# Optimal transport: introduction/panorama

(5/9-2013)

Robert Berman

**Main references:**

- [A-G][ Ambrosio, L; Gigli, N: "A user's guide to optimal transport" (2011) (http://cvgmt.sns.it/paper/195/)

- [V] Villani, "Topics in optimal transportation", AMS (2003)

We will cover parts of

A-G: Chapter 1-4 (80 pages)

V: Chapter 1,2,4,5,6,9

[In principle, will primarily use A-G and then V for further background]

**Further useful references:**

- Villani, "Optimal transp ort, old and new" Springer (2008)

- Ambrosio, L; Gigli, N; Savare, G: "Gradient Flows in Metric spaces and in the space of Probability Measures", Birkhasuer 2008 (second edition)

# Out-line of the main topics to be covered in the course

1. **Transportation theory**

2. **Relations to Monge-Ampère equations**

3. **The 2-Wasserstein space $W_2$**

4. **Otto calculus on $W_2$ and applications to dissipative PDEs**

5. **Applications to geometric functional inequalities**

# 1. Transportation theory

Monge's formulation (1781)

*In mathematics terms:* assume given two measures on $\mathbb{R}^n$ : $\mu$ and $\nu$ with the same total mass, normalized to be *one*, i.e.

$$\mu, \nu \in \mathcal{P}(\mathbb{R}^n) := \{probability \ \ measures \ \ on \, \mathbb{R}^n\}$$

and a *"cost function"* $c(x, p)$ on $\mathbb{R}^n \times \mathbb{R}^n$.

Definition: A *transport map* $T$ from $\mu$ to $\nu$ is a map such that

$$T : \ \mathbb{R}^n \to \mathbb{R}^n, \qquad T_* \mu = \nu,$$

i.e.

$$\int_B d\nu = \int_{T^{-1}(B)} d\mu$$

The *(total) cost* of $T$ is

$$C(T) := \int_{\mathbb{R}^n} c(x, T(x)) d\mu$$

A transport map $T$ is *optimal* (wrt $\mu, \nu$ and $c(x, y)$) if it minimizes the cost $C(T)$ over all maps transporting $\mu$ to $\nu$.

## The Monge Problem

- Prove the existence and uniqueness of an optimal map $T$

- "Describe" the optimal map $T$

## The standard setting

The most well-behaved case is the "quadratic case":

$$c(x, y) = \|x - y\|^2$$

(but Monge was, in fact, mainly interested in the much more difficult case $c(x, y) = \|x - y\| ...$)

General difficulties:

- *Non-linearity:* the cost functional

$$C(T) := \int_{\mathbb{R}^n} c(x, T(x)) d\mu$$

  is non-linear wrt $T$.

- *Non-compactness:* If $T_i$ is a sequence of transport maps, then $T_i$ may not converge to a transport map.

*Solution:* "Relaxation" of the Monge problem, i.e. enlarge $\{T\}$!

- Heuristically: allow "splitting of mass"

*Def:* A *transport plan* $\gamma$ is a probability measure on $\mathbb{R}^n \times \mathbb{R}^n$ whose first and second marginals are equal to $\mu$ and $\nu$ :

- $\gamma \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$
- $(\pi_1)_* \gamma = \mu$
- $(\pi_2)_* \gamma = \nu$

Then define the cost functional $C(\gamma) := \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y) d\gamma$

**Example:** If $T$ is a map transporting $\mu$ to $\nu$, i.e. $T_*\mu = \nu$, then

$$\gamma_T := (I \times T)_*\mu$$

is a transport plan (from $\mu$ to $\nu$).

- Note: the measure $\gamma_T$ is supported on the graph of $T$

## The Kantorovich ('42) problem

Given probability measures $\mu, \nu$ and a const function $c(x, y)$ find an *optimal transport plan* $\gamma$, i.e. a minimizer of

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

Advantage:

- $C(\gamma)$ is linear wrt $\gamma$
- "Compactness" (by weak compactness/tightness)

For any reasonable cost function $c(x, y)$ the existence of an optimal $\gamma$ is then easy.

*More good news*:

- There is a *dual* formulation of the Kantorovich problem

$$\gamma \longleftrightarrow (\phi, \psi)$$

  where $\phi(x)$ and $\psi(y)$ are *functions* on $X$.

- This leads to a variational formulation involving only a *c-convex* function $\phi(x)$ (using Legendre transforms).

*The new "regularity problem":*

Show that under suitable regularity assumptions on the data ($\mu, \nu$ and $c(x, y)$) the optimal transport plan $\gamma$ is realized by a *transport map* $T$, i.e. $\gamma = \gamma_T$.

## 2) Relations to Monge-Ampère equations

In the following we specialize to the "standard case":

$$c(x,y) = \|x-y\|^2$$

with $\mu$ and $\nu$ assumed "regular", say

$$\mu = f(x)dx, \quad \nu = g(y)dy, \quad f,g \in \mathcal{C}^\infty(\mathbb{R}^n)$$

One can then a priori show [Brenier'87, Caffarelli,...] that an optimal map exists and is uniquely determined by

$$T(x) = \nabla\phi(x): \quad \mathbb{R}^n \to \mathbb{R}^n$$

for a convex function $\phi$ on $\mathbb{R}^n$.

For $T = \nabla\phi(x)$ with $\phi$ smooth and strictly convex it is easy to see that

$$T_*(f(x)dx) = g(y)dy$$

if and only if

$$\det(\frac{\partial^2\phi}{\partial x_i \partial x_j})g(\nabla\phi) = f(x),$$

i.e. $\phi$ solves a *Monge-Ampère equation.*

Ex: prove this!

The Monge-Ampère operator

$$\phi \mapsto \det(\frac{\partial^2 \phi}{\partial x_i \partial x_j})$$

is a propotype of a fully non-linear partial differential operator

- Reversing the story: the transport problem with $c(x, y) = \|x - y\|^2$ thus provides a variational approach to real Monge-Ampère equations

- This can be related to more standard Dirichlet type variational principles *[B., "Statistical mechanics of permanents, real-Monge-Ampere equations and optimal transport", arXiv:1302.4045]*

## 3) The 2-Wasserstein space

The space $\mathbb{R}^n$ comes with a standard *metric*, the Euclidean one:

$$d(x, y) := \|x - y\|$$

It induces a metric on the infinite dimensional space $\mathcal{P}(\mathbb{R}^n)$ of all probability measures on $\mathbb{R}^n$ :

$$d_{W_2}(\mu, \nu) := \text{"the minimal cost to transport } \mu \text{to } \nu\text{"}$$

(defined in terms of the standard cost function $c(x, y) = \|x - y\|^2$), i.e.

$$d_{W_2}(\mu, \nu) = \inf_{\gamma} C(\gamma) (= C(\gamma_{optimal}))$$

over all transport plans $\gamma$ from $\mu$ to $\nu$.

More precisely, the Wasserstein 2-metric $d_{W_2}$ is a well-defined metric on the subspace

$$\mathcal{P}_2(\mathbb{R}^n) := \left\{ \mu \in \mathcal{P}(\mathbb{R}^n) : \quad \int |x|^2 d\mu < \infty \right\}$$

What about the geometric structure of the Wasserstein space $W_2 := (\mathcal{P}_2(\mathbb{R}^n), d_{W_2})$?

- It has *positive curvature* (in the sense of Alexandrov)

How to describe the geodesics in $W_2$ (i.e. curves $\mu_t$ with minimal length)?

They can be described in terms of optimal transport maps:

- The *geodesic* connecting ("regular") measures $\mu_0$ and $\mu_1$ is given by

$$\mu_t := ((1-t)I + tT)_* \mu_0,$$

where $T$ is the optimal transport map from $\mu_0$ to $\mu_1$

## 4) Otto calculus on $W_2$ and applications to dissipative PDEs

*Otto's key observation:* $d^2_{W_2}$ comes from a *Riemannian metric* on $\mathcal{P}_2(\mathbb{R}^n)$.

- Hence, we can "do calculs" on $\mathcal{P}_2(\mathbb{R}^n)$!

The definition of the length of the tangent vector of the curve

$$\mu_t = \rho_t dx$$

at $t = 0$, i.e.

$$\left\| \frac{\partial \rho_t}{\partial t} | t=0 \right\|_{W_2}$$

is inspired by fluid mechanics/kinetic theory:

Think of the changing density $\rho_t$ as induced by a velocity field $\overrightarrow{v_t}$

In mathematical terms:

$$\frac{\partial \rho_t}{\partial t}|_{t=0} = -\nabla \cdot (\rho_0 \overrightarrow{v}) \quad \text{(the continuity equation)}$$

for a vector field $\overrightarrow{v}$ (uniquely determined by $\nabla \times \overrightarrow{v} = 0$, i.e. $\overrightarrow{v} = \nabla\phi$ for some function $\phi$)

- Then define

$$\left\| \frac{\partial \rho_t}{\partial t}|_{t=0} \right\|_{W_2}^2 := \int_{\mathbb{R}^n} \|\overrightarrow{v}\|^2 \rho_0 dx$$

(=the total *kinetic energy* of the fluid)

## Applications to dissipative PDEs

Given a functional $F$ on the space $\mathcal{P}(\mathbb{R}^n)$ we can study its *gradient-flow* with respect to $d_{W_2}$ :

$$\frac{\partial \rho_t}{\partial t} = -\nabla F_{|\rho_t}, \quad \rho_{|t} = \rho_0$$

By construction the functional $F$ is mototone (decreasing) along the gradient-flow $\rho_t$.

- Many interesting "dissipative" evolution PDEs can be realized in this way

**Example1:** The gradient flow of the functional

$$F(\rho) = \int (\log \rho)\rho dx$$

(=-"Boltzmann entropy") gives the heat (diffusion) equation

$$\frac{\partial \rho_t}{\partial t} = \Delta \rho_t$$

Recall: it may be realized using Brownian motion $B_t$ on $\mathbb{R}^n$ (i.e. $\rho_t$ is the law of $B_t$)

- The Wasserstein point of view gives a useful bridge to probability ("interacting particle systems").

**Example2:** Given a (suitable) function $V$ on $\mathbb{R}^n$ the gradient flow of the functional

$$F(\rho) = \int (\log \rho)\rho dx + \int V \rho dx$$

is the *linear Fokker-Planck equation:*

$$\frac{\partial \rho_t}{\partial t} = \Delta \rho_t + \nabla \cdot (\rho_t \nabla V)$$

- *Probabilistic interpretation:* $\rho_t$ is the law of the stochastic process $X_t$ on $\mathbb{R}^n$ satisfying the SDO

$$\frac{dX_t}{dt} = -\nabla V + \frac{dB_t}{dt}$$

(=stochastic gradient flow of $V$ on Euclidean $\mathbb{R}^n$)

**Example3:** Assume given a (suitable) function $W$ on $\mathbb{R}^n$. Then the gradient flow of the functional

$$F(\rho) = \int (\log \rho)\rho dx + \int\int W(x-y)\rho(x)\rho(y)dxdy$$

is the following *non-linear* Fokker-Planck equation

$$\frac{\partial \rho_t}{\partial t} = \Delta \rho_t + \nabla \cdot (\rho_t \nabla V_t)$$

$$V_t(x) = \int W(x-y)\rho(y)dy$$

(this evolution equation appears naturally in chemotaxis, astrophysics, swarm aggregation,...)

The gradient flow formulation of the PDEs is also useful for proving *existence* and *uniqueness* of (weak) solutions.

- The starting point is a variational formulation of gradient-flows going back to De Georgi's "minimizing movements"

- Also useful from a *numerical* point of view (the method is based on a varational version of Euler's method)

## 5) Applications to geometric functional inequalities

Optimal transport techniques have led to spectacularly transperant proofs of a range of geometric inequalities such as

- The Brunn-Minskowski inequality
- Isoperimetric inequalities
- (log) Sobolev inequalities

In a nut shell the proofs follow one of the following three different techniques:

- Use a transport map to "linearize" the problem
- Use *convexity* arguments on the Wasserstein 2-space
- Use a *gradient flow* on the Wasserstein 2-space