gaussian case

Let γ be the standard Gaussian measure on $\mathbb{R}^d$, let $\mu = g \cdot \gamma$, $\nu = \gamma$ and $\nabla\Phi$ be the corresponding optimal transport, where g is smooth, bounded, strictly positive, $I_\gamma g < \infty$ and $-D^2 \log g \leq c \cdot \mathrm{Id}$. Then one has

$$I_\gamma g = 2Ent_\gamma g - 2\int \log\det{}_2(D^2\Phi - \mathrm{Id})g d\gamma$$

$$+\int \|D^2\Phi - \mathrm{Id}\,\|_{HS}^2 g d\gamma + \sum_{k=1}^{d}\int \mathrm{Tr}[(D^2\Phi)^{-1}D^2\Phi_{x_k}]^2 g d\gamma$$

16.4.2 Log-concave case

Theorem 8. *Let $\mu = ge^{-W}dx$, $\nu = e^{-W}dx$. Assume that for some $K > 0$, $D^2W \geq K \cdot \mathrm{Id}$ and*

$$W(x) - < \nabla W(y), x - y > - W(y) \geq \frac{K}{2} |\nabla W(x) - \nabla W(y)|^2$$

Then

$$\frac{K}{2} \int \| D^2\Phi - \mathrm{Id} \|_{HS}^2 g d\mu \leq \frac{2}{K} \int g \log g d\mu + \int \frac{|\nabla g|^2}{g} d\mu$$

In particular, the estimate holds for some $K > 0$ if $C_1 \cdot \mathrm{Id} \leq D^2W \leq C_2 \cdot \mathrm{Id}$.

16.5 the Caffarelli's Theorem

Theorem 9. *Assume that $D^2W \geq K \cdot \mathrm{Id}$. Then for every unit vector e, $p \geq 0$, and $r = \frac{p+2}{2}$, one has*

$$K \| \Phi_{ee}^2 \|_{L^r(\mu)} \leq \| (V_{ee})_+ \|_{L^r(\mu)}$$
$$K \| \Phi_{ee}^2 \|_{L^r(\mu)} \leq \frac{p+4}{4} \| V_e^2 \|_{L^r(\mu)}$$

Corollary 10. *In the limit $p \to \infty$ we obtain the contraction theorem of Caffarelli*

$$K \| \Phi_{ee} \|_{L^\infty(\mu)}^2 \leq \| (V_{ee})_+ \|_{L^\infty(\mu)}$$

16.6 Operator Norm Estimates

Lemma 11. *Assume that Φ is smooth. Then for every smooth vector field v and every nonnegative test function η the following inequality holds*

$$\int < D^2 V v, v > \eta d\mu \geq K \int \| D^2\Phi \cdot v \|^2 \eta d\mu + \int < (D^2\Phi)_v \cdot v, (D^2\Phi)^{-1} \nabla \eta > d\mu$$
$$+ 2 \int \mathrm{Tr}((D^2\Phi)_v \cdot Dv \cdot (D^2\Phi)^{-1}) \eta d\mu + \int \mathrm{Tr}([(D^2\Phi)^{-1}(D^2\Phi)_v]^2) \eta d\mu$$

Lemma 12. *Assume that Φ is convex and twice continuously differentiable. For every $\varepsilon > 0$ there exists a matrix $Q_\varepsilon \geq 0$ such that $\| Q_\varepsilon \| \leq \varepsilon$ and $D^2\Phi + Q_\varepsilon$ has no multiple eigenvalues almost everywhere.*

Theorem 13. *Assume that* $D^2W \geq K \cdot \mathrm{Id}$ *and* $(D^2V)_+ \in L^1(\mu)$*. Then the following inequality holds*

$$\int \|(D^2V)_+\| d\mu \geq K \int \|D^2\Phi\|^2 d\mu$$

Theorem 14. *Assume that* $D^2W \geq K \cdot \mathrm{Id}$*. Then for every* $r \geq 1$ *one has*

$$K(\int \|D^2\Phi\|^{2r} d\mu)^{\frac{1}{r}} \leq (\int \|(D^2V)_+\|^r d\mu)^{\frac{1}{r}}$$

References

[1] Kolesnikov, A. V. , *On Sobolev regularity of mass transport and transportation inequalities.* In:arXiv preprint arXiv:1007.1103 (2010).

Shuangjian Zhang, Toronto
email: **szhang@math.toronto.edu**