1 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.

1.1 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 1.3,1.4)

1.2 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.

1.3 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 $T'(x)$.

1.4 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*

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

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**