Optimal Transport

Basically follow the book Topics in Optimal Transportation
by C. Villani

non-Euclidean geometry

synthetic geometry
(Lott, Sturm, Villani)

non-smooth gradient flow

metric analysis

Aubry-Mather (Bernard, Fathi, Gangbo ...)

Fully nonlinear PDEs (Ma, Trudinger, Wang, Loeper ..)
(regularity theory of OT)

Numerical Methods

Preface of the 1st Edition

Optimal Transport: born in France 1781
as a research topic
Monge

Classical subject in probability theory, economics and Optimization

1987' Brenier: Polar decomposition
(interplay between PDE, fluid mechanics, geometry, ...)

Otto Calculus

Mean-Field limit in Statistical Physics

(importance of handling mass transport on ∞-dimensional space such as the Wiener space or the space of probability measures on some phase space.)

Dobrushin et al. ...

Basic ideas: Kantorovich duality
metric properties induced by optimal transport

Notation:
"Small set" in $\mathbb{R}^n$ means: Hausdorff dim $\leq n-1$

Given time, cover Numerical Methods for OT.

Brenier, Otto, Tanaka's work on Boltzmann eq.

Notations and overviews

- The identity map Id
- X a set, write $1_X(x) = \begin{cases} 1, & \text{if } x \in X; \\ 0, & \text{otherwise} \end{cases}$
- A^c: the complement of the set A
- $\mathbb{R}^n$, $n \geq 1$; $|A|_n$: n-dim Lebesgue measure if A is Lebesgue measurable

  $x \in \mathbb{R}^n$, $|x| = \sqrt{x \cdot x}$

  $$x \cdot y = \sum_{i=1}^{n} x_i y_i$$

- X: an abstract measure space

  $\mathcal{P}(X) \subseteq \underbrace{M(X)}$ finite signed measure

  $\mu \in M(X)$: $\|\mu\|_{TV} = \mu_+[X] + \mu_-[X]$.

  $\mu = \mu_+ - \mu_-$, μ_+, μ_- singular to each other

  ($\nu \in M_+(X)$, f measurable, $\Rightarrow \|f\|_{L^1(d\nu)} = \|f\nu\|_{TV}$)

  $(\mathcal{P}(X), \sigma(X))$ $\sigma(X)$: Borel σ-algebra

  $w^*-\mathcal{P}(X)$: $\mathcal{P}(X)$ equipped with the weak topology

$$\delta_x[A] = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise} \end{cases}$$

dirac mass at x, a measure

μ : measure on X. $p \geq 1$. $L^p(X)$ or $L^p(X, d\mu)$ or $L^p(d\mu)$

i.e. L^p space with reference measure ρ.

$\frac{1}{p} + \frac{1}{p'} = 1$, p' the conjugate of p

$T: X \to Y. \rightsquigarrow T: (X, \mu) \to (Y, T_\# \mu)$

$T_\# \mu$: push-forward measure of μ

$(T_\# \mu)[B] = \mu[T^{-1}(B)]$

preimage of B

$S(X) = \{T: X \circlearrowleft \mid T_\# \mu = \mu\}$: all maps T that keep μ invariant

Also use

$T_\# f = g$, where f, g are two density functions.

- X : topological space, equipped with $\mathcal{B}(X)$ Borel σ-algebra

$C(X)$, $C_b(X)$, $C_0(X)$, $C_c(X)$

$C(X;\mathbb{R})$ ...

$C_b(X)$ equipped with a natural norm,

$$\|u\|_\infty = \sup_{x\in X} |u(x)|$$

$A \subset X$, $\mathrm{Int}(A) = \mathring{A}$: inner part of A

$$\overline{A} = A \cup \partial A$$

support of a measure μ on X

= smallest closed set $F \subset X$, s.t. $\mu[X \setminus F] = 0$

= $\mathrm{Supp}\,\mu$

μ is concentrated on $A \subset X \iff \mu[X \setminus A] = 0$.

X : metric space $x \in X$, $B(x,r)$: a ball with radius r. and center x.

$\mathrm{Lip}(X)$: all Lipschitz functions on X.

$$P_p(X) = \{ \underset{\text{Borel}}{\mu \in P(X)} \mid \int_X d(x_0, x)^p \, d\mu(x) < \infty \}.$$
for some $x_0 \in X$.

X : Banach space, X^* : topological dual

$\langle \cdot, \cdot \rangle$ duality between X and X^*

$\varphi : X \to \mathbb{R}$ (or $\mathbb{C}$). convex

φ^* its dual function (Legendre - Fenchel)

$\partial\varphi$: subdifferential of φ

identified with its graph

X: smooth Riemannian manifold, $F : X \to \mathbb{R}$

(or general Banach) continuous.

DF: differential map $DF(x) \cdot v$: 1st order variation of

F at point $x \in X$, along the direction v

X: Riemanian manifold. $T_x X$, $\langle \cdot, \cdot \rangle_x$

$\mathcal{D}(X)$: C^∞ functions on X with compact support

$\mathcal{D}'(X)$: the space of distribution on X.

The gradient operator ∇ on $\mathcal{D}(X)$ by:

$$\langle \underset{\substack{\in \\ (T_x X)^*}}{\nabla F(x)}, v \rangle_x = \underbrace{DF(x) \cdot v}_{\text{Linear map}}$$

Linear map $\Rightarrow$ A representation.

$\nabla \cdot$: the divergence operator,

the adjoint of ∇

Laplace operator: $\Delta F = \nabla \cdot \nabla F$

If $X = \mathbb{R}^n$, then

$$\nabla F = \left(\frac{\partial F}{\partial x_1}, \cdots, \frac{\partial F}{\partial x_n}\right),$$

$$\nabla \cdot u = \sum_{i=1}^{n} \frac{\partial u_i}{\partial x_i} \qquad \Delta F = \sum_{i=1}^{n} \frac{\partial^2 F}{\partial x_i^2}$$

D^2: Hessian operator on X

(∇^2) The Euclidean case $(X = \mathbb{R}^n)$,

$$D^2 F(x) = \left(\frac{\partial^2 F(x)}{\partial x_i \partial x_j}\right)$$

$$\mathcal{P}_{ac}(\mathbb{R}^n) = \{ \mu \in \mathcal{P}(\mathbb{R}^n) \mid \mu \ll \mathrm{Leb}^n \}$$

subset in $L^1(\mathbb{R}^n)$

$$\mathcal{P}_{ac,2}(\mathbb{R}^n) = \mathcal{P}_{ac}(\mathbb{R}^n) \cap \{ \mu \mid \int |x|^2 d\mu(x) < \infty \}$$

finite 2-moment.

The Aleksandrov Hessian of a convex function φ on $\mathbb{R}^n$ will be denoted by $D_A^2 \varphi$: defined a.e. $\mathrm{int}(D(\varphi))$

(This is not $D_D^2 \varphi$, the distributional Hessian of φ.)

$\det_H D^2 \varphi$, $\mathrm{Trace}(D_A^2 \varphi) = \Delta_A \varphi$

$\mathrm{Trace}(D^2_{D'}\varphi) = \Delta_{D'}\varphi.$

$C^k(\Omega)$, k-integer $\qquad C^{k,\alpha}(\Omega)$, $\alpha\in(0,1)$

$M_n(\mathbb{R})$: all $n\times n$ matrices,

$\mathrm{tr}\, M$: trace of M.

I_n : $n\times n$ identity matrix; $\qquad M, M^T$ transpose

M : symmetric if $M = M^T$

$M \geq 0 \Leftrightarrow M$ symmetric with nonnegative eigenvalues

anti-symmetric if $M^T = -M$.

orthogonal: $MM^T = M^T M = I_n$

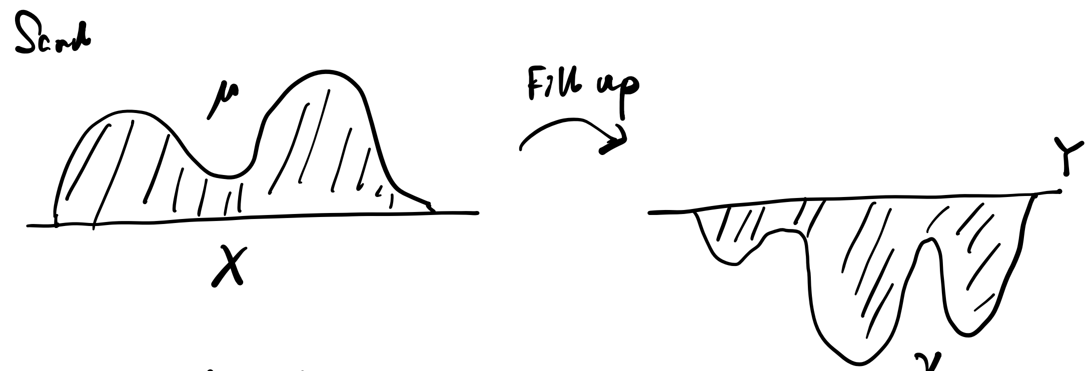
$S_n(\mathbb{R})$ (symmetric), $S_n^+(\mathbb{R})$ (symmetric + positive)

$A_n(\mathbb{R})$ (anti-symmetric) $\quad O_n(\mathbb{R})$: (orthogonal)

$\boxed{\text{We shall skim all above}}$

Lecture 1 : Introduction

1. Formulation of Optimal Transportation Problem.



$\int_X d\mu = \int_Y d\nu = 1$, both are probability measures

$A \subset X$, $B \subset Y$ > measurable

$\mu[A]$: how much sand in A

$\nu[A]$: how much sand can be piled in B

(moving) cost function : $c: X \times Y \to \mathbb{R}$

$c(x,y)$: how much it costs to transport one unit of mass from location x to location y

Assume c is measurable and ≥ 0.

In general : $c: X \times Y \to \mathbb{R}_+ \cup \{+\infty\}$

some $c(x,y)$ can take infinite values

Problem: Realize the transportation at minimal cost?

Transport map: $T: X \to Y$

More general, transference plan by prob. meas π on $X \times Y$.

$d\pi(x,y)$: # of mass transported from x to y

(It is possible that some mass located at x may be splitted into severla parts.

$\mu = \delta_0$, $\nu = \frac{1}{2}(\delta_1 + \delta_{-1})$)

Admissable plan: $\pi \in P(X \times Y)$ shall satisfy

$$\begin{cases} \int_Y d\pi(x,y) = d\mu(x) & \text{(all mass taken from } x \text{ conincide with } d\mu(x)) \\ \int_X d\pi(x,y) = d\nu(y) & \text{(all mass transferred to } y \text{ coincide with } \nu(y)) \end{cases}$$

or

$$(1) \begin{cases} \pi[A \times Y] = \mu[A], \\ \pi[X \times B] = \nu[B], \end{cases} \quad \forall A \subset X, \ B \subset Y \text{ measurable}$$

Or for all Borel measurable functions $\varphi = \varphi(x)$, $\psi = \psi(y)$

(2) $$\int_{X\times Y} (\varphi(x) + \psi(y))\, d\pi(x,y) = \int_X \varphi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y)$$

Def: We call that the probability measure $\pi \in P(X\times Y)$ have marginals μ and ν if (1) holds true.

$$\Pi(\mu,\nu) \triangleq \left\{ \pi \in P(X\times Y) \;\middle|\; \begin{array}{l} \pi[A\times Y] = \mu(A), \\ \pi[X\times B] = \nu(B), \end{array} A, B \text{ measurable} \right\}$$

- $\Pi(\mu,\nu) \neq \phi$

ex: the tensor product $\mu \otimes \nu \in \Pi(\mu,\nu)$

for each x, the mass $d\mu(x)$ is distributed according to $\nu(y)$.

<u>Kantorovich's formulation</u>

$$\text{Min } I[\pi] = \int_{X\times Y} c(x,y)\, d\pi(x,y), \quad \text{for } \pi \in \Pi(\mu,\nu)$$

(Nobel Prize in Economics: Linear Programming)

Economics: μ: a density of production units (生产者)

ν : - - of consumers

$I[\pi]$ = total transport cost

Optimal transport cost = $\inf_{\pi \in \Pi(\mu,\nu)} I[\pi]$

$\|$

$\mathcal{T}_c(\mu,\nu)$

Basic questions: Existence and Uniqueness

further: characterization of the (possible) optimal transport plans

<u>Probabilistic interpretations:</u>

$\boxed{\text{Coupling}}$

$$\boxed{\min_{\substack{(U,V) \\ \mathrm{Law}(U)=\mu\in\mathcal{P}(X) \\ \mathrm{Law}(V)=\nu\in\mathcal{P}(Y)}} I(U,V) \triangleq \mathbb{E}[c(U,V)]}$$

Recall: r.v. U in X is a measurable map with values in X, $\left(U : (\Omega, \mathbb{P}) \to (X, \underbrace{U_\# P}_{=\mu})\right)$

$\left(A \in \mathcal{B}(X), \ \mu(A) = \mathbb{P}[U^{-1}(A)] \right)$

$$U : (\Omega, \mathbb{P}) \to (X, \underbrace{U_\# \mathbb{P}}_{= \mu})$$

$$V : (\Omega; \mathbb{P}) \to (Y, \underbrace{V_\# \mathbb{P}}_{= \nu})$$

$\pi = \mathrm{Law}(U, V)$ ← joint law

(U, V): a coupling between U, V.

$\pi = \mathrm{Law}(U, V) = \mu \otimes \nu$ if U and V are independent

Kantorovich's problem = relaxed version of Monge's formulation of Optimal transport

$T : X \to Y$ measurable. $\quad T_\# \mu = \nu$

$$d\pi(x, y) = \underbrace{d\mu(x)\, \delta[y = T(x)]}_{\text{this}}$$

$$\pi = \underbrace{(Id \times T)_\# \mu}_{= \pi_T}$$

or in the probabilistic notation,

r.v. $V = T(U)$

- $\pi_T = (Id \times T)_\# \mu$ satisfies:

$$\int_{X \times Y} \varphi(x, y)\, d\pi_T(x, y) = \int_X \varphi(x, T(x))\, d\mu(x),$$

$\forall \varphi \geq 0$, measurable on $X \times Y$.

total transport cost $I[\pi_T]$

$$= \int_X c(x, T(x))\, d\mu(x)$$

$$\boxed{\text{Monge:} \quad \min_{\substack{T, \\ T_\#\mu = \nu}} \int_X c(x, T(x))\, d\mu(x)}$$

$$\Big(T_\#\mu = \nu \Leftarrow \int_{X\times Y} (\varphi(x) + \psi(y))\, d\pi_T(x,y) = \int_X \varphi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y)$$

$$\Updownarrow$$

$$\int_X (\varphi(x) + \psi(T(x)))\, d\mu(x) = \cdots$$

$$\Updownarrow$$

$$\int_X \psi(T(x))\, d\mu(x) = \int_Y \psi(y)\, d\nu(y)$$

$$\| \qquad\qquad \mathbin{/\!/}$$

$$\int_Y \psi(y)\, d\, T_\#\mu(y) \quad \Big)$$

$\nu = T_\#\mu$: ν is the push-forward of μ by T.

or T transports μ to ν.

Law of r.v. U $=$ $U_\#\mathbb{P}$

Ex: (Dirac mass) $\nu = \delta_a \Rightarrow \Pi(\mu, \nu) = \mu \otimes \delta_a$

All the mass should transported to a.

$$T_c(\mu, \delta_a) = \int_X c(x, a)\, d\mu(x).$$

Ex: (Discrete Cases)

Suppose $X = Y = \mathbb{R}^d$,

$$\mu = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \qquad \nu = \frac{1}{n}\sum_{j=1}^n \delta_{y_j},$$

now any measure in $\Pi(\mu,\nu)$ can be identified as a bistochastic $n \times n$ matrix $\pi = (\pi_{ij})_{i,j}$,

means: $\pi_{ij} \geq 0 \quad \forall i,j$

$$\underbrace{\sum_i \pi_{ij} = 1}_{j\text{-th column}}, \quad \forall j; \qquad \underbrace{\sum_j \pi_{ij} = 1}_{\text{the sum of } i\text{-th row}}, \quad \forall i$$

Kantorovich problem now reads:

$$(*) \quad \inf\left\{ \frac{1}{n}\sum_{i,j} \pi_{ij}\, c(x_i, x_j) \,;\ \pi - \text{bistochastic} \right\}$$

B_n : $n \times n$ bistochastic matric

$B_n \subset M_n(\mathbb{R})$, B_n bounded, convex

Choquet's minimization theorem :

(*) has solutions which are extremal points of B_n

(not a nontrivial convex combination of two points in B_n)

Birkhoff's theorem $\Rightarrow$ those extremal points in B_n

are permutation matrices. all permutations of $\{1, 2, \cdots n\}$

i.e. $\pi_{ij} = \underbrace{\delta_{j,\sigma(i)}}_{\text{Kronecker symbol}}$, $\sigma \in S_n$

$$\Downarrow$$

$$\frac{1}{n} \sum_{i,j} \pi_{ij}\, c(x_i, y_j) = \frac{1}{n} \sum_{i} c(x_i, y_{\sigma(i)}), \quad \sigma \in S_n.$$

Thus, (*) reduces to Monge's formulation

$$\inf \left\{ \frac{1}{n} \sum_{i} c(x_i, y_{\sigma(i)}) \; ; \; \sigma \in S_n \right\}$$

RK: The reasoning fails in continuous setting.

If μ, ν A.C. w.r.t. Lebesgue, $\exists$ extreme points in $\Pi(\mu, \nu)$ which are not concentrated on any graph.

($\Pi(\mu,\nu)$ are convex in $\mathcal{P}(X\times Y)$

$$\lambda \pi_1 + (1-\lambda)\pi_2 \in \Pi(\mu,\nu), \text{ if } \pi_1,\pi_2 \in \Pi(\mu,\nu).$$

How to study this ?)

Overview:

Basic questions:

Q1
- Existence of minimizers of Monge-Kantorovich problem. How to characterize them?

$\Downarrow$

Q2
- What information on μ,ν does the knowledge of the optimal transport cost $T_c(\mu, c)$ bring?

Depends on: $\boxed{X, \; Y, \; c(x,y)}$ regularity of μ and ν

Assume for the moment, $X = Y = \mathbb{R}^n$,

$c(x,y) = |x-y|^p, \quad 0 < p < +\infty,$

μ, ν compactly supported.

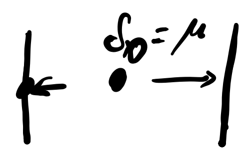
- $p>1$, c is strictly convex.
  If μ, ν are A.C. w.r.t. Lebesgue, then $\exists !$ solution to the Kantorovich problem.
  (μ does not charge sets with finite $(n-1)$-Hausdorff measure.)

- Geometric characterization of optimal maps T :

$p=2$ (dim $n>1$) $\quad T = \nabla\phi$, ϕ - convex

T: monotone and orientation preserving.

- If the source $\mu \in \mathcal{P}(\mathbb{R}^n)$ charge a "small" set (a set of Hausdorff dim $\leq n-1$ in $\mathbb{R}^n$, a point in 1D, a line segment in 2D etc.)

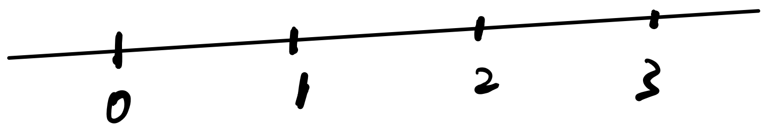


Optimal transport plans in Kantorovich problem have to split mass.

Monge $\neq$ Kantorovich
(no solution even)

- $p=1$, μ, ν AC. Existence of solutions to Monge-Kantorovich but no uniqueness

$\mu = 1_{[0,1]}$ $\qquad \nu = 1_{[2,3]}$



$c(x,y) = |x-y|$

- $p<1$: in general no solution of Monge Problem, except if μ and ν are concentrated on disjoint sets.

- (X, d) Polish space. $c(x,y) = d(x,y)^p$, $p \geq 1$
(complete, separable)

$$T_c^{1/p} = \left\{ \inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times X} d(x,y)^p \, d\pi(x,y) \right\}^{1/p}$$

is a metric on $\mathcal{P}(X)$:

metrizes weak convergence of probability measures

More precisely:

Let $(\mu_k)_{k\geq 1}^{\infty} \subseteq \mathcal{P}(\mathcal{X})$ and for some $x_0 \in X$,

$d(x, x_0)^p \, d\mu_k(x)$ is tight

then $\mu_k \to \mu \iff \mathcal{T}_c(\mu_k, \mu) \to 0.$

Tightness of $\{\mu_k\}$ of non-negative measures on X:

$\forall \varepsilon > 0$, $\exists$ compact set K_ε, s.t. $\sup_k \mu_k[X \setminus K_\varepsilon] \leq \varepsilon$.

- choose cost $c(x, y) = 1_{x \neq y}$

$$\mathcal{T}_c(\mu, \nu) = \frac{1}{2} \|\mu - \nu\|_{TV} \quad \text{total variation}$$

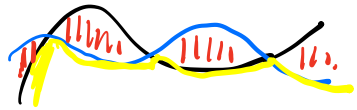
i.e. $\inf \mathbb{E}[X \neq Y] = \sup\{\mu(F) - \nu(F) : F \text{ closed}\}$

$\mathcal{T}_c$ now metrizes the strong topology on $\mathcal{P}(X)$.

Explaination: For $c(x, y) = 1_{x \neq y} = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}$

the optimal map is obtained when all the mass shared by μ and ν does indeed stay in place.

Then



$$\int_{\mu-\nu\geq 0} d(\mu-\nu) = \frac{1}{2}\left[\int_{\mu-\nu\geq 0} d(\mu-\nu) + \int_{\mu-\nu\leq 0} d(\nu-\mu)\right]$$

since $\displaystyle\int_{\mu-\nu=0} d(\mu-\nu) = 0$, $\displaystyle\int_X d(\mu-\nu) = 0$.

Probability: Optimal Transport distance has been used much earlier.

History:

Monge's original problem was extremely difficult.

① Monge's formulation is more tricky than Kantorovich's -- .

② $c(x,y) = |x-y|$ degenerate, from the convexity point of view.

Sudakov (70's) (Mistakes.) pointed by Alberti, Kirchheim and Preiss

Evans and Gangbo: ($c = |x-y|$) (theory of p-Laplace eqs.)

Why Monge's problem is in general tricky?

Assume
$\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, A.C. with

$$d\mu(x) = f(x)\,dx,$$
$$d\nu(y) = g(y)\,dy.$$

Assume that $T: \mathbb{R}^d \to \mathbb{R}^d$ is C^1-diffeomorphism, and $T_\# \mu = \nu$

then

$$\int_{\mathbb{R}^d} \phi(y)\, d\nu(y) = \int_{\mathbb{R}^d} \phi(y)\, d\, T_\#\mu(y)$$

$$\| \qquad\qquad = \int_{\mathbb{R}^d} \phi(T(x))\, f(x)\, dx$$

$$\int_{\mathbb{R}^d} \phi(y)\, g(y)\, dy \qquad \| \ \phi \text{ arbitrary}$$

$$= \int \phi(T(x))\, g(T(x))\, |\det \nabla T(x)|\, dx \quad \Downarrow$$

$$\star \quad \boxed{f(x) = g(T(x))\, |\det \nabla T(x)|}$$

Monge-Ampère Eqs.

highly - nonlinear

Kantorovich's problem can be seen as a relaxed version of Monge's problem.

(i.e. extends the class of objects on which the infimum is taken.

$$\begin{array}{c} (T : X \to Y) \\ \downarrow \mathcal{I} \\ [x \mapsto \delta_{T(x)}] \end{array} \leftarrow L^0(X; P(Y))$$

nonlinear constraint on T $\Rightarrow$ linear constraint on $\mathcal{I}(T)$;

(Young measures ...)

$\boxed{\text{Overview of the textbook}}$

Ch1: Kantorovich duality (Powerful tool, theoretically and numerically)

Very Common Practice

Ch2: Fundamental theory of Optimal Transport:

Existence, Characterization for Kantorovich

Monge's problem: Brenier, Evans, Gangbo, Knott, Smith, McCann, Rachev, Rüschendorf.

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

Ch3: Brenier's Polar factorization theorem

Ch4: Monge-Ampère Eq.

Ch5: Displacement interpolation/convexity

Ch6: Geometric inequalities

Ch7: Wasserstein distance

Ch8: Differential/Dynamical formulation of OT.

(Benamou & Brenier, Otto --)

Ch9: Logarithmic Sobolev inequalities

entropy - entropy production inequality

transport type inequality

Lecture 1 | Kantorovich Duality

Recall some basic notions in Optimal Transport

(X, μ) two probability spaces
(Y, ν)

$c : X \times Y \to \mathbb{R}_+ \cup \{+\infty\}$ measurable
cost function $\qquad c(x,y) = d^p(x,y)$

Kantorovich's formulation of O.T.

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} c(x,y)\, d\pi(x,y),$$

where $\Pi(\mu,\nu) := \{ \pi \in \mathcal{P}(X \times Y) \mid \pi$ has marginals μ on X and ν on $Y \}$

Equivalent characterizations of marginals:

$$(1.1) \quad \begin{aligned} \pi[A \times Y] &= \mu[A], \\ \pi[X \times B] &= \nu[B], \end{aligned} \qquad \forall\, A \subset X,\ B \subset Y \text{ measurable.}$$

Of course, $\pi[X \times Y] = \mu[X] = 1$.
π is a probability measure.

Equivalently,

$\pi \in \Pi(\mu, \nu) \Leftrightarrow (1.1)$

$\Updownarrow$

(1.2) $\pi \in M_+(X \times Y)$ s.t. for all $(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)$,

(nonnegative measures)

$$\int_{X \times Y} (\varphi(x) + \psi(y))\, d\pi(x,y) = \int_X \varphi\, d\mu + \int_Y \psi\, d\nu.$$

RK: Under some top. assumptions of (X, μ) and (Y, ν), one can use a narrower class of test functions

- We only consider:

Borel prob. measures.

$\mathcal{B}(X)$: Borel σ-algebra

$\mathcal{P}(X)$: the set of Borel Prob. measures on X.

When X and Y are Polish spaces (complete, separable metric space), and μ and ν are Borel, it is sufficient to impose (1.2) for $(\varphi, \psi) \in C_b(X) \times C_b(Y)$ only.

(Further X and Y are locally compact ($\mathbb{R}^d$ for ex.) (each point admits a compact nbhd), it is ok to impose $(\varphi, \psi) \in C_0(X) \times C_0(Y)$ . . .)

Check the exe. 1.1 & 1.2.

Duality:

Linear minimization problem with convex constraints, like (Kantorovich's formulation of OT.), admits a dual formulation.

OT's setting: Kantorovich 1942.
where he considered $c(x,y) = d(x,y)$ distance

But his duality theorem holds in much more general setting.

Thm 1.3 (Kantorovich duality)
Let X and Y be Polish spaces.

Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, and

$c : X \times Y \to \mathbb{R}_+ \cup \{+\infty\}$ be l.s.c. cost funt.

Whenever $\pi \in \mathcal{P}(X \times Y)$ and $(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)$, define

$$I[\pi] = \int_{X \times Y} c(x, y)\, d\pi(x, y),$$

$$J(\varphi, \psi) = \int_X \varphi\, d\mu + \int_Y \psi\, d\nu.$$

Def. $\Pi(\mu, \nu)$ as above — the admissiable transport plan.

$$\Phi_c = \{(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu) \mid \varphi(x) + \psi(y) \leq c(x, y)\}$$

for $d\mu$-a.e. $x \in X$
$d\nu$-a.e. $y \in Y$.

Then

$$\inf_{\pi \in \Pi(\mu, \nu)} I[\pi] = \sup_{(\varphi, \psi) \in \Phi_c} J(\varphi, \psi). \tag{1.4}$$

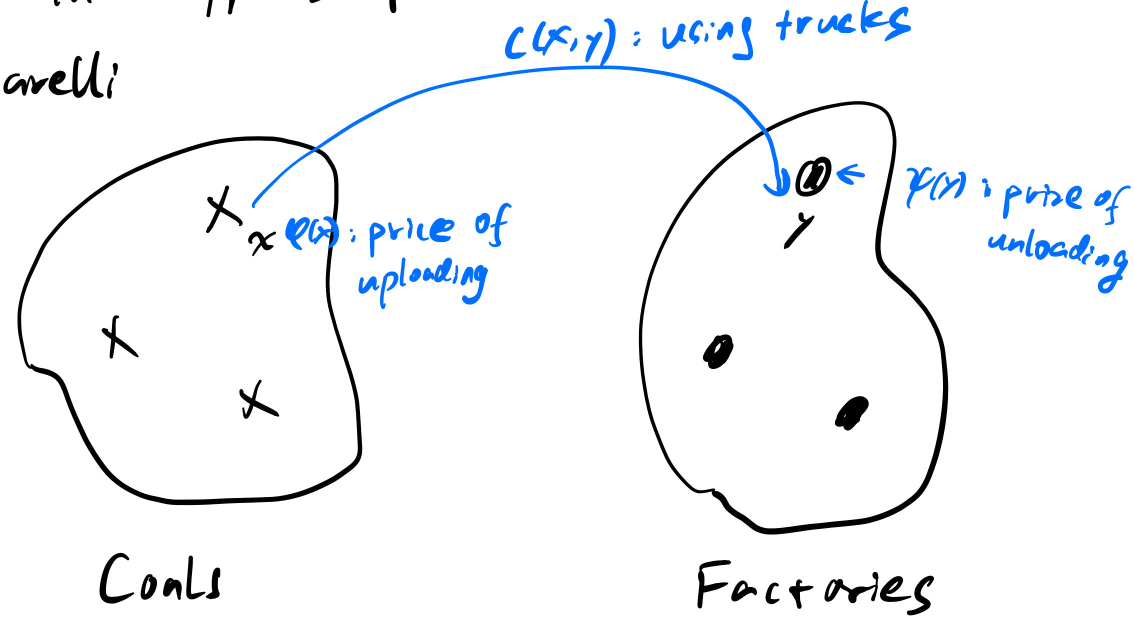
Moreover, the infimum in LHS above is attained. Furthermore, it does not change the value of supremum in RHS if one restricts the def. of Φ_c to those (φ, ψ) which are also bounded and continuous.

Recall: a function F on X (metric space) is said to be l.s.c. if

$$\forall x \in X, \quad F(x) \le \liminf_{y \to x} F(y).$$

Rk: The shipper's problem:

Caffarelli



Total cost by yourself: $\int_{X \times Y} c(x,y)\, d\pi(x,y)$

Free market cost:

$$\int_X \varphi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y).$$

Offer from the market:

$$\varphi(x) + \psi(y) \le c(x,y).$$

Kantorovich's duality: if the shipper is clever enough, then he can arrange the prices in such a way that you will pay him (almost) as much as you would have been ready to spend by the other method.

A Preliminary Observation:

$$(1.5) \quad \sup_{\Phi_c \cap C_b} J(\varphi, \psi) \leq \sup_{\Phi_c \cap L^1} J(\varphi, \psi) \leq \inf_{\Pi(\mu,\nu)} I[\pi].$$

Pf: The 1st inequ. is trivial.

For $(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)$, and $\pi \in \Pi(\mu, \nu)$,

$$J(\varphi, \psi) = \int_X \varphi \, d\mu + \int_Y \psi \, d\nu$$

$$= \int_{X \times Y} \underbrace{(\varphi(x) + \psi(y))} \, d\pi(x, y).$$

Since $\varphi \oplus \psi \leq c$, $d\pi(x,y)$-a.e.

$$\leq \int_{X \times Y} c(x, y) \, d\pi(x, y).$$

Indeed, let N_X, N_Y be measurable sets s.t.

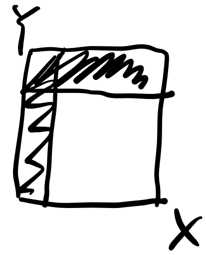
$$\mu[N_x] = 0, \quad \nu[N_y] = 0,$$

and $\quad \varphi(x) + \psi(y) \le c(x,y), \quad (x,y) \in N_x^c \times N_y^c.$

Also

$$\pi[N_x \times Y] = \mu[N_x] = 0,$$

$$\pi[X \times N_y] = \nu[N_y] = 0,$$



and hence $\pi[(N_x^c \times N_y^c)^c]$

$$\le \pi[N_x \times Y] + \pi[X \times N_y] = 0.$$

Consequently,

$$\int_{X\times Y} (\varphi(x) + \psi(y))\, d\pi(x,y) \underset{\substack{\uparrow \\ \text{taking inf}}}{\le} \inf \int_{X\times Y} c(x,y)\, d\pi(x,y) = I[\pi].$$

$\boxed{?}$

RK: Once one shows

$$\sup_{\Phi_c \cap C_b} J(\varphi, \psi) = \inf_{\pi \in \Pi(\mu,\nu)} I[\pi],$$

$$\hookrightarrow \quad \sup_{\Phi_c \cap L^1} J(\varphi, \psi) = \sup_{\Phi_c \cap C_b} J(\varphi, \psi).$$

<u>A formal proof</u> of Thm 1.3.

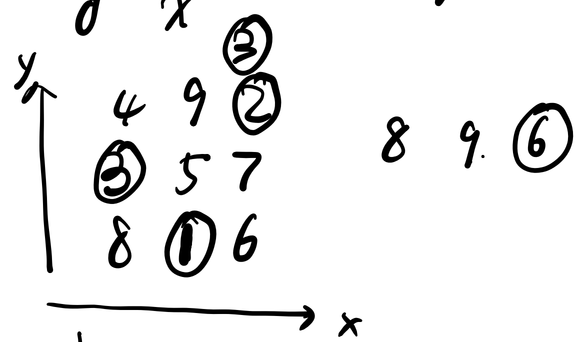
Basic idea: Rewrite the constrained infimum prob.

as an inf-sup problem,
(min-max)

exchange the two operations "inf-sup" by a "sup-inf".

(But notice in general, one only has

$$\min_x \max_y f(x,y) \geq \max_y \min_x f(x,y) \quad)$$



Major \ Class	1	2	3	→ x
MATH	x		x	
Physics		x		
Chemistry				

↓ y

Write

$$\inf_{\pi \in \Pi(\mu,\nu)} I[\pi] = \inf_{\pi \in M_+(\mu,\nu)} \left(I[\pi] + \begin{cases} 0 & \text{if } \pi \in \Pi(\mu,\nu) \\ +\infty & \text{else} \end{cases} \right).$$

Further

$$\begin{cases} 0 & \text{if } \pi \in \Pi(\mu,\nu) \\ +\infty & \text{else} \end{cases} = \sup_{(\varphi,\psi)} \Big[\int_X \varphi \, d\mu + \int_Y \psi \, d\nu \underbrace{- \int_{X\times Y} [\varphi(x) + \psi(y)] \, d\pi(x,y)}_{} \Big]$$

if $\neq 0$, then $\sup_{(\varphi,\psi)} \{\cdot\} = +\infty$,

where the supremum runs over all $(\varphi,\psi) \in C_b(X) \times C_b(Y)$.

Hence

$$\inf_{\Pi(\mu,\nu)} I[\pi] = \inf_{\pi \in M_+} \sup_{(\varphi,\psi)} \Big\{ \int_{X\times Y} c \, d\pi + \int_X \varphi \, d\mu + \int_Y \psi \, d\nu - \int_{X\times Y} (\varphi \oplus \psi) \, d\pi \Big\}$$

|| mini-max principle, but in general not true.

$$= \sup_{(\varphi,\psi)} \inf_{\pi \in M_+} \Big\{ \int_X \varphi \, d\mu + \int_Y \psi \, d\nu - \int_{X\times Y} (\varphi \oplus \psi - c) \, d\pi \Big\}$$

$$= \sup_{(\varphi,\psi)} \Big\{ \int_X \varphi \, d\mu + \int_Y \psi \, d\nu - \underline{\underline{\sup_{\pi \in M_+} \int_{X\times Y} (\varphi \oplus \psi - c) \, d\pi}} \Big\}$$

$= (DP)$.

if $\exists (x_0, y_0)$ s.t. $\zeta(x_0,y_0) = \varphi(x_0) + \psi(y_0) - c(x_0,y_0) > 0$, then choosing $\pi = \lambda \delta_{(x_0,y_0)}$, and letting $\lambda \to +\infty$, then $\sup\limits_{\lambda \in M_f} \{\cdot\} = +\infty$.

On the other hand, if $\varphi \oplus \psi \le c$,

then the supremum is 0. (taken when $\lambda = 0$).

Hence $(1.8) = \sup\limits_{\substack{(\varphi, \psi) \\ \varphi \oplus \psi \le c}} J(\varphi, \psi)$ okay.

$\blacksquare$

Lecture 17

Exercise 1.7 (Linear Programming)

Study of the minimization/maximization of linear problems subject to inequalities defined by linear functions.
(constraint)

Finite dimension case:

For $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, $A \in M_{m \times n}(\mathbb{R})$,

$$\sup_{Ax \leq b} c \cdot x = \inf_{y \geq 0, A^T y = c} b \cdot y,$$

if one of these extrema is achieved.

Here $Ax \leq b$, $y \geq 0$ hold component-wisely.

Computations:

$$\text{Left hand} = \sup_x \Big[c \cdot x + \inf_{y \geq 0} \underbrace{(b - Ax)}_{\geq 0} \cdot \underset{\geq 0}{y} \Big] \quad (b - Ax \in \mathbb{R}^m,\ y \in \mathbb{R}^m)$$

(if $Ax \leq b$, i.e. $(b - Ax) \geq 0$, $\inf_{y \geq 0} (b - Ax) \cdot y = 0$.

otherwise, $\exists (b - Ax)_i < 0$, $\inf_{y \geq 0} (b - Ax) \cdot y = -\infty$)

That is

$$\text{LHS} = \sup_{x\in\mathbb{R}^n} \inf_{\substack{y\in\mathbb{R}^m \\ y\geq 0}} \left[(c - A^T y)\cdot x + b\cdot y \right]$$

$$\overset{\text{mini-max ✓}}{=} \inf_{y\geq 0} \left\{ b\cdot y + \sup_{x} \left[(c - A^T y)\cdot x \right] \right\}$$

$$= \inf_{\substack{y\geq 0 \\ A^T y = c}} \{ b\cdot y \}$$

(NOT Rigorous)

<u>Rigorous Minimax principle:</u>

Convex analysis

Def (Legendre-Fenchel Transform)

Let E be normed vector space, Θ a convex function on E with values in $\mathbb{R}\cup\{+\infty\}$.

$$\left(\Theta(\lambda z_1 + (1-\lambda) z_2) \leq \lambda \Theta(z_1) + (1-\lambda)\Theta(z_2) \right)$$
$$\lambda\in[0,1],\quad z_1\ z_2 \in E$$

Legendre-Fenchel Transform of Θ is the function Θ^* defined on the topological dual E^* of E by

$$\Theta^*(z^*) = \sup_{z\in E} \left[\langle z^*, z\rangle - \Theta(z) \right]$$

Thm (Fenchel - Rockafeller duality)

Let E be a normed vector space,

E^* its top. dual space,

and Θ, Ξ two convex functions on E (with values in $\mathbb{R} \cup \{+\infty\}$. Let Θ^*, Ξ^* be the Legendre-Fenchel transforms of Θ, Ξ respectively.

Further assume $\exists\, z_0 \in E$, s.t. $\begin{cases} \Theta(z_0) < +\infty, \ \Xi(z_0) < +\infty, \\ \Theta \text{ is continuous at } z_0 \end{cases}$

Then,

$$\inf_E [\Theta + E] = \max_{z^* \in E^*} [-\Theta^*(-z^*) - \Xi^*(z^*)]$$

part of the theorem.

Proof:

$$-\Xi^*(z^*) = \inf_{z \in E} (-\langle z^*, z\rangle + \Xi(z))$$

$z \rightsquigarrow y$

sum up

$$-\Theta^*(-z^*) = \inf_{z \in E} (\langle z^*, z\rangle + \Theta(z))$$

$z \to x$

$\Downarrow$

$$LHS = \sup_{z^* \in E^*} \inf_{x, y \in E} \left(\Theta(x) + \Xi(y) + \langle z^*, x - y\rangle \right)$$

Need to prove

$$\inf_{x \in E} \{\Theta(x) + \Xi(x)\}.$$

$=$ RHS.

1. $\text{LHS} \underset{\forall x}{\overset{x=y}{\leq}} \sup_{z^* \in E^*} \left(\Theta(x) + \Xi(x) \right) = \Theta(x) + \Xi(x)$

so $\quad \text{LHS} \leq \text{RHS}$

Only need to show $\exists\ z^* \in E^*$, s.t. $\forall\ x, y \in E$,

$\Theta(x) + \Xi(y) + \langle z^*, x - y \rangle \geq \underline{\underline{m}} = \inf(\Theta + \Xi)$.

Finite, since $\Theta(z_0) + \Xi(z_0) < +\infty$.

2. Let $\quad C \equiv \{ (x, \lambda) \in E \times \mathbb{R} ;\ \lambda > \Theta(x) \}$, $\rightsquigarrow$ convex

$C' \equiv \{ (y, \mu) \in E \times \mathbb{R} ;\ \mu \leq m - \Xi(y) \}$ $\rightsquigarrow$ convex

(

$\lambda_1 > \Theta(x_1)$ $\qquad (x_1, \lambda_1) \quad (x_2, \lambda_2) \qquad \left(\frac{x_1 + x_2}{2}, \frac{\lambda_1 + \lambda_2}{2} \right)$

$\lambda_2 > \Theta(x_2)$ $\qquad \frac{\lambda_1 + \lambda_2}{2} > \frac{1}{2}\left(\Theta(x_1) + \Theta(x_2) \right) \geq \Theta\left(\frac{x_1 + x_2}{2} \right)$ ✓

$\mu_1 \leq m - \Xi(y_1)$

$\mu_2 \leq m - \Xi(y_2)$ $\qquad \frac{\mu_1 + \mu_2}{2} \leq m - \frac{1}{2}\left(\Xi(y_1) + \Xi(y_2) \right) \leq m - \Xi\left(\frac{y_1 + y_2}{2} \right)$)

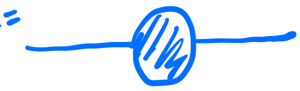
- $(z_0, \Theta(z_0) + 1) \in \text{Int}(C)$ since Θ is cont. at z_0

$z \sim z_0 \quad \lambda \sim \Theta(z_0) + 1, \quad \Rightarrow \quad \lambda > \Theta(z_0)$.

or $\text{Int}(C) \neq \phi \quad \Longrightarrow \quad \overline{C} = \overline{\text{Int}(C)}$

(This needs C convex).

(Of course, in general $\text{Int}(C) \neq \phi \not\Rightarrow \overline{C} = \overline{\text{Int}(C)}$.

e.g. $C=$  in $\mathbb{R}^2$, but C is not convex)

$C \cap C' = \phi$ ($\left.\begin{array}{l} \lambda > \Theta(x) \\ -\lambda > -m + \Xi(x) \end{array}\right\} \Rightarrow m > \Theta(x) + \Xi(x)$)

By Hahn–Banach theorem,

$\exists$ a nontrivial linear form $\ell \in (E \times \mathbb{R})^*$,

s.t.

$$\inf_{c \in C} \langle \ell, c \rangle = \inf_{c \in \mathrm{Int}(C)} \langle \ell, c \rangle \geq \sup_{c' \in C'} \langle \ell, c' \rangle.$$

Or $\exists\, w^* \in E^*$, $\alpha \in \mathbb{R}$, $(w^*, \alpha) \neq (0, 0)$, s.t.

$$\langle w^*, x \rangle + \underline{\underline{\alpha}}\,\lambda \geq \langle w^*, y \rangle + \underline{\underline{\alpha}}\,\mu,$$

where $\lambda > \Theta(x)$, $\mu \leq m - \Xi(y)$. Necessarily. $\alpha > 0$.

($\alpha < 0$, not possible; $\alpha = 0$, $\langle w^*, x \rangle \geq \langle w^*, y \rangle$ ✗)

Dividing by α, i.e. $z^* = w^*/\alpha$, we have

$$\langle z^*, x \rangle + \lambda \geq \langle z^*, y \rangle + \mu.$$

in particular:

$$\langle z^*, x \rangle + \Theta(x) \geq \langle z^*, y \rangle + m - \Xi(y)$$

i.e. $\forall x, y \in E$,

$$\underline{\Theta(x) + \Xi(y) + \langle z^*, x - y \rangle \geq m, \quad \text{okay}}$$

Recall Hahn–Banach theorem: [See in Ch1. Brezis]

Thm 1.6 in Brezis: Let $A \subset E$ and $B \subset E$ be two non-empty convex subsets s.t. $A \cap B = \phi$. Assume that one of them is open. Then $\exists$ a closed hyperplane that separates A and B.

Exercises: In the proof, why $\overline{C} = \overline{\mathrm{Int}(C)}$?

Proof of the Kantorovich duality:

Some properties of Polish space

- A Borel probability mea. μ on Polish space X is regular, i.e. $\forall$ Borel A,

$$\mu[A] = \sup\{\mu[K];\ K \text{ cpt},\ K \subset A\}$$
$$= \inf\{\mu[O];\ O \text{ open},\ A \subset O\}.$$

Riesz theorem (Rudin: Real and Complex Analysis.)

- $\mu \in P(X)$, X Polish; then μ is concentrated on a σ-compact set. ($\exists\ S$ - measurable,

$$S = \bigcup_{n=1}^{\infty} K_n,\ K_n \text{ cpt},\ \text{s.t. } \mu[S] = 1)$$

Also, μ is tight.

($\forall \varepsilon > 0$, $\exists$ K_ε cpt, s.t. $\mu[K_\varepsilon^c] \le \varepsilon$.)

- A family $\mathcal{A}$ of prob. meas. on top. space X is said to be tight if for any $\varepsilon > 0$, $\exists$ cpt $K_\varepsilon \subset X$, s.t.

$$\sup_{\mu \in \mathcal{A}} \mu[K_\varepsilon^c] \le \varepsilon.$$

Prokhorov's theorem Let X be Polish. Then any tight family $\mathcal{A}$ in $P(X)$ is relatively sequentially cpt in $P(X)$: for any sequence (μ_k) in $\mathcal{A}$, one can extract a sub-sequence, still denoted (μ_k), and a $\mu_* \in P(X)$, s.t. $\forall \varphi \in C_b(X)$,

$$\left.\begin{aligned} &\\ &\lim_{k\to\infty} \int_X \varphi \, d\mu_k = \int_X \varphi \, d\mu_* . \end{aligned}\right\} \text{Narrow Convergence.}$$

- Let (X, d) be a metric space, $F \ge 0$. Lower semi-continuous on X, then

$$F(x) = \lim_{n\to\infty} F_n(x),$$

where $\boxed{F_n(x) = \inf_{y\in X} [F(y) + n d(x, y)].}$

Check: ①

F_n is well-defined:

(Looking at the special $y=x$, $0 \le F_n(x) \le F(x)$;
$F_n(x) \le F_{n+1}(x)$ — $F_n \nearrow$)

② F_n is n-Lipschitz

$\forall x, y, \quad F_n(x) \le F(y) + n\,d(x,y)$

Fix x, $\forall \varepsilon > 0$, $\exists z$, st

$$F(z) + n\,d(x,z) - \varepsilon \le F_n(x)$$

Hence $F_n(y) \le F(z) + n\,d(y,z)$

$\le F(z) + n\,d(x,z) + n\,d(x,y)$

$\le F_n(x) + \varepsilon + n\,d(x,y)$

i.e. $F_n(y) - F_n(x) \le n\,d(x,y)$

Switching x and y leads to $|F_n(x) - F_n(y)| \le n\,d(x,y)$.

$\boxed{\checkmark}$

- If K is compact metric space, then $C(K)$ is separable.

Proof of Thm1.3 (Kantorovich's duality)

Three Steps:

Step I: Showing that minimax principle works under the assumptions that

$\begin{cases} X \text{ and } Y \text{ are compact;} \\ c \text{ is continuous} \end{cases}$

Step II and III: relax these two assumptions

Step I: Assume $\begin{cases} X \text{ and } Y \text{ compact} \\ c - \text{continuous} \end{cases}$

Write $E = C_b(X \times Y)$: bounded continuous, equipped with supremum norm $\|\cdot\|_\infty$.

By Riesz' theorem, its top. dual $\cong E^* = M(X \times Y)$, the space of Radon measures, normed by total variation.

(Recall the Riesz' Representation Theorem:

Thm 2.14 in Rudin: Let X be locally compact Hausdorff space, and Λ be a positive linear functional on $C_c(X)$. Then $\exists$ a σ-algebra $\mathfrak{M}$ on X which contains all Borel sets in X, and $\exists!$ positive measure μ on $\mathfrak{M}$ which represents Λ in the sense that

a) $\Lambda f = \int_X f \, d\mu, \quad \forall f \in C_c(X)$;

and TFHT:

b) $\mu(K) < \infty, \quad \forall$ cpt $K \subset X$;

c) $\forall E \in \mathfrak{M}$ (measurable),

$$\mu(E) = \inf\{\mu(V) : E \subset V \ \ V\text{-open}\}$$
(outer-regular)

d) The relation

$$\mu(E) = \sup\{\mu(K) \mid K \subset E, \ K\text{-cpt}\}$$

holds for every open set E, and for every $E \in \mathfrak{M}$, with $\mu(E) < \infty$;

e) If $E \in \mathfrak{M}$, $A \subset E$, and $\mu(E) = 0$, then $A \in \mathfrak{M}$.

($\Lambda_1 f = \int_X f \, d\mu_1$, $\Lambda_2 f = \int_X f \, d\mu_2$,

$(\Lambda_1 - \Lambda_2) f = \int_X f \, d(\mu_1 - \mu_2)$.))

Now we go back to the proof

Introduce

$$\Theta : u \in C_b(X \times Y) \mapsto \begin{cases} 0 & \text{if } u(x,y) \geq -c(x,y), \\ +\infty & \text{else} \end{cases}$$

$$\Xi_1 : u \in C_b(X \times Y) \longmapsto \begin{cases} \int_X \varphi \, d\mu + \int_Y \psi \, d\nu, & \text{if } u(x,y) = \varphi(x) + \psi(y), \\ +\infty, & \text{else.} \end{cases}$$

well-defined:

(if $\varphi(x) + \psi(y) = \tilde{\varphi}(x) + \tilde{\psi}(y)$, $\forall x, y$, then $\varphi(x) - \tilde{\varphi}(x) = \tilde{\psi}(y) - \psi(y)$ $\forall x, y$

$\Rightarrow \varphi - \tilde{\varphi} = \tilde{\psi} - \psi = s \in \mathbb{R}$. okay)

The assumptions on Thm 9 (Fenchel–Rockafeller) hold true: Θ – convex

Ξ_1 – linear $\Rightarrow$ convex

Choose z_0 as constant function 1.

Then formula (1.9) holds true:

$$(1.9) \quad \inf_{E} [\Theta + \Xi_1] = \max_{z^* \in E^*} [-\Theta^*(-z^*) - \Xi_1^*(z^*)],$$

where $\Theta^*(z^*) = \sup_{z \in E} [\langle z^*, z \rangle - \Theta(z)]$

$$\text{LHS} = \inf_{\substack{u \in C_b(X \times Y) \\ u \geq -c(x,y) \\ u = \varphi \oplus \psi}} \left\{ \int_X \varphi \, d\mu + \int_Y \psi \, d\nu \right\}$$

$-\varphi = \tilde{\varphi}, \quad -\psi = \tilde{\psi}$

$\tilde{\varphi} + \tilde{\psi} \leq c$

$-\varphi - \psi \leq c(x,y)$

$$= \inf \left\{ \int_X \varphi \, d\mu + \int_Y \psi \, d\nu \;\middle|\; \varphi(x) + \psi(y) \geq -c(x,y) \right\}$$

$$= -\sup \left\{ \int_X \varphi \, d\mu + \int_Y \psi \, d\nu \;\middle|\; \varphi(x) + \psi(y) \leq c(x,y) \right\}$$

$$= -\sup \{ J(\varphi, \psi) \mid (\varphi, \psi) \in \Phi_c \}$$

Next, we compute the Legendre–Fenchel transform of Θ and Ξ_1.

Recall:

$$E = C_b(X\times Y), \quad E^* = M(X\times Y) \text{ — Radon measures}$$

$$\Theta^*(-\pi) = \sup_{u\in C_b(X\times Y)} \left\{ -\int u(x,y)\,d\pi(x,y) \,;\; u(x,y) \geq -c(x,y) \right\}$$

$$\left(\Theta^*(z^*) = \sup_{z\in E} \left[\langle z^*, z\rangle - \Theta(z) \right] \right)$$

$$= \sup_{u\in C_b(X\times Y)} \left\{ \int u(x,y)\,d\pi(x,y) \,;\; u(x,y) \leq c(x,y) \right\}$$

- If π is not nonnegative measure, then $\exists$ a non-positive function $v \in C_b(X\times Y)$, s.t. $\int v\,d\pi > 0$. Then $u = \lambda v$, $\lambda \to +\infty$, shows that the sup. $= +\infty$.
- If π is non-negative, then the sup. $= \int c\,d\pi$.

Thus
$$\Theta^*(-\pi) = \begin{cases} \int c(x,y)\,d\pi(x,y), & \text{if } \pi \in M_+(X\times Y), \\ +\infty, & \text{else.} \end{cases}$$

Similarly, for $\pi \in M(X\times Y)$

$$\Xi_1^*(\pi) = \sup_{u\in C_b(X\times Y)} \left\{ \langle \pi, u\rangle - \Xi_1(u) \right\}$$

Note π is fixed, while u is arbitrary

For any u, admitting a decomposition $u = \varphi\oplus\psi$, $\{\cdot\} = \int u\,d\pi - \underbrace{\left(\int_X \varphi\,d\mu + \int_Y \psi\,d\nu\right)}$

$\underbrace{\leq 0}_{-\infty}$ $\quad = 0$ Only case. $\quad$ or > 0 $\quad u \to \lambda u$, $+\infty$ $\lambda \to +\infty$

$+\infty$

$u \to \lambda u,\ \lambda \to -\infty$

For u, if $u \neq \varphi \oplus \psi$, $\Xi_1(u) = +\infty$, no contribution

To sum it up:

$$\Xi_1^*(\pi) = \begin{cases} 0 & \text{if } \forall (\varphi, \psi) \in C_b(X) \times C_b(Y), \\ & \int (\varphi(x) + \psi(y))\, d\pi(x,y) = \int \varphi\, d\mu + \int \psi\, d\nu \\ +\infty & \text{else} \end{cases}$$

Θ^* : indicates $\pi \in M_+(X \times Y)$

Ξ_1^* : indicates that $\pi \in \Pi(\mu, \nu)$

And

$$\max_{\pi \in M(X \times Y)} -\left[\Theta^*(-\pi) + \Xi_1^*(\pi)\right]$$

$$= -\min_{\pi \in M(X \times Y)} \left\{ \int c(x,y)\, d\pi(x,y) \;\middle|\; \begin{array}{l} \pi \in M_+(X \times Y) \\ \pi \in \Pi(\mu, \nu) \end{array} \right\}$$

$$= -\min_{\pi \in \Pi(\mu,\nu)} I[\pi]$$

Finally $$\boxed{\min_{\pi \in \Pi(\mu,\nu)} I[\pi] = \sup_{\Phi_c \cap C_b} J(\varphi, \psi)}$$

Step II & III: Relaxations.

Chapter 1 of "A user's guide to optimal transport".

Give a quick mention of the many aspects of optimal transport theory

Ch1: introduce to optimal transport problem and its formulation in terms of transport maps and transport plans

- duality formula
- c - monotonicity
- existence of optimal maps in the model case cost = distance^2.

Ch2: Wasserstein W_2. $(P(X), W_2)$.

X Polish.

geodesics in $P(X)$ [illegible] (geodesics in X)

time evolution of Kantorovich potentials and Hopf-Lax semigroups

$X = M$ = Riemannian manifold $\Rightarrow$ time-dependent optimal transport problem geodesics.

Benamou-Brenier formula

$(\mathcal{P}_2(M), W_2)$ (Otto's work)

Ch3: gradient flows

① classical theory for λ-convex functionals in Hilbert spaces;

② Equivalent formulations that involve only the distance $\xrightarrow[\text{to}]{\text{Apply}}$ general metric space

③ EVI, EDE.

discrete version of Gradient Flow: given by implicit Euler scheme.

(Convergence of the scheme to continuous solution as τ (time discretization) $\downarrow 0$.

(Some research by Carrillo, Figalli, ...)

Ch4: Applications to classical functional / geometric inequalities.

- Brunn-Minkowski

- isoperimetric inequality
- Log Sobolev inequality

optimal effective versions

Ch5 : Variants of optimal transport

① Branched transportation

② Modification in the action functional on curves

③ with unequal mass

Chapter 1

1. The optimal transport problem

Two Formulations

(X, d) Polish $\quad \mathcal{P}(X)$ Borel $\quad \operatorname{supp}(\mu)$

two Polish

$T: X \to Y$ Borel map, $\mu \in \mathcal{P}(X)$

$$T: (X, \mu) \to (Y, T_\# \mu)$$

$$T_\# \mu(E) = \mu(T^{-1}(E)) \quad \forall E \subset Y \text{ Borel.}$$

The push forward is characterized by

$$\int_Y f(y)\, d\, T_\#\mu(y) = \int_X f(Tx)\, d\mu(x),$$

$$\forall f: Y \to \mathbb{R} \cup \{\pm\infty\} \text{ Borel}$$

Borel cost function:

$c: X \times Y \to \mathbb{R} \cup \{+\infty\}$.

Monge version:

Prob 1.1 For $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$.

$$\min_{T : T_\# \mu = \nu} \int_X c(x, T(x))\, d\mu(x)$$

Sometimes ill-posed:

① $\mu = \delta_0$. $\nu = \frac{1}{2}(\delta_1 + \delta_{-1})$

no admissable T.

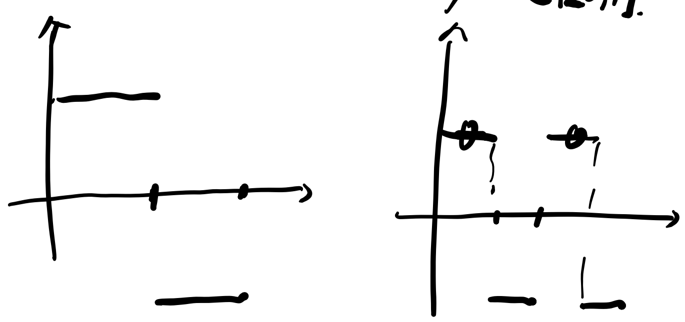
② $T_\# \mu = \nu$ is not weakly sequentially closed.

($f_n(x) = f(nx)$ $f : \mathbb{R} \to \mathbb{R}$ 1-periodic

$$f(x) = \begin{cases} 1 & \text{on } [0, \frac{1}{2}) \\ -1 & \text{on } [\frac{1}{2}, 1) \end{cases}$$

f

$\mu = \mathcal{L}|_{[0,1]}$ $\nu = \frac{1}{2}(\delta_{-1} + \delta_1)$



f_1

$(f_n)_\# \mu = \nu$

$f_n \rightharpoonup f \equiv 0$

$f_\# \mu = \delta_0 \neq \nu$.

Kantorovich's formulation

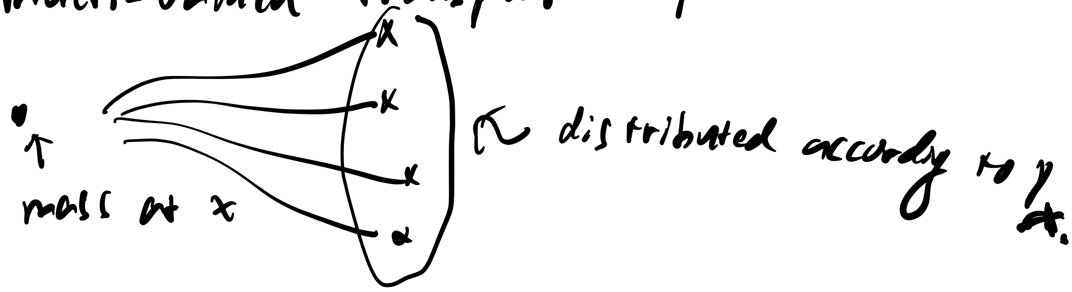
$$\min_{\gamma \in \Pi(\mu,\nu)} \int_{X\times Y} c(x,y)\, d\gamma(x,y)$$

$$\Pi(\mu,\nu) = \{ \gamma \in \mathcal{P}(X\times Y) \mid \gamma(A\times Y) = \mu(A),\ \forall A \in \mathcal{B}(X);\ \gamma(X\times B) = \nu(B),\ \forall B \in \mathcal{B}(Y) \}$$

$$\pi^X_\# \gamma = \mu, \quad \pi^Y_\# \gamma = \nu.$$

$\gamma = \int \gamma_x \, d\mu(x)$, $\gamma_x \in \mathcal{P}(\{x\}\times Y)$

multi-valued transport maps.



- $\Pi(\mu,\nu) \neq \emptyset$
- $\Pi(\mu,\nu)$ convex; compact w.r.t. the narrow topology in $\mathcal{P}(X\times Y)$
- $\gamma \longmapsto \int c \, d\gamma = \langle c, \gamma \rangle$ linear.
- minima always exists under mild

assumptions on c

- $T_\# \mu = \nu \Rightarrow (Id, T)_\# \mu = \gamma \in \Pi(\mu, \nu)$

Notions concerning analysis over Polish space:

$(\mu_n) \subset \mathcal{P}(X)$ narrowly converge to μ

if $\int \varphi \, d\mu_n \to \int \varphi \, d\mu$, $\forall \varphi \in C_b(X)$.

The topology of narrow convergence is metrizable.

A set $\mathcal{K} \subset \mathcal{P}(X)$ is called tight

if $\forall \varepsilon > 0$, $\exists$ cpt set $K_\varepsilon \subset X$, s.t.

$$\mu[X \setminus K_\varepsilon] \leq \varepsilon \quad \forall \mu \in \mathcal{K}.$$

<u>Prokhorov</u>:

(X, d) Polish

$\mathcal{K} \subset \mathcal{P}(X)$ relatively compact w.r.t. narrow topology

$\Leftrightarrow \mathcal{K}$ is tight

$K = \{\mu\}$. $\Rightarrow$ Ulam's theorem: any Borel probability measure on a Polish space is concentrated on a σ-compact set.

Claim: $\Pi(\mu,\nu)$ is tight, given X, Y Polish, $\mu \in P(X)$, $\nu \in P(Y)$.

Pf: $\forall\, \gamma \in \Pi(\mu,\nu) \Leftrightarrow$ tight.

$$\gamma(X \times Y \setminus K_1 \times K_2) \leq \underbrace{\mu(X \setminus K_1)}_{<\varepsilon/2} + \underbrace{\nu(X \setminus K_2)}_{<\varepsilon/2}.$$

Present this theory first:

Existence of minimizers for Kantorovich's formulation

(comes from $\begin{cases} \text{l.s.c.} \\ \text{compactness} \end{cases}$ arguments.

Thm 1.5 Assume c is l.s.c. and c is bounded from below.

Then $\exists$ a minimizer for

$$\inf_{\gamma \in \Pi(\mu,\nu)} \int c(x,y)\, d\gamma(x,y).$$

Pf: • $\Pi(\mu,\nu)$ is tight in $P(X\times Y)$

and hence relatively compact by Prokhorov.

• Further, for a sequence $\{\gamma_n\}\subset\Pi(\mu,\nu)$,

up to a subsequence, one can assume that

$\gamma_n\to\gamma$ narrowly,

we claim $\gamma\in\Pi(\mu,\nu)$.

Take $\varphi\in C_b(X)\subset C_b(X\times Y)$, then

$$\int\varphi(x)\,d\gamma(x,y)$$

$$=\lim_{n\to\infty}\int\varphi(x)\,d\gamma_n(x,y)=\lim_{n\to\infty}\int\varphi(x)\,d\mu(x)$$

$$=\int\varphi(x)\,d\mu(x)$$

$$\left.\begin{array}{ll}\text{i.e.} & \pi^X_{\#}\gamma=\mu.\\ \text{Similarly} & \pi^Y_{\#}\gamma=\nu.\end{array}\right\}\Rightarrow\gamma\in\Pi(\mu,\nu).$$

• The functional $\gamma\mapsto\int c\,d\gamma$ is l.s.c.

w.r.t. narrow convergence

c l.s.c. $c \geq -b$

$\hookrightarrow$ $c(x,y) = \lim_n c_n(x,y)$, $c_n \uparrow$,

$c_n \in L^\infty$.

c_n - uniformly continuous.

By monotone convergence theorem

$$\int c \, d\gamma = \sup_n \int c_n \, d\gamma = \lim_{n\to\infty} \int c_n d\gamma.$$

Since by construction,

$\gamma \mapsto \int \underset{\substack{\cap \\ C_b(X\times Y)}}{\underline{c_n}} \, d\gamma$ is narrowly continuous,

Say for $\gamma_m \to \gamma$ narrowly as $m \to \infty$,

$$\begin{aligned}\int c \, d\gamma &= \sup_n \int c_n \, d\gamma \\ &= \sup_n \lim_{m\to\infty} \underbrace{\int c_n \, d\gamma_m}_{\leq \int c \, d\gamma_m} \\ &\leq \liminf_{m\to\infty} \int c \, d\gamma_m ,\end{aligned}$$

i.e. $\gamma \mapsto \int c \, d\gamma$ is l.s.c w.r.t.

narrow convergence.

Take a minimizing sequence (γ_n) and assuming up to a subsequence, $\gamma_n \to \gamma$ narrowly, then

$$\int c \, d\gamma \le \liminf_{n\to\infty} \int c \, d\gamma_n = \inf_{\tilde{\gamma}\in\Pi(\mu,\nu)} \int c \, d\tilde{\gamma}$$

clearly.

$OPT(\mu,\nu)$: optimal plans

(cost c is clear from the context.)

$\boxed{\text{Page 7 - Page 15.}}$ All but optimal maps

Once we have the existence, many questions arise

1. uniqueness
2. characterization of optimal plan
3. regularity theory . maps.

@ difference between Monge and Kantorovich

When $\inf(\text{Monge}) = \inf(\text{Kantorovich})$?

1.2. Necessary and sufficient conditions

Motivation Examples:

Take $X = Y = \mathbb{R}^d$,

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

$\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ supported on finite sets

Claim: $\gamma \in \Pi(\mu,\nu)$ is optimal

$$\Updownarrow$$

$$\sum_{i=1}^{N} \frac{|x_i - y_i|^2}{2} \le \sum_{i=1}^{N} \frac{|x_i - y_{\sigma(i)}|^2}{2},$$

for any $N = 1, 2, 3, \cdots$, $(x_i, y_i) \in \operatorname{supp}(\gamma)$,

and any $\sigma \in [N]$.

Question: why this has nothing to do

with the mass?

Expanding the squares we get

$$\sum_{i=1}^{N} \langle x_i, y_i \rangle \geq \sum_{i=1}^{N} \langle x_i, y_{\sigma(i)} \rangle,$$

or $\mathrm{supp}(\gamma)$ is cyclically monotone.

Recall

Rockafeller: A set $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ is cyclically monotone iff $\exists$ a convex and l.s.c. $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ s.t.

$$\Gamma \subset \mathrm{graph}(\partial \varphi)$$

sub-differential of φ.

Equivalence between the following three notions:

- $\gamma \in \Pi(\mu, \nu)$ optimal,

- $\mathrm{supp}(\gamma)$ is cyclically monotone ;
- $\exists$ convex, l.s.c. φ, s.t.

$$\mathrm{supp}(\gamma) \subset \mathrm{graph}(\partial\varphi)$$

Hold in much more general setting.

Some notations for general Borel and real-valued cost :

Def : (c-cyclical monotonicity)

We say that $\Gamma \subset X \times Y$ is c-cyclically monotone if $(x_i, y_i) \in \Gamma$, $1 \le i \le N$, $N = 1, 2, 3, \dots$, implies

$$\sum_{i=1}^{N} c(x_i, y_i) \le \sum_{i=1}^{N} c(x_i, y_{\sigma(i)}),$$

for any $\sigma \in S_N$.

Def : (c-transform)

(Generalizations of Legendre transform)

Let $\psi : Y \to \mathbb{R} \cup \{\pm\infty\}$ be any function.

Its c_+ - transform $\psi^{c_+} : X \to \mathbb{R} \cup \{-\infty\}$ is defined as

$$\psi^{c_+}(x) := \inf_{y \in Y} \{ c(x,y) - \psi(y) \}$$

Similarly, given $\varphi : X \to \mathbb{R} \cup \{\pm\infty\}$, its c_+ - transform is the function $\varphi^{c_+} : X \to \mathbb{R} \cup \{-\infty\}$ defined by

$$\varphi^{c_+}(y) := \inf_{x \in X} \{ c(x,y) - \varphi(x) \}.$$

The c_- - transform $\psi^{c_-} : X \to \mathbb{R} \cup \{+\infty\}$ of a function ψ on Y is given by

$$\psi^{c_-}(x) := \sup_{y \in Y} \{ \underline{\underline{-c(x,y)}} - \psi(y) \}$$

Def: (c - concavity and c - convexity)

$\varphi: X \to \mathbb{R} \cup \{-\infty\}$ is c-concave

if $\exists\, \psi: Y \to \mathbb{R} \cup \{-\infty\}$ s.t. $\underline{\varphi = \psi^{c_+}}$.

($\psi: Y \to \mathbb{R} \cup \{-\infty\}$ is c-concave, if

$\exists\, \varphi: X \to \mathbb{R} \cup \{-\infty\}$, s.t. $\psi = \varphi^{c_+}$.)

Symmetrically, $\varphi: X \to \mathbb{R} \cup \{+\infty\}$ is c-convex if

$\exists\, \psi: Y \to \mathbb{R} \cup \{+\infty\}$, s.t. $\varphi = \psi^{c_-}$; and

$\psi: Y \to \mathbb{R} \cup \{+\infty\}$ is c-convex if $\exists\, \varphi: X \to \mathbb{R} \cup \{+\infty\}$

s.t. $\psi = \varphi^{c_-}$.

<u>Claim</u>: $\varphi: X \to \mathbb{R} \cup \{-\infty\}$ is c-concave

$$\Updownarrow$$

$$(\varphi^{c_+})^{c_+} = \varphi^{c_+ c_+} = \varphi$$

Pf: Indeed, for any function $\psi: Y \to \mathbb{R} \cup \{\pm\infty\}$

it holds $\boxed{\psi^{c_+} = \psi^{c_+ c_+ c_+}}$

$\psi: Y \to \mathbb{R} \cup \{\pm\infty\}$

$$\psi^{c_+}(x) = \inf_{y \in Y} \left[c(x,y) - \psi(y) \right] \quad x \in X$$

$$\psi^{c_+c_+}(z) = \inf_{\tilde{x}\in X} \left[c(\tilde{x}, z) - \psi^{c_+}(\tilde{x}) \right]$$

$z \in Y$

$$\psi^{c_+c_+c_+}(x) = \inf_{\bar{y}\in Y} \left[c(x, \bar{y}) - \psi^{c_+c_+}(\bar{y}) \right]$$

$$= \inf_{\bar{y}\in Y} \sup_{\bar{x}\in X} \left[c(x,\bar{y}) - c(\bar{x},\bar{y}) + \psi^{c_+}(\bar{x}) \right]$$

$$= \inf_{\bar{y}\in Y} \sup_{\bar{x}\in X} \inf_{y\in Y} \left[c(x,\bar{y}) - c(\bar{x},\bar{y}) + c(\bar{x}, y) - \psi(y) \right]$$

Choosing $x = \bar{x}$

$$\psi^{c_+c_+c_+}(x) \geq \inf_{y\in Y} \left[c(x,y) - \psi(y) \right] = \psi^{c_+}$$

Choosing $y = \bar{y}$,

$$\psi^{c_+c_+c_+}(x) \leq \inf_{y\in Y} \left[c(x,y) - \psi(y) \right]$$

(taking infimum over smaller set!!)

If ψ is c-concave,

then $\exists\, \psi$, s.t. $\varphi = \psi^{c_+}$,

$$\varphi^{c_+c_+} = \psi^{c_+c_+c_+} = \psi^{c_+} = \varphi. \quad \text{Okay}$$

$\boxtimes$

Note

$$\psi^{c_-}(x) = \sup_{y\in Y}\{-c(x,y) - \psi(y)\}$$

$$= -\inf_{y\in Y}\{c(x,y) - (-\psi(y))\}$$

$$-\psi^{c_-}(x) = \inf_{y\in Y}\{c(x,y) - (-\psi)(y)\}$$

$$= (-\psi)^{c_+}(x)$$

i.e. $$\boxed{\psi^{c_-} = -(-\psi)^{c_+}}$$

$$\psi^{c_-c_-} = -(-\psi)^{c_+c_+}$$

$$\psi^{c_-c_-c_-} = -(-\psi)^{c_+c_+c_+} \quad \text{etc..}$$

Def 1.10 (c-superdifferential and c-subdifferential)

Let $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ be a c-concave function.

The c-superdifferential

$\partial^{c+}\varphi \subset X \times Y$ is defined as

$$\partial^{c+}\varphi := \{(x,y) \in X \times Y \mid \varphi(x) + \varphi^{c+}(y) = c(x,y)\}.$$

The c-super-differential $\partial^{c+}\varphi(x)$ at x is

$$\partial^{c+}\varphi(x) = \{x \in X \mid (x,y) \in \partial^{c+}\varphi\}$$

One can also define $\partial^{c+}\psi$ for a c-concave functions $\psi : Y \to \mathbb{R} \cup \{-\infty\}$.

c-subdifferential $\partial^{c-}\varphi$ for a c-convex function $\varphi : X \to \{+\infty\}$ is defined as

$$\partial^{c-}\varphi := \{(x,y) \in X \times Y \mid \varphi(x) + \varphi^{c-}(y) = -c(x,y)\}$$

$\Big(\Leftarrow \; \varphi^{c-}(y) := \sup_{x \in X} \left[-c(x,y) - \varphi(x) \right]$

The pair (x,y) s.t. $\varphi^{c-}(y) = -c(x,y) - \varphi(x)$

or $\varphi(x) + \varphi^{c-}(y) = -c(x,y). \Big)$

<u>Rk]</u>

- The classical case

$c(x,y) = -\langle x, y \rangle, \quad X = Y = \mathbb{R}^d$

$\left(\text{or } c(x,y) = \frac{|x-y|^2}{2} \right)$

Now φ is "c-convex" means that

$\exists \psi$, s.t. $\varphi = \psi^{c-}$, or

$$\varphi(x) = \sup_{y \in \mathbb{R}^d} \{ \langle x, y \rangle - \psi(y) \}$$

i.e φ is the Legendre transform of ψ

now φ is convex. and l.s.c.

(the superior envelop of those functions, y runs over $\mathbb{R}^d$)

$$\varphi_{v}\left(\frac{x_1+x_2}{2}\right) \le \frac{1}{2}\left(\varphi_{v}(x_1) + \varphi_{v}(x_2)\right)$$

$$\le \frac{1}{2}\left[(\sup_{v} \varphi_{v})(x_1) + (\sup_{v} \varphi_{v})(x_2)\right] \checkmark$$

- the c-subdifferential (super...) of the c-convex (concave) function is the classical sub-differential. (super...)

- c-transform = Legendre transform

$\boxed{\text{We just use "c-concave"}}$

$$\underline{\underline{y \in \partial^{c+}\varphi(x)}} \iff \boxed{\begin{array}{l} \varphi(x) + \varphi^{c+}(y) = c(x,y) \\ \varphi(z) + \varphi^{+}(y) \le c(z,y) \quad \forall z \in X \end{array}}$$

$$\Updownarrow$$

$$\begin{cases} \varphi(x) = c(x,y) - \varphi^{c+}(y), \\ \varphi(z) \le c(z,y) - \varphi^{c+}(y), \quad \forall z \in X \end{cases}$$

$$\Updownarrow$$

$$\underline{\underline{\varphi(x) - c(x,y) \ge \varphi(z) - c(z,y), \quad \forall z \in X.}}$$

Consequently,

the c-superdiff of a c-concave function is always a c-cyclically monotone set:

Indeed, if $(x_i, y_i) \in \partial^{c\varphi} \varphi$,

$$\sum_i c(x_i, y_i) = \sum_i \left(\varphi(x_i) + \varphi^{c\varphi}(y_i)\right)$$
$$= \sum_i \left(\varphi(x_i) + \varphi^{c\varphi}(y_{\sigma(i)})\right)$$
$$\leq \sum_i c(x_i, y_{\sigma(i)}),$$

for any permutation σ.

More generally, one has the following thm:

Thm (Fundamental theorem of optimal transport)
(Kind of characterization theorem)

Assume that $c : X \times Y \to \mathbb{R}$ is continuous and bounded from below and let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ be s.t. $c(x,y) \leq a(x) + b(y)$, for some $a \in L^1(d\mu)$

$b \in L^1(d\nu)$.

And Let $\gamma \in \Pi(\mu,\nu)$ (Admissable).

Then TFAE:

i) The plan γ is optimal;

ii) The set $\mathrm{supp}(\gamma)$ is c-cyclically monotone;

iii) $\exists$ a c-concave function φ, s.t.

$\max\{\varphi, 0\} \in L^1(d\mu)$ and $\mathrm{supp}(\gamma) \subset \partial^{c+}\varphi$.

Pf:

Observation: $\forall \tilde{\gamma} \in \Pi(\mu,\nu)$

$$\int c(x,y)\, d\tilde{\gamma}(x,y) \le \int (a(x)+b(y))\, d\tilde{\gamma}(x,y)$$

$$= \int a(x)\, d\mu(x) + \int b(y)\, d\nu(y) < +\infty$$

$\Rightarrow \forall \tilde{\gamma} \in \Pi(\mu,\nu)$, $\max\{c, 0\}$ is integrable.

Since c is bounded from below,

$c \in L^1(\tilde{\gamma})$.

i) $\Rightarrow$ ii) We argue by contradiction:

Assume that $supp(\gamma)$ is not c-cyclically monotone. Then we can find $N \in \mathbb{N}$, $\{(x_i, y_i)\}_{i=1}^N \subset supp(\gamma)$ and some permutation $\sigma \in S_N$, s.t.

$$\sum_{i=1}^N c(x_i, y_i) > \sum_{i=1}^N c(x_i, y_{\sigma(i)}).$$

By continuity, we can find nbhds $x_i \in U_i$, $y_i \in V_i$ with

$$\sum_{i=1}^N \left[c(u_i, v_{\sigma(i)}) - c(u_i, v_i) \right] < 0,$$

$$\forall (u_i, v_i) \in U_i \times V_i, \quad 1 \le i \le N.$$

Idea: build a "variation" $\tilde{\gamma} = \gamma + \eta$ of γ in such a way that minimality of γ is violated. To this end, we need a signed measure η with

(A) $\quad \eta^- \le \gamma$ (so that $\tilde{\gamma} \in M_+(X \times Y)$)

(B) $\int \eta \, dx = 0$ $\xrightarrow{} $ Null $\quad$ (so that $\tilde{\gamma} \in \Pi(\mu,\nu)$)

$\int \eta \, dy = 0$

(C) $\int c \, d\eta < 0$ (so that γ is not optimal).

Let $\Omega := \prod_{i=1}^{N} U_i \times V_i$,

let $P \in \mathcal{P}(\Omega)$ be defined as the product measure

$$\frac{1}{m_i} \gamma|_{U_i \times V_i}, \quad \text{where } m_i = \gamma(U_i \times V_i).$$

π^{U_i}, π^{V_i} : natural projections of Ω to U_i and V_i respectively and define

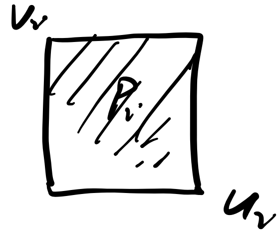
$$\eta := \frac{\min_i m_i}{N} \sum_{i=1}^{N} \left\{ (\pi^{U_i}, \pi^{V_{\sigma(i)}})_{\#} P - (\pi^{U_i}, \pi^{V_i})_{\#} P \right\}$$

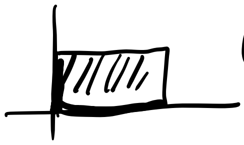
Checking η satisfies A), B) and C).

(The key part is actually to understand that:

$$\frac{c}{N} \sum_i \left\{ (\pi^{U_i}, \pi^{V_{\sigma(i)}})_{\#} P - (\pi^{U_i}, \pi^{V_i})_{\#} P \right\}$$

really means





$$(\mu\times\nu)(A\times B) = \mu(A)\,\nu(B)$$

$$\int_{X\times Y} c(x,y)\, d\eta(x,y) = \frac{c}{N}\sum_i \Big[\int c(x,y)\, d(\pi^{U_i}, \pi^{V_{\sigma(i)}})_{\#}\mathbb{P} - \underbrace{\int c(x,y)\, d(\pi^{U_i}, \pi^{V_i})_{\#}\mathbb{P}}_{\displaystyle = \frac{1}{m_i}\int_{U_i\times V_i} c(x,y)\, d\gamma(x,y)}$$

$$\mathbb{P}\in\mathcal{P}(\Omega) \qquad \Omega = \prod_{i=1}^{N} U_i\times V_i = \Big\{ (\underbrace{x_1, y_1, \cdots x_N, y_N}_{X^N}) \;\Big|\; \begin{matrix} x_i\in U_i \\ y_i\in V_i \end{matrix} \Big\}$$

$$\int \phi \, d\, T_{\#}\mu = \int \phi(Tx)\, d\mu$$

$$\int c \, d\Big\{ \frac{c}{N}\sum_i (\pi^{U_i}, \pi^{V_{\sigma(i)}})_{\#} - (\pi^{U_i}, \pi^{V_i})_{\#} \Big\}\mathbb{P}$$

$$= \frac{c}{N}\sum_i \int \Big[c(x_i, y_{\sigma(i)}) - c(x_i, y_i) \Big]\, d\mathbb{P}(x_1, y_1, \cdots x_N, y_N)$$

$$< 0$$

ii) $\Rightarrow$ iii)

We now prove that if $\Gamma \subset X \times Y$ is c-cyclically monotone, then $\exists$ a c-concave function φ, s.t. $\partial^{c+}\varphi \supset \Gamma$, and

$$\max\{\varphi, 0\} \in L^1(d\mu).$$

Fix $(\bar{x}, \bar{y}) \in \Gamma$, for perspective c-concave function φ s.t. $\partial^c\varphi \supset \Gamma$, we need impose that $\forall$ $(x_i, y_i) \in \Gamma$, $i = 1, 2, \cdots N$,

$$\varphi(x) \le c(x, y) - \varphi^{c+}(y_1) \quad (\Leftarrow y_1 \in \partial^{c+}\varphi(x_1))$$

$$= c(x, y) - c(x_1, y_1) + \varphi(x_1)$$

$$\le (c(x, y) - c(x_1, y_1)) + c(x_1, y_2) - \varphi^{c+}(y_2)$$

$$= (c(x, y_1) - c(x_1, y_1)) + (c(x_1, y_2) - c(x_2, y_2)) + \varphi(x_2)$$

$$\le \cdots$$

$$\le (c(x, y_1) - c(x_1, y_1)) + (c(x_1, y_2) - c(x_2, y_2)) + \cdots$$
$$+ \cdots + (c(x_N, \bar{y}) - c(\bar{x}, \bar{y})) + \varphi(\bar{x})$$

Define φ as the infimum of the above expression as $\{(x_i, y_i)\}_{i=1}^{N}$ vary among all N-pairs in Γ and $N = 1, 2, 3, \ldots$. We are free to add a constant to φ, so

define:

$$\varphi(x) := \inf \Big\{ \big(c(x, y_1) - c(x_1, y_1)\big) + \big(c(x_1, y_2) - c(x_2, y_2)\big) + \cdots + \big(c(x_N, \bar{y}) - c(\bar{x}, \bar{y})\big) \Big\}.$$

Choosing $N = 1$, and $(x_1, y_1) = (\bar{x}, \bar{y})$, $x = \bar{x}$,

we get $\varphi(\bar{x}) \leq 0$.

Conversely, from the c-cyclical monotonicity of Γ

$$\left(\left.\begin{aligned} &\big(c(\bar{x}, y_1) - c(x_1, y_1)\big) \\ &+ c(x_1, y_2) - c(x_2, y_2) \\ &\quad \vdots \\ &+ c(x_{N-1}, y_N) - c(x_N, y_N) \\ &+ c(x_N, \bar{y}) - c(\bar{x}, \bar{y}) \end{aligned}\right\} \geq 0 \right)$$

$$\begin{array}{ccccccc} \bar{x} & & x_1, & \cdots & x_N & & \\ \bar{y} & \searrow & y_1 & \searrow \cdots & y_N & \searrow & \bar{y} \end{array}$$

one has $\boxed{\varphi(\bar{x}) \geq 0}$

Thus $\varphi(\bar{x}) = 0$.

Clearly, from the construction, φ is c-concave.

Taking $N=1$, $(x_1, y_1) = (\bar{x}, \bar{y})$,

$$\varphi(x) \le \underbrace{c(x, \bar{y})} - c(\bar{x}, \bar{y}) < a(x) + b(\bar{y}) - c(\bar{x}, \bar{y})$$

Since $a \in L^1(\mu)$, $\max\{\varphi, 0\} \in L^1(d\mu)$.

Thus we only need to show that

$$\partial^{c+}\varphi \supset \Gamma.$$

To do this, take $(\tilde{x}, \tilde{y}) \in \Gamma$, let

$(x_1, y_1) = (\tilde{x}, \tilde{y})$, by the definition of φ,

one has

$$\begin{aligned}\varphi(x) \le\ & c(x, \tilde{y}) - c(\tilde{x}, \tilde{y}) \\ & + \inf\Big(\big(c(\tilde{x}, y_2) - c(x_2, y_2)\big) \\ & \qquad + \cdots + c(x_m, \bar{y}) - c(\bar{x}, \bar{y})\Big) \\ =\ & c(x, \tilde{y}) - c(\tilde{x}, \tilde{y}) + \varphi(\tilde{x}).\end{aligned}$$

i.e. $$\boxed{\varphi(\tilde{x}) - c(\tilde{x},\tilde{y}) \geq \varphi(x) - c(x,\tilde{y}), \quad \forall x \in X.}$$

$\Downarrow$

Eq (1.3) Characterization of $\tilde{y} \in \partial^{c_+}\varphi(\tilde{x})$.

iii) $\Rightarrow$ i) Let $\tilde{\gamma} \in \Pi(\mu,\nu)$ Admissible

We claim: $\int c\, d\gamma \leq \int c\, d\tilde{\gamma}$.

Since $\mathrm{supp}(\gamma) \subset \partial^{c_+}\varphi$, for any $(x,y) \in \mathrm{supp}(\gamma)$

$$\varphi(x) + \varphi^{c_+}(y) = c(x,y)$$

$$\varphi(x) + \varphi^{c_+}(y) \leq c(x,y), \quad \forall x \in X, y \in Y.$$

Hence

$$\int c(x,y)\, d\gamma(x,y) = \int \left(\varphi(x) + \varphi^{c_+}(y)\right) d\gamma(x,y)$$

$$= \int \varphi(x)\, d\mu(x) + \int \varphi^{c_+}(y)\, d\nu(y)$$

$$= \int \left(\varphi(x) + \varphi^{c_+}(y)\right) d\tilde{\gamma}(x,y) \leq \int c(x,y)\, d\tilde{\gamma}(x,y)$$

Okay. $\boxtimes$

RK:

Consequence:

- $\gamma \in \Pi(\mu,\nu)$ being optimal depends only on $\operatorname{supp}\gamma$, and not on how the mass is distributed.

  For example: if γ is optimal, $\tilde{\gamma} \in \Pi(\mu,\nu)$, s.t.

  $\operatorname{supp}(\tilde{\gamma}) \subset \operatorname{supp}(\gamma)$, then

  $\tilde{\gamma}$ is also optimal.

- Assume $T: X \to Y$ is a map s.t.

  $T(x) \in \partial^{c+}\varphi(x)$, for some c-concave φ for all x.

  Then under proper assumptions ($c \leq a \oplus b$,

  $a \in L^1(d\mu)$, $b \in L^1(d\nu)$), the map T is optimal

  $\nu = T_\#\mu$ between μ and $T_\#\mu$.

  Therefore it makes perfect sense to say that T is an optimal map, without explicit mention to the reference measures.

<u>**Rk:**</u> FT of O.T. tells us

$\forall \gamma \in OPT(\mu,\nu)$, $\exists$ a c-concave φ,

s.t. $supp(\gamma) \subset \partial^{c+}\varphi$.

Even a stronger statement holds:

if $supp(\gamma) \subset \partial^{c+}\varphi$ for some optimal γ,

then $supp(\gamma') \subset \partial^{c+}\varphi$ for every $\gamma' \in OPT(\mu,\nu)$.

$\max\{\varphi, 0\} \in L^1(d\mu)$,

$$\varphi^{c+}(y) = \inf_{x\in X}\left(c(x,y) - \varphi(x)\right)$$

$$\varphi^{c+}(y) \le c(\bar{x},y) - \varphi(\bar{x})$$

$$\le a(\bar{x}) + b(y) - \varphi(\bar{x})$$

i.e. $\max\{\varphi^{c+}, 0\} \in L^1(d\nu)$

Thus it holds

$$\int \varphi \, d\mu + \int \varphi^{c+} d\nu = \int \left(\varphi(x) + \varphi^{c+}(y)\right) d\gamma(x,y)$$

$$\forall \gamma \in \Pi(\mu,\nu)$$

$$\le \int c(x,y)\,d\gamma'(x,y) \;\overset{\text{Both being optimal}}{=}\; \int c(x,y)\,d\gamma(x,y)$$

should be "="

$$\Downarrow$$

γ'-a.e. (x,y), $(x,y) \in \partial^{c}\varphi$, i.e. $\mathrm{supp}(\gamma') \subset \partial^{c}\varphi$. ∎

---

§1.3

The dual Problem

Kantorovich's formulation

$$\text{(K)} \qquad \inf_{\underbrace{\gamma \in \Pi(\mu,\nu)}_{\text{affine constraint, (convex)}}} \underbrace{\int_{X\times Y} c(x,y)\,d\gamma(x,y)}_{=\ \langle c,\gamma\rangle \text{ linear functional}}$$

This kind optimization problem admits a natural dual problem, where we maximize a linear functional with affine constraints.

Kantorovich's problem has a dual problem reads For $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$.

Dual:
$$\sup_{\substack{\varphi(x)+\psi(y) \le c(x,y),\ \forall x,y. \\ \varphi \in L^1(d\mu) \\ \psi \in L^1(d\nu)}} \left\{ \int \varphi(x)\, d\mu(x) + \int \psi(y)\, d\nu(y) \right\}$$

Theorem (Kantorovich duality)

$$\inf_{\gamma \in \Pi(\mu,\nu)} \int c(x,y)\, d\gamma(x,y) = \sup_{\substack{(\varphi,\psi) \\ \varphi \oplus \psi \le c \\ \varphi \in L^1(\mu) \\ \psi \in L^1(\nu)}} \left[\int \varphi(x)\, d\mu(x) + \int \psi(y)\, d\nu(y) \right].$$

(Proof can be deduced by Fundamental theorem of O.T.

$$\inf = \int c(x,y)\, d\gamma(x,y) = \int \left(\varphi(x) + \varphi^{c+}(y) \right) d\gamma(x,y),$$

$\gamma \in Opt(\mu,\nu)$ choosing (φ, φ^{c+}) as the

$supp(\gamma) \subset \partial^{c+}\varphi$. optimal pair.)

Heuristic argument first:

based on the Min-Max Principle.

The constraint $\gamma \in \Pi(\mu,\nu)$ "translates" to the functional to maximize in the dual problem and the functional to minimize $\int c\, d\gamma = \langle c, \gamma\rangle$ becomes the constraint in the dual.

Calculations:

$$\inf_{\gamma \in \Pi(\mu,\nu)} \langle c, \gamma\rangle = \inf_{\gamma \in M_+(X\times Y)} \{\langle c,\gamma\rangle + \chi(\gamma)\},$$

where
$$\chi(\gamma) = \begin{cases} 0 & \text{if } \gamma \in \Pi(\mu,\nu); \\ +\infty & \text{otherwise} \end{cases}$$

Claim: $\chi(\gamma)$ may be written as

$$\chi(\gamma) = \sup_{(\varphi,\psi)} \{\langle \varphi, \mu\rangle + \langle \psi, \nu\rangle - \langle \varphi \oplus \psi, \gamma\rangle\}$$

$C_b(X)$ $C_b(Y)$ (As in Villani's short book)

Thus

$$\inf_{\gamma\in\Pi(\mu,\nu)} \int c(x,y)\, d\gamma(x,y)$$

$$= \inf_{\gamma\in M_+(X\times Y)} \sup_{(\varphi,\psi)} \{ \underbrace{\langle c-\varphi\oplus\psi, \gamma\rangle + \langle\varphi,\mu\rangle + \langle\psi,\nu\rangle}_{= F(\gamma,\varphi,\psi)} \}$$

Since $\gamma\mapsto F(\gamma,\varphi,\psi)$ is convex (actually linear)

and $(\varphi,\psi)\mapsto F(\gamma,\varphi,\psi)$ is concave (actually linear),

the min-max principle holds and we have

$$\inf_{\gamma\in M_+} \sup_{(\varphi,\psi)} F(\gamma,\varphi,\psi)$$

$$= \sup_{(\varphi,\psi)} \inf_{\gamma\in M_+} F(\gamma,\varphi,\psi)$$

$$= \sup_{(\varphi,\psi)} \{ \langle\varphi,\mu\rangle + \langle\psi,\nu\rangle + \boxed{\inf_{\gamma\in M_+} \langle c-(\varphi\oplus\psi), \gamma\rangle} \}$$

$$= \begin{cases} 0 & \text{if } \varphi\oplus\psi\le c \\ -\infty & \text{otherwise} \end{cases}$$

Hence we proved

$$\inf_{\gamma \in \Pi(\mu,\nu)} \int c(x,y)\, d\gamma(x,y) = \sup_{(\varphi,\psi)} \left\{ \int \varphi(x)\, d\mu(x) + \int \psi(y)\, d\nu(y) \right\}.$$

Question: What is the min-max principle?

Is it a vague "thumb-of rule"?

Give some more examples: ...

Let us give a rigorous proof independent of min-max principle.

Thm (Kantorovich duality)

Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and $c : X \times Y \to \mathbb{R}$ a continuous and bounded from below.

Assume that $c(x,y) \le a(x) + b(y)$, $\forall x, \forall y$, for some $a \in L^1(d\mu)$, $b \in L^1(d\nu)$.

Then

$$\inf_{\pi \in \Pi(\mu,\nu)} \int c \, d\gamma(x,y) = \sup_{(\varphi,\psi)} \left\{ \int \varphi(x) d\mu(x) + \int \psi(y) d\nu(y) \right\}.$$

Further, the supremum of the dual problem is attained, and the maximizing couple (φ, ψ) is of the form (φ, φ^{c+}) for some c-concave function φ. (Compare it to Theorem 1.3 in Villani)

Pf: Here we adopt the same assumptions as in Fundamental Theorem of O.T.

For any $\gamma \in \Pi(\mu,\nu)$, $\varphi \in L^1(d\mu)$ $\psi \in L^1(d\nu)$,

and $\varphi \oplus \psi \leq c$ pointwise,

we observe

$$\int c(x,y) \, d\gamma(x,y) \geq \int (\varphi(x) + \psi(y)) d\gamma(x,y)$$

$$= \int \varphi(x) d\mu(x) + \int \psi(y) d\nu(y).$$

i.e.

$$\inf_{\gamma \in \Pi(\mu,\nu)} \int c \, d\gamma \geq \sup_{(\varphi,\psi)} \{ \langle \varphi, \mu \rangle + \langle \psi, \nu \rangle \}.$$

Now it is much easier to prove the converse inequality

Pick $\gamma \in OPT(\mu,\nu)$, then $\exists$ a c-concave function φ, s.t. $\mathrm{supp}(\gamma) \subset \partial^{c+}\varphi$, with $\max\{\varphi, 0\} \in L^1(d\mu)$, and $\max\{\varphi^{c+}, 0\} \in L^1(d\nu)$.

Then

$$\underbrace{\int c(x,y)\, d\gamma(x,y)}_{\substack{\text{finite} \\ \mathbb{R}}} = \int (\varphi(x) + \varphi^{c+}(y))\, d\gamma(x,y)$$

$$= \underbrace{\int \varphi(x)\, d\mu(x)}_{\Downarrow} + \int \varphi^{c+}(y)\, d\nu(y)$$

$\varphi \in L^1(d\mu)$ $\quad$ $\varphi^{c+} \in L^1(d\nu)$

i.e. (φ, φ^{c+}) is an admissible couple (potentials) in the problem.

$\boxtimes$

RK: Under all the assumptions as above, for any c-concave couple of function (φ, φ^{c_+}) maximizing the dual problem, and any optimal plan γ, it holds

$$\mathrm{supp}(\gamma) \subset \partial^{c_+}\varphi.$$

Firstly, we have already known $\exists$ some c-concave φ s.t. $\varphi \in L^1(d\mu)$, $\psi = \varphi^{c_+} \in L^1(d\nu)$ and

$$\mathrm{supp}(\gamma) \subset \partial^{c_+}\varphi, \quad \forall \gamma \in OPT(\mu, \nu).$$

For other maximizing couple $(\tilde{\varphi}, \tilde{\psi})$ for the dual problem,

constraint: $\tilde{\varphi}(x) + \underline{\underline{\tilde{\psi}(y)}} \leq c(x,y)$

$$\tilde{\psi}(y) \leq \tilde{\varphi}^{c_+}(y) = \inf_{x \in X}\{c(x,y) - \tilde{\varphi}(x)\}$$

$\hookrightarrow$ $(\tilde{\varphi}, \tilde{\varphi}^{c_+})$ is a maximizing couple as well.

($\tilde{\varphi}^{c_+} \in L^1(d\nu)$ as well).

Now for any $\gamma \in OPT(\mu, \nu)$,

$$\int \tilde{\varphi}\, d\mu + \int \tilde{\varphi}^{c+} d\nu = \int \varphi\, d\mu + \int \varphi^{c+} d\nu$$

$$= \int \left(\varphi(x) + \varphi^{c+}(y)\right) d\gamma(x,y) \quad \operatorname{supp}(\gamma) \subseteq \partial^{c+}\varphi.$$

$$\Downarrow$$

$$= \int c(x,y)\, d\gamma(x,y)$$

$$\geq \int \tilde{\varphi}\, d\mu + \int \tilde{\varphi}^{c+} d\nu.$$

$$\text{"="} \quad \int \left(\tilde{\varphi}(x) + \tilde{\varphi}^{c+}(y)\right) d\gamma(x,y)$$

$$\Leftrightarrow \quad \operatorname{supp}(\gamma) \subset \partial^{c+}\tilde{\varphi}.$$

okay.

Def: A c-concave function φ s.t.
(φ, φ^{c+}) is a maximizing pair for the dual problem is called a $\boxed{\text{c-concave}}$ Kantorovich potential, (or simply Kantorovich potential), for the couple μ and ν.

A c-convex function φ is called c-convex Kantorovich potential if $-\varphi$ is a c-concave Kantorovich potential.

((c-concave Kantorovich potential

φ is used in.

① $\gamma \in OPT(\mu,\nu)$, $supp(\gamma) \subset \partial^{c+}\varphi$.

② (φ, φ^{c+}) solves the dual problem.))

Lecture 3

Continuation of the Proof.

Step 2: Relax the assumption of compactness

Now: $c \in C_b(X\times Y)$ and uniformly continuous.

Def: $\|c\|_\infty = \sup_{X\times Y} c(x,y)$

Claim: There exists an optimal transference plan π_* for the Kantorovich problem, i.e.

$$I[\pi_*] = \inf_{\pi\in\Pi(\mu,\nu)} I[\pi].$$

($c\in L^\infty \Rightarrow$ infimum is finite. $(0\le c\le L_0)$

the existence of $\pi_* \Leftarrow$ compactness of $\Pi(\mu,\nu)$,

we will prove it in step 3.)

Choose $\delta>0$ arbitrarily small.

X, Y Polish (Complete, seperable) $\Rightarrow X\times Y$ Polish.

Hence π_* is tight.

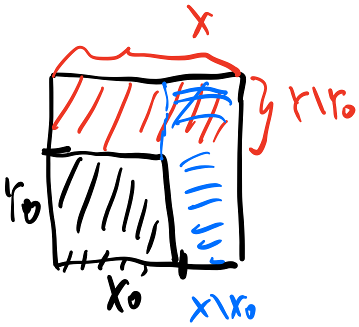
Since $\mu\in P(X)$, $\nu\in P(Y)$, μ (or ν) as a set is tight.

Hence $\exists\ X_0\subset X$, $Y_0\subset Y$ cpt, s.t.

$$\mu[X\setminus X_0]\le\delta,\quad \nu[Y\setminus Y_0]\le\delta$$

Thus ($\pi_* \in \Pi(\mu,\nu)$ of course)

$$\pi_*[(X\times Y)\setminus(X_0\times Y_0)] \le \pi_*[X\times(Y\setminus Y_0)] + \pi_*[(X\setminus X_0)\times Y] \le 2\delta$$



Def: $$\pi_{*0} = \frac{1_{X_0\times Y_0}}{\pi_*[X_0\times Y_0]}\,\pi_*$$

(Probability measure supported on $X_0\times Y_0$)

$\mu_0 = (P_X)_\# \pi_{*0}$, $\nu_0 = (P_Y)_\# \pi_{*0}$, marginals

$\Pi_0(\mu_0,\nu_0) \triangleq \{\tilde{\pi} \in P(X_0\times Y_0) \mid \tilde{\pi}$ has marginals μ_0, ν_0 respectively$\}$.

Def: $$I_0[\pi_0] = \int_{X_0\times Y_0} c(x,y)\,d\pi_0(x,y).$$

Let $\tilde{\pi}_0 \in \Pi_0(\mu_0,\nu_0)$ be one optimal transference plan solving the restrictive Kantorovich's problem

$$\inf_{\pi_0\in\Pi_0(\mu_0,\nu_0)} I_0[\pi_0],$$

i.e. $$I_0[\tilde{\pi}_0] = \inf_{\pi_0\in\Pi_0(\mu_0,\nu_0)} I_0[\pi_0].$$

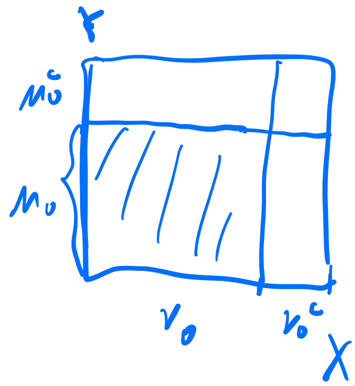
Checking.

Construct a $\tilde{\pi} \in \Pi(\mu, \nu)$ as

$$\tilde{\pi} = \pi_*[X_0 \times Y_0]\, \tilde{\pi}_0 + 1_{(X_0 \times Y_0)^c}\, \pi_*$$

(Idea:

Construct an approximation $\tilde{\pi}$ from a local optimizer!).



So

$$I[\tilde{\pi}] = \langle c, \tilde{\pi} \rangle_{X \times Y}$$

$$= \pi_*[X_0 \times Y_0]\, I_0[\tilde{\pi}_0] + \int_{(X_0 \times Y_0)^c} c(x,y)\, d\pi(x,y)$$

$$\leq I_0[\tilde{\pi}_0] + 2\|c\|_\infty \delta.$$

It follows that

$$\inf_{\pi \in \Pi(\mu,\nu)} I[\pi] \leq \inf I_0 + 2\|c\|_\infty \delta.$$

Now introduce the functional

$$J_0(\varphi_0, \psi_0) = \int_{X_0} \varphi_0 \, d\mu_0 + \int_{Y_0} \psi_0 \, d\nu_0 ,$$

defined on $L^1(d\mu_0) \times L^1(d\nu_0)$.

By the Prop 2,

$$\inf I_0 = \sup J_0,$$

where the supremum runs over all admissable couples $(\varphi_0, \psi_0) \in L^1(d\mu_0) \times L^1(d\nu_0)$, s.t.

$$\varphi_0(x) + \psi_0(y) \le c(x, y), \quad \text{a.e. } x, y.$$

in measure sense.

In particular, $\exists\, (\tilde{\varphi}_0, \tilde{\psi}_0)$ s.t.

$$J_0(\tilde{\psi}_0, \tilde{\psi}_0) \ge \sup J_0 - \delta$$

Now: construct a couple (φ, ψ) from $(\tilde{\varphi}_0, \tilde{\psi}_0)$, which would be a "very good approximation of maximization of $J(\varphi, \psi)$.

WLOG, assume that

$$\tilde{\varphi}_0(x) + \tilde{\psi}_0(y) \le c(x, y), \text{ for all } x, y$$

(Just take $\tilde{\varphi}_0(x) = -\infty$, or $\tilde{\psi}_0(y) = -\infty$ for those $x \in N_x^c$, $y \in N_y^c$).

Firstly, control $\tilde{\varphi}_0$, $\tilde{\psi}_0$ from below at some point in $X \times Y$. WLOG, assume $\delta \le 1$. Since $J_0(0, 0) = 0$, we have $\sup J_0 \ge 0$.

hence $J_0(\tilde{\varphi}_0, \tilde{\psi}_0) \ge -\delta \ge -1$.

$$\Big\| \int_{X_0 \times Y_0} (\tilde{\varphi}_0(x) + \tilde{\psi}_0(y))\, d\pi_0(x, y) \ge -1$$

$\Rightarrow$ $\exists$ $(x_0, y_0) \in X_0 \times Y_0$, s.t.

$$\tilde{\varphi}_0(x_0) + \tilde{\psi}_0(y_0) \geq -1.$$

Note: if we replace $(\tilde{\varphi}_0, \tilde{\psi}_0)$ by $(\tilde{\varphi}_0 + s, \tilde{\psi}_0 - s)$ for some $s \in \mathbb{R}$, we do not change the value of $J_0(\tilde{\varphi}_0, \tilde{\psi}_0)$, and the resulting couple is $\checkmark$.

Up to a constant s, we can ensure.

$$\tilde{\varphi}_0(x_0) \geq -\tfrac{1}{2} \qquad \tilde{\psi}_0(y_0) \geq -\tfrac{1}{2}.$$

Hence

for all $(x, y) \in X_0 \times Y_0$,

$$\tilde{\varphi}_0(x) \leq c(x, y_0) - \tilde{\psi}_0(y_0) \leq c(x, y_0) + \tfrac{1}{2},$$

$$\tilde{\psi}_0(y) \leq c(x_0, y) - \tilde{\varphi}_0(x_0) \leq c(x_0, y) + \tfrac{1}{2}.$$

"Rüschendorf's trick" for improving admissable pairs:

Previously

$$\tilde{\varphi}_0(x) \leq \boxed{c(x, y) - \tilde{\psi}_0(y)}$$

Now define: for $x \in X$

$$\bar{\varphi}_0(x) := \inf_{y \in Y_0} \{c(x, y) - \tilde{\psi}_0(y)\}.$$

Hence $\boxed{\tilde{\varphi}_0 \leq \bar{\varphi}_0}$ on X_0

it follows $J(\bar{\varphi}_0, \tilde{\psi}_0) \geq J(\tilde{\varphi}_0, \tilde{\psi}_0)$.

Control for $\bar{\varphi}_0$ for $x \in X$:

$$\bar{\varphi}_0(x) = \inf_{y \in Y_0} [c(x,y) - \underbrace{\tilde{\psi}_0(y)}_{\geq -c(x_0,y) - 1/2}]$$

$(x \in X)$

$$= \inf_{y \in Y_0} [c(x,y) - c(x_0,y)] - 1/2$$

$$\bar{\varphi}_0(x) \leq c(x, y_0) - \underbrace{\tilde{\psi}(y_0)}_{\geq -1/2} \leq c(x, y_0) + \frac{1}{2}.$$

(again $x \in X$)

Now we define for $y \in Y$,

$$\bar{\psi}_0(y) = \inf_{x \in X} [c(x,y) - \bar{\varphi}_0(x)],$$

Again: $\bar{\varphi}_0(x) + \bar{\psi}_0(y) \leq c(x,y)$,

i.e. $(\bar{\varphi}_0, \bar{\psi}_0) \in \Phi_c$

And $J_0(\bar{\varphi}_0, \bar{\psi}_0) \geq J_0(\bar{\varphi}_0, \tilde{\psi}_0) \geq J_0(\tilde{\varphi}_0, \tilde{\psi}_0)$

($\Uparrow$ clear. $(\bar{\varphi}_0 \geq \tilde{\varphi}_0$ on $X_0)$

$$\bar{\varphi}_0(x) \leq c(x,y) - \tilde{\psi}_0(y) \quad \forall y \in Y_0$$

i.e. $\tilde{\psi}_0(y) \leq \underbrace{c(x,y) - \bar{\varphi}_0(x)}_{}$ $\forall y \in Y_0$ $\forall x \in X$.

$$\Rightarrow \tilde{\psi}_0(y) \leq \inf_{x \in X} [\;\cdot\;] = \bar{\psi}_0(y_0) \quad)$$

Moreover, for $y \in Y$,

$$\bar{\psi}_0(y) \geq \inf_{x \in X} \left[c(x,y) - c(x,y_0) \right] - 1/2.$$

(since $\bar{\varphi}_0(x) \leq c(x,y_0) + 1/2$)

In particular,

$$\bar{\varphi}_0(x) \geq -\|c\|_\infty - 1/2$$
$$\bar{\psi}_0(y) \geq -\|c\|_\infty - 1/2.$$

Once we have those bounds, we are almost done!

Indeed,

$$J(\bar{\varphi}_0, \bar{\psi}_0) = \int_X \bar{\varphi}_0 \, d\mu + \int_Y \bar{\psi}_0 \, d\nu$$

$$= \int_{X \times Y} \left[\bar{\varphi}_0(x) + \bar{\psi}_0(y) \right] d\pi_*(x,y) \quad \left(\begin{array}{l} \text{As long as} \\ \pi_* \in \Pi(\mu,\nu) \end{array} \right)$$

$$= \pi_*[X_0 \times Y_0] \int_{X_0 \times Y_0} \left[\bar{\varphi}_0(x) + \bar{\psi}_0(y) \right] d\pi_{*0}(x,y)$$

$$+ \int_{(X_0 \times Y_0)^c} \left[\bar{\varphi}_0(x) + \bar{\psi}_0(y) \right] d\pi_*(x,y)$$

$$\geq (1-2\delta) \left(\int_{X_0} \bar{\varphi}_0 \, d\mu_0 + \int_{Y_0} \bar{\psi}_0 \, d\nu_0 \right) - (2\|c\|_\infty + 1) \, \pi_*[(X_0 \times Y_0)^c]$$

$$\geq (1-2\delta) \, J_0(\bar{\varphi}_0, \bar{\psi}_0) - 2(2\|c\|_\infty + 1)\, \delta$$

$$\geq (1-2\delta)\underbrace{J_0(\tilde{\varphi}_0, \tilde{\psi}_0)}_{\geq \sup J_0 - \delta = \inf I_0 - \delta} - 2(2\|c\|_\infty + 1)\delta$$

$$\inf I_0 \geq \inf_{\lambda \in \Pi(\mu,\nu)} I[\lambda] - (2\|c\|_\infty + 1)\delta.$$

$$\geq (1-2\delta)\left(\inf I - (2\|c\|_\infty+1)\delta\right) - 2(2\|c\|_\infty + 1)\delta.$$

Since δ is arbitrarily small, we conclude that

$$\boxed{\sup J(\varphi, \psi) \geq \inf I} \quad \star$$

okay.

RK: Since c is uniform continuous,

$$\bar{\varphi}_0(x) \triangleq \inf_{y \in Y_0} \left[c(x,y) - \tilde{\psi}_0(y) \right]$$

$$\bar{\psi}_0(y) \triangleq \inf_{x \in X} \left[c(x,y) - \bar{\varphi}_0(x) \right].$$

are uniformly continuous on the whole X (Good exercise)

and Y respectively.

Therefore, it does not matter whether the supremum of J is taken over $\Phi_c \cap L^1$ over $\Phi_c \cap C_b$.

Step II: Approximating the cost fun. c.

Write $c = \sup_n c_n$,

$0 \le c_1 \le c_2 \le \cdots$

where c_n is non-decreasing sequence of non-negative, uniformly continuous functions;

Upon replacing c_n by $\min\{c_n, n\}$, one can assume each c_n is bounded.

The following is the standard approximation technique.

Define $I_n[\pi] = \int_{X\times Y} c_n \, d\pi$, $\pi \in \Pi(\mu,\nu)$.

From Step II,

$$\inf_{\pi\in\Pi(\mu,\nu)} I_n[\pi] = \sup_{(\varphi,\psi)\in\Phi_{c_n}} J(\varphi,\psi).$$

We will conclude by showing that.

$$(*) \quad \inf_{\pi\in\Pi(\mu,\nu)} I[\pi] = \sup_n \inf_{\pi\in\Pi(\mu,\nu)} I_n[\pi]$$

Indeed, for each n,

$$\sup_{(\varphi,\psi)\in\Phi_{c_n}} J(\varphi,\psi) \le \sup_{(\varphi,\psi)\in\Phi_{c}} J(\varphi,\psi).$$

$\shortparallel$

$\Leftarrow$) $\sup_n \inf_\pi I_n[\pi] \leq \sup_{(\varphi,\psi)\in\Phi_c} J(\varphi,\psi)$

If (*) holds true, then

$$\inf_{\pi\in\Pi(\mu,\nu)} I[\pi] \leq \sup_{(\varphi,\psi)\in\Phi_c} J(\varphi,\psi)$$

while the other direction is trivial.
Hence, we only need to show Eq. (*).

Note. I_n is a nondecreasing sequence of functionals, $(I_n[\pi] \leq I_{n+1}[\pi] \leq \cdots \leq I[\pi])$
so $\inf I_n \nearrow$ and bounded above by $\inf I$.
Thus, we only have to prove that

$$\lim_{n\to\infty}\left(\inf_\pi I_n[\pi]\right) \geq \inf_{\pi\in\Pi(\mu,\nu)} I[\pi].$$

- $\Pi(\mu,\nu)$ is tight.

($\Leftarrow$ Since both μ and ν as a set is tight,

$\forall \varepsilon>0$, $\exists$ $K_\varepsilon \subset X$, $L_\varepsilon \subset Y$ cpt, s.t.

$\mu[X\setminus K_\varepsilon] < \varepsilon/2$, $\nu[Y\setminus L_\varepsilon] < \varepsilon/2$

Then for any $\pi \in \Pi(\mu,\nu)$,

$$\pi\left[(K_\varepsilon \times L_\varepsilon)^c\right] \leq \pi[K_\varepsilon^c \times Y] + \pi[X \times L_\varepsilon^c]$$
$$= \mu[K_\varepsilon^c] + \nu[L_\varepsilon^c] \leq \varepsilon.$$

Prokhorov's theorem $\Rightarrow$ $\Pi(\mu,\nu)$ is relatively cpt. for the weak topology

That is, if $(\pi_n^k)_{k\in\mathbb{N}}$ is a minimizing sequence for the problem $\inf I_n[\pi]$, then up to extraction of a subsequence, π_n^k converges weakly to $\pi_n \in \mathcal{P}(X\times Y)$, i.e. for $\theta \in C_b(X\times Y)$, (narrowly)

$$\int_{X\times Y} \theta(x,y)\, d\pi_n^k(x,y) \xrightarrow{k\to\infty} \int_{X\times Y} \theta(x,y)\, d\pi_n(x,y).$$

From this, one obtains immediately

- $\pi_n \in \Pi(\mu,\nu)$ (taking $\theta(x,y) = \varphi(x) + \psi(y)$);
- $\inf I_n = \lim_{k\to\infty} \int c_n\, d\pi_n^k = \int c_n\, d\pi_n$,

which shows the existence of a minimizing probability measure π_n. (Go back to the proof in Step II).

Similarly, those optimizers π_n for $\inf_\pi I_n[\pi]$

admits a cluster point π_* by compactness of $\Pi(\mu,\nu)$ as well.

Whenever $n \geq m$, one has

$$I_n[\pi_n] \geq I_m[\pi_n]. \quad (c_n \geq c_m)$$

By continuity of I_m,

$$\lim_{n\to\infty} I_n[\pi_n] \geq \overline{\lim_{n\to\infty}} I_m[\pi_n] \geq I_m[\pi_*]$$

one cluster of $\{\pi_n\}$.

By monotone convergence theorem,

$$I_m[\pi_*] \to I[\pi_*], \quad \text{as } m \to \infty.$$

So.

$$\lim_{n\to\infty} I_n[\pi_n] \geq \lim_{m\to\infty} I_m[\pi_*] = I[\pi_*]$$

$$\geq \inf_{\pi\in\Pi(\mu,\nu)} I[\pi].$$

which then concludes

$$\inf_{\pi\in\Pi(\mu,\nu)} I[\pi] = \sup_n \inf_{\pi\in\Pi(\mu,\nu)} I_n[\pi].$$

Note that: π_* above is actually one of the minimizers.

Step IV: Let us confirm again that the infimum is attained.

Again this is a consequence of the compactness of $\Pi(\mu,\nu)$.

Taking (π_k) a minimizing sequence of $I[\pi]$, and let π_* be any weak cluster of (π_k); then

$$
\begin{aligned}
I[\pi_*] &= \lim_{n\to\infty} I_n[\pi_*] \quad \text{(Monotone Convergence)}\\
&\le \lim_{n\to\infty} \limsup_{k\to\infty} I_n[\pi_k] \quad \le I[\pi_k]\\
&\le \limsup_{k\to\infty} I[\pi_k] = \inf I \quad \text{(Minimizing sequence)}.
\end{aligned}
$$

$\square$

RK: (c-concave functions)

When $c \in L^\infty$, one can restrict the supremum in

$$
\text{Eq}(\mu,\nu) \quad \inf_{\Pi(\mu,\nu)} I[\pi] = \sup_{\Phi_c} J(\varphi,\psi)
$$

to those pairs $(\varphi^{cc}, \varphi^c)$, where φ is bounded and

$$
\varphi^c(y) = \inf_{x\in X} [c(x,y) - \varphi(x)]
$$

(Compared to Legendre transform)

$$\varphi^{cc}(x) = \inf_{y \in Y} [c(x,y) - \varphi^c(y)].$$

Checking that: $(\varphi^{cc})^c = \varphi^c$.

The pair $(\varphi^{cc}, \varphi^c)$: a pair of conjugate c-concave functions.

Note that φ^c is measurable, since it can be written as $\lim_{L \to \infty} \psi_L$, where

$$\psi_L(y) = \inf_{x \in X} [c_L(x,y) - \varphi(x)],$$

and $c_L \nearrow$ bounded, uniformly continuous. $c_L(x) \to c(x)$. pointwise.

More discussions on Kantorovich duality via the study of c-concave functions. in Chapter 2.

See also User's guide to OT. by Ambrosio and Gigli.

RK: In the case when $0 \le c \le b$, $(c \in L^\infty)$, it is useful to note that the supremum can be further restricted:

$$\sup\{J(\varphi,\psi) : (\varphi,\psi)\in\Phi_c\}$$

$$= \sup\{J(\varphi,\psi) : (\varphi,\psi)\in\Phi_c,\ 0\le\varphi\le\|c\|_\infty,\ -\|c\|_\infty\le\psi\le 0.\}$$

Lecture 4

- Fundamental Theorem of O.T.

Follow Ch1 of "Users' guide to Optimal Transport". by Ambrosio and Gigli

Notions:

Cyclical Monotonicity

A set $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ is said to be cyclically monotone,

if $\forall N = 1, 2, 3, \dots$, $(x_i, y_i) \in \Gamma$, $i = 1, \dots N$,

(*) $$\sum_{i=1}^{N} \underbrace{\langle x_i, y_i \rangle}_{x_i \cdot y_i} \geq \sum_{i=1}^{N} \langle x_i, y_{\sigma(i)} \rangle,$$

$\forall \sigma \in S_N$ (any permutation)

A classical Result in Convex analysis:

Rochafeller theorem:

A Set $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ is cyclically monotone,

$\iff$ $\exists$ a convex and l.s.c. $\varphi : \mathbb{R}^d \to \mathbb{R} \cup$

$\{+\infty\}$ $(\varphi \not\equiv +\infty)$, st

$$\Gamma \subset \text{Graph}(\partial\varphi)$$

subdifferential

Def:

$$\partial\varphi \triangleq \{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d \mid \varphi(x) + \varphi^*(y) = \langle x, y\rangle\}$$

Legendre transform

Subdifferential at x

$$\partial\varphi(x) = \{y \in \mathbb{R}^d \mid \varphi(x) + \varphi^*(y) = \langle x, y\rangle\}$$

Recall $\varphi^*(y) = \sup_{x \in \mathbb{R}^d} \{\underbrace{\langle x, y\rangle}_{\text{linear}} - \underbrace{\varphi(x)}_{\text{convex}}\}$

$y \in \mathbb{R}^d$

φ^*: <u>l.s.c.</u> (a supremum of a family of l.s.c. functions)

<u>convex</u> $\sup_{\alpha} \underbrace{f_\alpha(y)}_{\text{convex}} = \text{convex}$

We have

① $\forall y, \forall z, \quad \varphi(z) + \varphi^*(y) \geq \langle z, y\rangle$

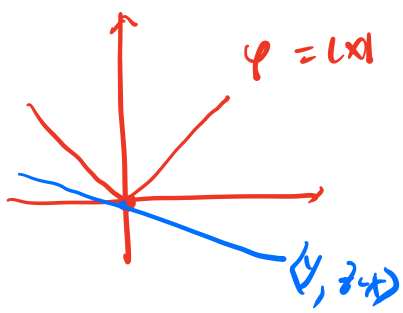
② $\varphi(x) + \varphi^*(y) = \langle x, y\rangle, \ y \in \partial\varphi(x)$

① $\Rightarrow \varphi^*(y) \geq \langle z, y\rangle - \varphi(z) \quad \forall\ z \in \mathbb{R}^d$

$\varphi^*(y) = \langle x, y\rangle - \varphi(x)$

Then $\langle x, y\rangle - \varphi(x) \geq \langle z, y\rangle - \varphi(z), \quad \forall z$

$$\boxed{\forall z, \quad \varphi(z) \geq \underbrace{\varphi(x) + \langle y, z-x\rangle}_{\text{supporting line}}, \quad \text{for } y \in \partial\varphi(x)}$$



$x = 0$

$y \in \partial\varphi(0) = [-1, 1]$

<u>Convex function</u>: $\varphi : \mathbb{R}^d \to \mathbb{R}$.

$\varphi \in C^1$

$$\varphi(z) \geq \underbrace{\varphi(x) + \langle \nabla\varphi(x), z-x\rangle}_{\text{supporting line}}$$

Generalization of convex functional

Take $c(x,y) = \frac{1}{2}|x-y|^2$ or $-\langle x, y\rangle$ in $\mathbb{R}^d \times \mathbb{R}^d$

$\frac{1}{2}|x|^2$

for $\pi \in \Pi(\mu,\nu)$, $\int c \, d\pi = \int \frac{1}{2}|x-y|^2 \, d\pi(x,y)$

fixed

$$= \boxed{\frac{1}{2}\int |x|^2 \, d\mu(x) + \frac{1}{2}\int |y|^2 \, d\nu(x)} - \int \langle x, y\rangle \, d\pi(x,y)$$

• Def (c- cyclical monotone)

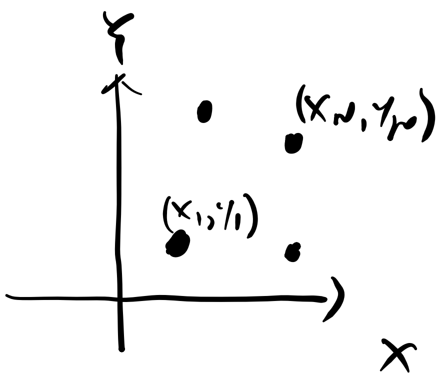
Change the (*) above to:

$\Gamma \subset X \times Y$, for any N, any $(x_i, y_i)_{i=1}^N \in \Gamma$,

we have:

$$\sum_{i=1}^N c(x_i, y_i) \le \sum_{i=1}^N c(x_i, y_{\sigma(i)}),$$

$$\forall \sigma \in S_N$$



• Def: $\boxed{\text{c-transform}}$

($c: X \times Y \to [0, +\infty)$ & c: continuous
(sometimes can be $+\infty$)

But don't need such strong assumption)

• $\underline{\underline{\varphi}}: X \to \mathbb{R} \cup \{\pm\infty\}$ but proper ($\varphi(x_0) \in \mathbb{R}$
$\exists x_0$)

$$\varphi^c(y) = \inf_{x \in X} \{ c(x,y) - \varphi(x) \}$$

(Go back to classical one:

$c = \frac{1}{2}|x-y|^2$ (or $c = -\langle x, y \rangle$) $X = Y = \mathbb{R}^d$

$$\varphi^c(y) = \inf_{x \in \mathbb{R}^d} \{ \frac{1}{2}|x-y|^2 - \varphi(x) \}$$

$$= \inf_x \{ \underline{\underline{\frac{1}{2}|x|^2}} - x \cdot y + \frac{1}{2}y^2 - \underline{\underline{\varphi(x)}} \}$$

$$= \frac{1}{2}y^2 - \sup_x \{ x \cdot y - (\frac{1}{2}|x|^2 - \varphi(x)) \}$$

i.e. $\frac{1}{2}y^2 - \varphi^c(y) = (\frac{1}{2}|x|^2 - \varphi(x))^*(y)$

(Somehow explain why $c = \frac{1}{2}|x-y|^2$ has better properties.)

Def: A function $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ is <u>c-concave</u>

if $\exists\, \psi : Y \to \mathbb{R} \cup \{\pm\infty\}$, $\varphi = \psi^c$.

$$\left(\frac{1}{2}|y|^2 - \varphi^c = \left(\frac{1}{2}|x|^2 - \varphi(x)\right)^*(y) \right)$$

φ^c is c-concave $\Leftrightarrow$ $\frac{1}{2}|y|^2 - \varphi^c$ is convex

Or

φ is c-concave

$\Updownarrow$ for $c = \frac{1}{2}|x-y|^2$

$\frac{1}{2}|y|^2 - \varphi(y)$ is convex.

Claim: $\boxed{\psi^c = \psi^{ccc}}$ $\leftarrow$ for any ψ. $(\varphi^* = \varphi^{***})$

Exercise.

Exercise. φ - convex $\Updownarrow$ $\varphi = g^*$

and

$$\boxed{\begin{gathered}\psi \text{ is c-concave} \\ \Updownarrow \\ \psi^{cc} = \psi\end{gathered}}$$

($\Leftarrow$ Trivial since $\psi = \varphi^c$, $\psi^{cc} = \varphi^{ccc} = \varphi^c = \psi$)

(φ - convex $\varphi^{**} = \varphi$)

Pf: Assume $\psi : Y \to \mathbb{R} \cup \{\pm\infty\}$,

$$\psi^c(x) = \inf_{y \in Y} \left[c(x,y) - \psi(y) \right]$$

$$\psi^c : X \to \overline{\mathbb{R}}$$

$$\psi^{cc} : Y \to \overline{\mathbb{R}}, \quad \psi^{ccc} : X \to \overline{\mathbb{R}}$$

$$\psi^{ccc}(x) = \inf_{\bar y \in Y} \Big\{ c(x, \bar y) - \underbrace{\psi^{cc}(\bar y)}_{\inf_{\bar x \in X} [c(\bar x, \bar y) - \psi^c(\bar x)]} \Big\}$$

$$= \inf_{\bar y \in Y} \sup_{\bar x \in X} \left\{ c(x, \bar y) - c(\bar x, \bar y) + \psi^c(\bar x) \right\}$$

$$\psi^{ccc}(x) = \inf_{\bar y \in \bar Y} \sup_{\bar x \in X} \inf_{y \in Y} \Big\{ c(\underset{\bar x}{x}, \overset{y}{\bar y}) - c(\bar x, \overset{y}{\bar y}) + c(\overset{x}{\bar x}, y) - \psi(y) \Big\}$$

Choose $\bar x = x$,

$$\psi^{ccc}(x) \geq \inf_{y \in Y} \left\{ c(x,y) - \psi(y) \right\} = \psi^c(x)$$

Choose $y = \bar y$,

$$\psi^{ccc}(x) \leq \inf_{y} \{ \dots \} = \psi^c(x)$$

i.e. $\psi^{ccc} = \psi^{c}$. $\square$

Def: (c-superdifferential
(sub-...)
for a c-concave function φ)
(c-convex)

$\partial^{c+}\varphi \subset X \times Y$
$(\partial^{c-}\varphi)$

$$\partial^{c+}\varphi = \{(x,y) \in X \times Y \mid \varphi(x) + \varphi^{c}(y) = c(x,y)\}$$
$(-c(x,y))$

$\partial^{c}\varphi$, $\partial^{c}\varphi(x)$

$$\partial^{c}\varphi(x) = \{y \in Y \mid (x,y) \in \partial^{c}\varphi(x)\}$$
$$= \{y \in Y \mid \underline{\varphi(x) + \varphi^{c}(y) = c(x,y)}\}$$

$\underline{(\varphi, \varphi^{c})}$ will be an admissible pair for the dual problem (if we don't consider the integrability)

$(\varphi, \varphi^{c}) \in \Phi_c = \{(\varphi, \psi) \mid \varphi \oplus \psi \le c\}$

$$\varphi^c(y) = \inf_{x \in X} \{ c(x,y) - \varphi(x) \}$$

i.e. $\forall x, \forall y,$

$$\varphi^c(y) \le c(x,y) - \varphi(x)$$

or ① $$\boxed{\varphi(\cancel{x}\,z) + \varphi^c(y) \le c(\cancel{x}\,z, y) \quad \cancel{\forall x}\ \forall z, \forall y}$$

$\underline{\underline{\Phi_c}}$: $(\varphi, \varphi^c) \in \Phi_c$ (NEED check integrability)

If further $\boxed{y \in \partial^c \varphi(x),}$

② $$\boxed{\varphi(x) + \varphi^c(y) = c(x,y)} \longleftarrow \text{NOT for all } x, y$$

Characterization of $(x,y) \in \partial^c \varphi$,

$\Updownarrow$

or $y \in \partial^c \varphi(x)$

$$\Longleftrightarrow \ ☆ \begin{cases} \varphi(z) + \varphi^c(y) \le c(z,y), & \forall z \in X \\ \varphi(x) + \varphi^c(y) = c(x,y). \end{cases}$$

$$\Longleftrightarrow \quad \cancel{\varphi^c(y)} = c(x,y) - \varphi(x) \le c(z,y) - \varphi(z), \quad \forall z \in X$$

$$\Longleftrightarrow \quad ☆ \ \underline{\varphi(x) - c(x,y) \ge \varphi(z) - c(z,y)}, \quad \forall z \in X.$$

Observation:

φ : c-concave $\Rightarrow$ $\partial^c\varphi \subset X\times Y$ is c-cyclically monotone

Pf: $(x_i, y_i)_{i=1}^N \subset \partial^c\varphi$ $\left(\begin{array}{l} \forall i \in [N] = \{1,2,\dots N\} \\ y_i \in \partial^c\varphi(x_i) \end{array}\right.$

$$\sum_i c(x_i, y_i) = \sum_i \left(\varphi(x_i) + \varphi^c(y_i)\right)$$

$$= \sum_i \left(\varphi(x_i) + \varphi^c(y_{\sigma(i)})\right), \quad \forall \sigma \in S_N$$

$$\leq \sum_i c(x_i, y_{\sigma(i)})$$

$\boxtimes$

Thm (Fundamental Theorem of O.T.)

Assume $c : X\times Y \to [0, +\infty)$, c-continuous,

And $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, X, Y Polish,

$$c(x,y) \leq a(x) + b(y),$$

$$a \in L^1(d\mu)$$

$$b \in L^1(d\nu),$$

Thm Let $\gamma \in \Pi(\mu, \nu)$, TFAE
(下面等价)

i) The plan γ is optimal;

ii) $\mathrm{supp}(\gamma)$ is c-cyclically monotone;

$\Downarrow$

iii) $\exists$ a c-concave function φ, s.t.
$\max\{\varphi, 0\} \in L^1(d\mu)$, $\mathrm{supp}(\gamma) \subset \partial^c \varphi$.

Pf: $\forall\ \tilde{\gamma} \in \Pi(\mu, \nu)$,

$$\int \underbrace{c(x,y)}_{\substack{c \geq 0 \\ (c \geq -L)}} d\tilde{\gamma}(x,y) \leq \int (a(x) + b(y))\, d\tilde{\gamma}(x,y)$$

$$= \underbrace{\int_X a(x)\, d\mu(x)}_{} + \underbrace{\int_Y b(y)\, d\nu(y)}_{} < +\infty$$

(somehow moment assumption on μ)

Of course, $c \in L^1(d\tilde{\gamma})$

iii) $\Rightarrow$ i) If $\mathrm{supp}(\gamma) \subseteq \partial^c \varphi$, for some $\gamma \in \Pi(\mu,\nu)$, φ : c-concave,

then $\gamma \in \mathrm{OPT}(\mu,\nu)$.

$\Updownarrow$

Only NEED TO SHOW: $\boxed{\langle c, \gamma \rangle \le \langle c, \tilde{\gamma} \rangle, \ \forall \tilde{\gamma} \in \Pi(\mu,\nu)}$

$\forall (x,y) \in \mathrm{supp}(\gamma)$, $\underline{\varphi(x) + \varphi^c(y) = c(x,y)}$ (A)

But for all $(x,y) \in X \times Y$, one has

$$\underline{\varphi(x) + \varphi^c(y) \overset{(B)}{\le} c(x,y)}. \quad (\varphi, \varphi^c \overset{??}{\in} \Phi_c \text{ check integrability})$$

Hence $\langle c, \gamma \rangle$

$$\| $$

$$\int c(x,y)\, d\gamma(x,y) \overset{(A)}{=} \int (\varphi(x) + \varphi^c(y))\, d\gamma(x,y)$$

$$= \int \varphi(x)\, d\mu(x) + \int \varphi^c(y)\, d\nu(y) \quad (\gamma \in \Pi(\mu,\nu))$$

$$= \int (\varphi(x) + \varphi^c(y))\, d\tilde{\gamma}(x,y) \quad (\forall \tilde{\gamma} \in \Pi(\mu,\nu))$$

$$\overset{(B)}{\le} \int c(x,y)\, d\tilde{\gamma}(x,y)$$

($0 \le c$, $c \in L^1(\tilde{\gamma})$, $\forall \tilde{\gamma} \in \Pi(\mu,\nu)$,

Lecture 5

Observations:

$(\tilde{\varphi}, \tilde{\psi}) \in \Phi_c$, i.e. $\tilde{\varphi}(x) + \tilde{\psi}(y) \leq c(x,y)$ — $\mu\otimes\nu$ - a.e.

$\langle \tilde{\varphi}, \mu \rangle + \langle \tilde{\psi}, \nu \rangle \in \mathbb{R}$

If $c(x,y) \leq c_X(x) + c_Y(y)$, with

$$c_X \in L^1(d\mu), \quad c_Y \in L^1(d\nu),$$

then $\exists\, a \in \mathbb{R}$, s.t.

$$(\tilde{\varphi}^{cc} - a, \ \tilde{\varphi}^{c} + a) \in \Phi_c^{\pi}, \quad \begin{array}{l} \varphi \in L^1(d\mu) \\ \psi \in L^1(d\nu) \end{array}$$

$$(\varphi, \quad \psi)$$

and
$$\begin{cases} \underline{\underline{\varphi \leq c_X}} & \text{- pointwise} \\ \underline{\underline{\psi \leq c_Y}} & \text{- pointwise} \end{cases}$$

Checking: $(\tilde{\varphi}, \tilde{\psi}) \in \Phi_c$

$\hookrightarrow \underline{\underline{\tilde{\varphi}(x)}} + \tilde{\psi}(y) \leq c(x,y) \quad \mu\otimes\nu$ - a.e.

$$\tilde{\psi}(y) \le c(x,y) - \tilde{\varphi}(x) \quad \forall x.$$

$$\tilde{\psi}(y) \le \inf_{x \in X} \left(c(x,y) - \tilde{\varphi}(x) \right) \triangleq \tilde{\varphi}^c(y)$$

$$(\tilde{\varphi}, \tilde{\psi}) \rightsquigarrow (\tilde{\varphi}, \tilde{\varphi}^c) \rightsquigarrow (\tilde{\varphi}^{cc}, \tilde{\varphi}^c)$$

$$\tilde{\varphi}^{cc}(x) + \tilde{\varphi}^c(y) \le c(x,y) \quad \checkmark$$

- $$\underset{\substack{\rotatebox{90}{$\le$} \\ C_X(x)}}{\left(\tilde{\varphi}^{cc}(x) - a\right)} + \underset{\substack{\rotatebox{90}{$\le$} \\ C_Y(y)}}{\underline{\left(\tilde{\varphi}^c(y) + a\right)}} \le c(x,y) \quad \forall a \in \mathbb{R}.$$

$\Uparrow$ We need impose

$$\tilde{\varphi}^c(y) + a \le C_Y(y)$$

$$a \le C_Y(y) - \tilde{\varphi}^c(y) \quad \forall y$$

Take $$a = \inf_{y \in Y} \left(C_Y(y) - \tilde{\varphi}^c(y) \right) \underline{\underline{\in \mathbb{R}}}$$

(Need to check, but we omit it).

Now we have

$$\underline{\tilde{\varphi}^{cc}(x) - a - C_X(x)} \le 0$$

$$= \inf_{y \in Y} \Big(\underbrace{c(x,y)}_{\le \cancel{C_X(x)} + C_Y(y)} - \tilde{\varphi}^{c}(y) \Big) - a - \cancel{C_X(x)}$$

$$= \inf_{y \in Y} \left(C_Y(y) - \tilde{\varphi}^{c}(y) \right) - a = 0$$

Recall $-\infty < \langle \tilde{\varphi}, \mu \rangle + \langle \tilde{\psi}, \nu \rangle$

$$\le \langle \tilde{\varphi}^{cc} - a, \mu \rangle + \langle \tilde{\varphi}^{c} + a, \nu \rangle$$

$$= \langle \varphi, \mu \rangle + \langle \psi, \nu \rangle$$

$$\le \langle C_X, \mu \rangle + \langle C_Y, \nu \rangle \le M < \infty$$

$$\Rightarrow 0 \le \langle \underbrace{C_X - \varphi}_{\ge 0}, \mu \rangle + \langle \underbrace{C_Y - \psi}_{\ge 0}, \nu \rangle < +\infty$$

Hence

$C_X - \varphi \in L^1(d\mu)$ + $C_X \in L^1(d\mu) \Rightarrow \varphi \in L^1(d\mu)$

$C_Y - \psi \in L^1(d\nu)$ + $C_Y \in L^1(d\nu) \Rightarrow \psi \in L^1(d\nu)$.

$$\boxed{(\varphi, \psi) \doteq (\tilde{\varphi}^{cc}, \tilde{\varphi}^{c}) \in \Phi_c}$$

Go back to FT of O.T.

Thm (FT of O.T.)

X, Y - Polish. $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$.

$c: X \times Y \to [0, +\infty)$ & c is continuous

cost function

$$\inf_{\pi \in \Pi(\mu,\nu)} \langle c, \pi \rangle = \inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} c(x,y)\, d\pi(x,y)$$

$$= \langle c, \gamma \rangle, \quad \underline{\underline{\gamma}} \in \Pi(\mu, \nu)$$ (Existence is clear now)

Assume that $c(x,y) \le a(x) + b(y)$, with $a \in L^1(d\mu)$, $b \in L^1(d\nu)$,

Assume $\gamma \in \Pi(\mu, \nu)$ = 先验的.

Conclusion: TFAE.

i) $\gamma \in OPT(\mu, \nu)$;

ii) $\mathrm{Supp}(\gamma)$ is c-cyclically monotone;

iii) $\exists$ a c-concave function φ, s.t.

φ: c-concave

$\max\{\varphi, 0\} \in L^1(d\mu)$,

and $\mathrm{supp}(\gamma) \subset \boxed{\partial^c \varphi}$ = c-super-differential of φ.

($\mathrm{Supp}(\gamma) \subset \mathrm{Graph}(\partial\varphi)$)

Pf: iii) $\Rightarrow$ i)

★ If for $\gamma \in \Pi(\mu,\nu)$, $\mathrm{supp}(\gamma) \subset \partial^c \varphi$,

then $\langle c, \tilde{\gamma} \rangle \geq \langle c, \gamma \rangle$,

$\forall\ \tilde{\gamma} \in \Pi(\mu,\nu)$

Checking: $\mathrm{supp}(\gamma) \subset \partial^c \varphi$

$\Rightarrow$ $\forall (x, y) \in \mathrm{supp}(\gamma)$,
$\varphi(x) + \varphi^c(y) = c(x, y)$ $\quad(\tilde{A})$

Also. $\forall\ x \in X,\ y \in Y$, we have
$\varphi(x) + \varphi^c(y) \leq c(x, y)$ $\quad(\tilde{B})$

Now
$$\int_{X\times Y} c(x,y)\, d\gamma(x,y) \overset{(\tilde{A})}{=} \int_{X\times Y} (\varphi(x) + \varphi^c(y))\, d\gamma(x,y)$$

$$= \int_X \varphi(x)\, d\mu(x) + \int_Y \varphi^c(y)\, d\nu(y) \quad \left(\text{marginals of } \gamma\right)$$

$$= \int_{X\times Y} \underline{\underline{(\varphi(x) + \varphi^c(y))}}\, d\tilde{\gamma}(x,y) \quad \left(\text{marginals of } \tilde{\gamma} \in \Pi(\mu,\nu)\right)$$

$$\overset{(\tilde{B})}{\leq} \int_{X\times Y} c(x,y)\, d\tilde{\gamma}(x,y) \quad \text{Okay}$$

i) $\Rightarrow$ ii)

If $\gamma \in OPT(\mu,\nu) \subseteq \Pi(\mu,\nu)$, then $supp(\gamma)$ is c-cyclically monotone.

$\Uparrow$

$\subset X \times Y$ ($\forall N.$ $(x_i, y_i)_{i=1}^{N} \subset supp(\gamma)$,

$$\sum_{i=1}^{N} c(x_i, y_i) \le \sum_{i=1}^{N} c(x_i, y_{\sigma(i)})$$

$\forall \sigma \in S_N.$)

反证：

$\Downarrow$

Assume:

$\exists N \in \mathbb{N}$, $\{(x_i, y_i)\}_{i=1}^{N} \subset supp(\gamma)$,

s.t.

$$\sum_{i=1}^{N} c(x_i, y_i) > \sum_{i=1}^{N} c(x_i, y_{\sigma(i)})$$

Using the fact c is continuous,

$\exists$ nbhd of x_i, say U_i,

- - - - of y_i, say V_i, s.t.

$$\sum_{i=1}^{N} \left[c(u_i, v_{\sigma(i)}) - c(u_i, v_i) \right] < 0$$

$\forall (u_i, v_i) \in U_i \times V_i$

$i = 1, 2, \cdots N.$

$\gamma \rightsquigarrow$ construct a variation of γ

$$\underset{\sim}{\tilde{\gamma}} = \underset{\sim}{\gamma} + \underbrace{\underset{\sim}{\eta}}_{\text{signed measure}} \quad \text{s.t.} \quad \langle c, \tilde{\gamma} \rangle < \langle c, \gamma \rangle$$

s.t.

$$\begin{cases} \bullet \displaystyle\int_{X\times Y} \eta \, dx\,dy = \eta(X\times Y) = 0. \\ \bullet \ \tilde{\gamma} \in \mathcal{P}(X\times Y) \Rightarrow \boxed{\eta^- \leq \gamma} \\ \bullet \displaystyle\int \eta \, dx = 0, \ \int \eta \, dy = 0 \Rightarrow \tilde{\gamma} \in \Pi(\mu, \nu) \end{cases}$$

Let $\underbrace{\Omega}_{\text{whole set}} = (\underbrace{U_1 \times V_1}_{\gamma^{m_1}}) \times \overset{\gamma^{m_2}}{(U_2 \times V_2)} \times \cdots \times \overset{}{(U_N \times V_N)}_{\gamma^{m_N}}$

$\boxed{U_i \times V_i} \subset X \times Y$

$P \in \mathcal{P}(\Omega)$

$\gamma^{m_i} = \frac{1}{m_i} \gamma\big|_{U_i \times V_i}$, $m_i = \gamma(U_i \times V_i)$

$\gamma|_{U_i \times V_i}$: a probability measure on $U_i \times V_i$

s.t.

$$P = \gamma^{m_1} \otimes \gamma^{m_2} \otimes \cdots \otimes \gamma^{m_N}$$

$$\Omega = \{ \overbrace{(x_1, y_1, x_2, y_2, \cdots x_N, y_N)}^{(X_1^N, Y^N)} \mid x_i \in U_i, \ y_i \in V_i, \ i = 1, 2, \ldots N \}$$

(Ω, P) coordinates.

$$\pi^{U_i} : (X^N, Y^N) \longmapsto x_i$$

$$\pi^{V_i} : (X^N, Y^N) \longmapsto y_i$$

<u>**Define:**</u>

$$\eta = \frac{\min_{i=1}^{N} m_i}{N} \sum_{i=1}^{N} \left\{ \left(\pi^{U_i}, \pi^{V_{\sigma(i)}}\right)_\# P - \left(\pi^{U_i}, \pi^{V_i}\right)_\# P \right\}$$

Only Check $\boxed{\langle c, \eta \rangle < 0}$

$$\boxed{\begin{aligned} & T : X_\mu \to Y_{T_\# \mu} \\ & \int_Y \phi \, d T_\# \mu \\ &= \int_X \phi \circ T(x) \, d\mu(x) \end{aligned}}$$

$$\int_{X \times Y} c(x,y) \, d\eta(x,y) = \langle c, \eta \rangle$$

$$= \frac{\min_{i=1}^{N} m_i}{N} \sum_{i=1}^{N} \left\{ \int_{X\times Y} c(x,y) \, d\left(\pi^{U_i}, \pi^{V_{\sigma(i)}}\right)_\# P - \int_{X \times Y} c(x,y) \, d\left(\pi^{U_i}, \pi^{V_i}\right)_\# P \right\}$$

$$= \frac{\min_{i=1}^{N} m_i}{N} \sum_{i=1}^{N} \int_\Omega \left[c(X_i, Y_{\sigma(i)}) - c(X_i, Y_i) \right] d\mathbb{P}$$

$$= \left(\min_{i=1}^{N} m_i \right) \int_\Omega d\mathbb{P} \left\{ \underbrace{\frac{1}{N} \sum_{i=1}^{N} c(x_i, y_{\sigma(i)}) - c(x_i, y_i)}_{< 0} \right\}$$

$$< 0.$$

ii) $\Rightarrow$ iii) (Γ : 循环单调 $\Rightarrow$ $\Gamma \subset \underline{\underline{\partial^c \varphi}}$ φ- c-concave $\propto$ super-differential

c- cyclically monotone

We now prove that if $\Gamma \subset X \times Y$ is c-cyclically monotone, then $\exists$ a c-concave function φ, st $\partial^{c+}\varphi \supset \Gamma$, and

(General version of Rockafeller.) $\max\{\varphi, 0\} \in L^1(d\mu)$.

Fix $(\bar{x}, \bar{y}) \in \Gamma$, for perspective c-concave function φ s.t. $\partial^c \varphi \supset \Gamma$, we need impose that $\forall\ (x_i, y_i) \in \Gamma$, $i = 1, 2, \cdots N$,

$$\underline{\underline{\varphi(x)}} \le c(x, y) - \varphi^{c+}(y_1) \quad (\Leftarrow y_1 \in \partial^{c+}\varphi(x_1))$$

$$= c(x, y) - c(x_1, y_1) + \underline{\underline{\varphi(x_1)}}$$

$$\le (c(x, y) - c(x_1, y_1)) + c(x_1, y_2) - \underline{\varphi^{c+}(y_2)}$$

$$= (c(x, y_1) - c(x_1, y_1)) + (c(x_1, y_2) - \underline{c(x_2, y_2)}) + \underline{\varphi(x_2)}$$

$\le \cdots\cdots$

x-free

(x_1, y_1) $(x_2, y_2) \cdots$ (x_N, y_N)

$\underline{\underline{(\bar{x}, \bar{y})}}$ Fix

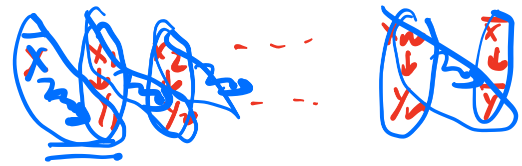
$$\le (c(x, y_1) - c(x_1, y_1)) + (c(x_1, y_2) - c(x_2, y_2)) + \cdots$$

$$+ \cdots + (c(x_N, \bar{y}) - c(\bar{x}, \bar{y})) + \underline{\underline{\varphi(\bar{x})}}$$

$(c(x_{N-1}, y_N) - c(x_N, y_N))$

0

Define φ as the infimum of the above expression as $\{(x_i, y_i)\}_{i=1}^{N}$ vary among all N-pairs in Γ and $N = 1, 2, 3, \ldots$. We are free to add a constant to φ, so



define:

$$\varphi(x) := \inf \Big\{ \big(c(\overset{\bar{x}}{x}, \overset{\bar{y}}{y_1}) - c(\overset{\bar{x}}{x_1}, \overset{\bar{y}}{y_1})\big) + \big(c(x_1, y_2) - c(x_2, y_2)\big) + \cdots + \big(c(x_N, \bar{y}) - c(\bar{x}, \bar{y})\big) \Big\}.$$

$\forall N$, $(x_1, y_1) \cdots (x_N, y_N), (\bar{x}, \bar{y}) \in \operatorname{supp}(\gamma)$

Choosing $N = 1$, and $(x_1, y_1) = (\bar{x}, \bar{y})$, $x = \bar{x}$, $\varphi(\bar{x}) = 0$

we get $\boxed{\varphi(\bar{x}) \le 0.}$

Conversely, from the c-cyclical monotonicity of Γ

$$\left(\left. \begin{aligned} &\big(c(\bar{x}, y_1) - c(x_1, y_1)\big) \\ &+ c(x_1, y_2) - c(x_2, y_2) \\ &\quad\vdots \\ &+ c(x_{N-1}, y_N) - c(x_N, y_N) \\ &+ c(x_N, \bar{y}) - c(\bar{x}, \bar{y}) \end{aligned} \right\} \ge 0 \right)$$

$\bar{x} \quad x_1, \ldots \quad x_N$

$\bar{y} \searrow y_1 \searrow \ldots \quad \searrow y_N \searrow \bar{y}$

one has $\boxed{\varphi(\bar{x}) \ge 0}$

$\varphi(x)$ is a c-concave fun

$\Updownarrow$

$\varphi(x) = \inf_{y \in \Gamma} \big(c(x, y) - \tilde{\psi}(y) \big)$

$\boxed{\tilde{\psi}(y) = \begin{cases} \psi(y) & y \in \operatorname{Supp} \\ -\infty & \text{otherwise.} \end{cases}}$

Thus $\boxed{\varphi(\bar{x}) = 0.}$ (φ is proper)

Clearly, from the construction, φ is c-concave.

Taking $N=1$, $(x_1, y_1) = (\bar{x}, \bar{y})$,

$$\varphi(x) \le c(x, \bar{y}) - c(\bar{x}, \bar{y}) < a(x) + b(\bar{y}) - c(\bar{x}, \bar{y})$$

Since $a \in L^1(\mu)$, $\max\{\varphi, 0\} \in L^1(d\mu)$.

Thus we only need to show that

$\partial^{c+}\varphi \supset \Gamma$.

To do this, take $(\tilde{x}, \tilde{y}) \in \Gamma$, let

$(x_1, y_1) = (\tilde{x}, \tilde{y})$, by the definition of φ,

one has

$$\varphi(x) \le c(x, \tilde{y}) - c(\tilde{x}, \tilde{y}) + \inf_{\substack{\forall N,\ (x_2, y_2) \cdots \\ (\bar{x}, \bar{y}) \in \mathrm{supp}(\gamma)}} \left((c(\tilde{x}, y_2) - c(x_2, y_2)) + \cdots + c(x_N, \bar{y}) - c(\bar{x}, \bar{y}) \right)$$

$$= c(x, \tilde{y}) - c(\tilde{x}, \tilde{y}) + \varphi(\tilde{x}).$$

i.e. $\varphi(\tilde{x}) - c(\tilde{x},\tilde{y}) \geq \varphi(x) - c(x,\tilde{y})$, $\forall x \in X$.

This implies that $\tilde{y} \in \partial^{c+}\varphi(\tilde{x})$.

($\Leftarrow$ $\begin{cases} \varphi(x) + \varphi^c(\tilde{y}) \leq c(x,\tilde{y}) & \forall x \\ \varphi(\tilde{x}) + \varphi^c(\tilde{y}) = c(\tilde{x},\tilde{y}) & \text{if } \tilde{y} \in \partial^c\varphi(\tilde{x}) \end{cases}$

so it is equivalent to

$$c(x,\tilde{y}) - \varphi(x) \geq c(\tilde{x},\tilde{y}) - \varphi(\tilde{x}))$$ okay

<u>Remark:</u>

Recall that the definition of φ.

$$\varphi(x) := \inf_{\substack{\forall N, \\ \left.\begin{matrix}(x_1,y_1)\\ \vdots \\ (x_N,y_N)\end{matrix}\right\} \in \Gamma = \mathrm{supp}(\gamma)}} \Big\{ \big(c(x,y_1) - c(x_1,y_1)\big) + \big(c(x_2,y_2) - c(x_2,y_2)\big) + \cdots + \cdots + \big(c(x_N,\bar{y}) - c(\bar{x},\bar{y})\big) \Big\}$$

only $c(x,y_1)$ depends on x

$$= \inf_{\substack{\alpha = (x_1,y_1,\ldots \\ x_N,y_N \in \text{Index} \\ \ldots)}} \left\{ c(x,y_1) + f(\alpha) \right\}$$

Of course the function $h(x) = c(x,y_1)$ is c-concave

since by definition, $h = \psi^c$, with

$$\psi^c(x) = \inf_{y \in Y} \left(c(x,y) - \psi(y) \right)$$

define: $\psi(y) = \begin{cases} 0 & \text{if } y = y_1 \\ -\infty & \text{if } y \neq y_1 \end{cases}$

Then now

φ is just an infimum of a family of c-concave functions on X, which <u>is</u> also c-concave.

$$\varphi(x) = \inf_{\alpha \in \Gamma} \left(\varphi_\alpha(x) \right) = \inf_{\alpha \in \Gamma} \inf_{y \in Y} \left(c(x,y) + \psi_\alpha(y) \right)$$

$$\overset{\vee}{=} \inf_{y \in Y} \inf_{\alpha \in \Gamma} \left(c(x,y) + \psi_\alpha(y) \right)$$

$\square$

Lecture 6.

We have proved the Fundamental Theory of O.T.

Some remarks:

- $\gamma \in \Pi(\mu,\nu)$ optimal, depends only on $supp(\gamma)$, not explicitly on how the mass is distributed.

ex: if $\gamma \in OPT(\mu,\nu)$, $\tilde{\gamma} \in \Pi(\mu,\nu)$ s.t.

$$supp(\tilde{\gamma}) \subseteq supp(\gamma),$$

then $$\tilde{\gamma} \in OPT(\mu,\nu).$$

- If $T: X \to Y$ is a map

s.t. $T(x) \in \partial^c \varphi(x)$, (super-differential of φ at x)

for some c- concave φ,

Then under assumptions $\begin{cases} c(x,y) \le a(x) + b(y), \\ \qquad a \in L^1(d\mu),\ b \in L^1(d\nu), \\ \nu = T_\# \mu, \end{cases}$

the map T is optimal between μ and $\nu = T_\# \mu$.

(or $(Id, T)_\# \mu = \gamma \in OPT(\mu,\nu)$).

Hence it makes perfect sense to say that

T is an optimal map, without explicit mention

of the reference measures.

- <u>A stronger statement holds:</u>

If $\mathrm{supp}(\gamma) \subset \partial^c \varphi$ for some optimal γ,

then $\mathrm{supp}(\gamma') \subset \partial^c \varphi$ for all $\gamma' \in OPT(\mu, \nu)$.

Since $\varphi \vee 0 \in L^1(d\mu)$,

$$\varphi^c(y) = \inf_{x \in X} \left(c(x, y) - \varphi(x) \right)$$

$$\varphi^c(y) \leq c(\bar{x}, y) - \varphi(\bar{x})$$

$$\leq a(\bar{x}) + b(y) - \varphi(\bar{x})$$

i.e. $\max\{\varphi^c, 0\} \in L^1(d\nu)$.

Thus $\int_X \varphi \, d\mu + \int_Y \varphi^c \, d\nu$

$$\overset{\gamma' \in \Pi(\mu,\nu)}{=\!=} \int_{X \times Y} (\varphi(x) + \varphi^c(y)) \, d\gamma'(x, y)$$

should be "=" $\uparrow$ $\leq$ $\int_{X \times Y} c(x, y) \, d\gamma'(x, y)$

$\Rightarrow (x, y) \in \partial^c \varphi$, for γ'-a.e. (x, y).

$$= \int_{X\times Y} c(x,y)\, d\gamma'(x,y)$$

$\uparrow$ Since both γ and γ' are optimal.

Going back to Kantorovich duality

Primal:

$$\inf_{\gamma\in\Pi(\mu,\nu)} \int_{X\times Y} c(x,y)\, d\gamma(x,y) \quad = \langle c, \gamma^* \rangle, \quad \gamma^* \in OPT(\mu,\nu)$$

linear functional

affine constraint

$\|$ Kantorovich duality $\qquad\qquad \|$

Dual

$$\sup_{\substack{\varphi\oplus\psi\le c,\\ \varphi\in L^1(d\mu),\\ \psi\in L^1(d\nu)}} \left\{ \int_X \varphi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y) \right\} = \langle \varphi^*, \mu\rangle + \langle \varphi^{*c}, \nu\rangle$$

$\uparrow$ c-concave.

(and $\operatorname{supp}(\gamma^*) \subset \partial^c \varphi^*$

Theorem (Kantorovich duality)

Assumptions as in F.T. of O.T.

- $\mu\in P(X)$, $\nu\in P(Y)$
  $c: X\times Y \to [0,+\infty)$ and continuous;
- $c \le a\oplus b$ pointwise for $a\in L^1(d\mu)$, $b\in L^1(d\nu)$

Then

$$\inf_{\pi \in \Pi(\mu,\nu)} \langle c, \pi \rangle = \sup_{(\varphi,\psi)\in\Phi_c} \{ \langle \varphi, \mu \rangle + \langle \psi, \nu \rangle \}$$

and

- The primal problem admits a minimizer $\gamma \in OPT(\mu,\nu)$;
- The dual problem admits a maximizing couple (φ, ψ), that can be chosen in the form of (φ, φ^c) for φ - c-concave.

Pf: Trivially, since for any pair $(\varphi,\psi) \in \Phi_c$,

$$\varphi(x) + \psi(y) \leq c(x,y) \quad \mu\otimes\nu\text{ - a.e.}$$

$$\sup_{\Phi_c} \{ \langle \varphi, \mu \rangle + \langle \psi, \nu \rangle \} \leq \inf_{\pi\in\Pi(\mu,\nu)} \langle c, \pi \rangle .$$

Now it is much easier to prove the converse direction.

Choose any $\gamma \in OPT(\mu,\nu)$, by F.T. of O.T,

$\exists$ a c-concave function φ, s.t.

$$\mathrm{supp}(\gamma) \subset \partial^c \varphi,$$

and $\begin{cases} \max\{\varphi, 0\} \in L^1(d\mu) \\ \max\{\varphi^c, 0\} \in L^1(d\nu). \end{cases}$

Then $\boxed{c \geq 0}$ $\qquad\qquad \mathrm{supp}(\gamma) \subset \partial^c\varphi$

$$0 \leq \int c(x,y)\, d\gamma(x,y) = \int (\varphi(x) + \varphi^c(y))\, d\gamma(x,y)$$

$$= \int_X \varphi(x)\, d\mu(x) + \int_Y \varphi^c(y)\, d\nu(y) < +\infty$$

($\langle \varphi_+ - \varphi_-, \mu \rangle + \langle \varphi^c_+ - \varphi^c_-, \nu \rangle \in \mathbb{R}$.

while $\langle \varphi_+, \mu \rangle + \langle \varphi^c_+, \nu \rangle \in \mathbb{R}$

$\hookrightarrow \varphi \in L^1(d\mu)$ & $\varphi^c \in L^1(d\nu)$)

i.e. $(\varphi, \varphi^c) \in \Phi_c$ is an admissable pair

which maximizing the dual problem.

$\boxtimes$

Remark: Under all assumptions as above,

for any c-concave couple of functions (φ, φ^c)

maximizing the dual problem, and any optimal plan γ, we have

$$\mathrm{supp}(\gamma) \subset \partial^c \varphi.$$

Pf: By F.T. of O.T., $\exists$ some c-concave function φ, s.t. $\varphi \in L^1(d\mu)$, $\psi = \varphi^c \in L^1(d\nu)$ s.t.

$$\underline{\mathrm{supp}(\gamma) \subset \partial^c \varphi, \quad \forall \gamma \in OPT(\mu, \nu)}$$

(See the remark following F.T. of O.T.)

For other maximizing couple $(\tilde{\varphi}, \tilde{\psi})$ for the dual problem,

$$(\tilde{\varphi}, \tilde{\psi}) \xrightarrow{\text{improve}} (\tilde{\varphi}, \tilde{\varphi}^c) \text{ is also a maximizing couple.}$$

$$(\tilde{\varphi}^c \in L^1(d\nu)).$$

Now $\forall \gamma \in OPT(\mu, \nu)$,

$$\int_X \tilde{\varphi}\, d\mu + \int_Y \tilde{\varphi}^c\, d\nu = \int_X \varphi\, d\mu + \int_Y \psi\, d\nu$$

$$= \int_{X \times Y} (\varphi(x) + \varphi^c(y))\, d\gamma(x,y)$$

$$= \int_{X \times Y} c(x,y)\, d\gamma(x,y)$$

$$= \geq \int \tilde{\varphi}\, d\mu + \int \tilde{\varphi}^{c+}\, d\nu$$

$$\Rightarrow \quad \mathrm{supp}(\gamma) \subset \partial^c \tilde{\varphi}.$$

$\boxtimes$

Def: A c-concave function φ, s.t. (φ, φ^c) is a maximizing pair for the dual problem is called a c-concave <u>Kantorovich potential</u>, for the couple μ and ν.

Brenier's Theorem:

Taking $X = Y = \mathbb{R}^d$, $\mu \in \mathcal{P}_2(\mathbb{R}^d) \leftarrow$ source measure

$\nu \in \mathcal{P}_2(\mathbb{R}^d) \sim$ target measure

If μ is regular (a particular case is that $\mu \ll Leb$), then there exists only one transport plan in $\Pi(\mu, \nu)$ that is induced by a map T, and the optimal map $T = \nabla \bar{\varphi}$, $\bar{\varphi}$ - a convex function.

What does "regular" mean here?

[Def ($c-c$ hypersurface) "$c-c$" means: "convex - convex"

A set $E \subseteq \mathbb{R}^d$ is called $c-c$ hypersurface, if in a suitable system of coordinates, it is the graph of the difference of two real-valued convex functions, i.e. $\exists\, f, g : \mathbb{R}^{d-1} \to \mathbb{R}$, s.t.

$$E = \{ (y, t) \in \mathbb{R}^d \mid y \in \mathbb{R}^{d-1},\ t = f(y) - g(y) \}.$$

Thm (Structures of sets of non-differentiability of convex functions)

Let $A \subset \mathbb{R}^d$.

- $A \subset \{x \in \mathbb{R}^d \mid \bar{\varphi} \text{ is not differentiable at } x\}$
  $\bar{\varphi}: \mathbb{R}^d \to \mathbb{R}$ convex

$\Updownarrow$

- $A \subset \bigcup_{i=1}^{\infty} H_i$, H_i is c-c hypersurface.

We call a probability measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ is regular, if $\mu(E) = 0$ for any c-c hypersurface $E \subset \mathbb{R}^d$.

By the structure theorem above (we only cite), for any $\bar{\varphi}: \mathbb{R}^d \to \mathbb{R}$ convex, any regular $\mu \in \mathcal{P}(\mathbb{R}^d)$

$$\mu(\mathcal{N}_{\bar{\varphi}}) = 0, \quad \text{where } \mathcal{N}_{\bar{\varphi}} = \{x \in \mathbb{R}^d \mid \bar{\varphi} \text{ is not differentiable at } x\}.$$

Examples of regular measures:

- Lebesgue Measures;
- measures A. C. w.r.t. Lebesgue;

- measures give 0 mass to Lipschitz hypersurfaces. (convex functions are locally Lipschitz $\xrightarrow{\text{Radmecher}}$ a.e. differentiable)

Now we proceed to prove Brenier's theorem:

Pf: Take $a(x) = b(x) = |x|^2$ as in the proof of F.T. of O.T. $a \in L^1(d\mu)$, $b \in L^1(d\nu)$ since $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$.

By F.T. of O.T. (in particular the remarks), for any c-concave Kantorovich potential φ and any optimal plan $\gamma \in OPT(\mu, \nu)$, one has:

$$\operatorname{supp}(\gamma) \subset \partial^c \varphi.$$

Easy claim: for $c = \frac{1}{2}|x-y|^2$, φ is c-concave $\iff \frac{1}{2}|x|^2 - \varphi \triangleq \bar{\varphi}$ is convex. and $\partial^c \varphi = \partial \bar{\varphi}$.

Since $\bar{\varphi} = \frac{1}{2}|x|^2 - \varphi$ is convex

and μ is regular,

$\mu(E) = 0$, for $E = \{x \in \mathbb{R}^d \mid \bar{\varphi}$ is not differentiable at $x\}$.

(also $\mathcal{M}(E) = 0$ in Lebesgue.)

Hence

$\nabla\bar{\varphi} : \mathbb{R}^d \to \mathbb{R}^d$ is well-defined μ-a.e. $x \in \mathbb{R}^d$

(also. $\mathcal{L}^d$-a.e).

and every optimal plan $\gamma \in OPT(\mu, \nu)$ has the property: $\text{supp}(\gamma) \subset \text{Graph}(\nabla\bar{\varphi})$.

Hence, the optimal plan is unique, and it is induced by the map $T = \nabla\bar{\varphi}$.

Remark: (Perturbations of the identity via smooth gradients are optimal)

Given $\psi \in C_c^\infty(\mathbb{R}^d)$, choose $\bar{\varepsilon}$ small enough, $(|\bar{\varepsilon}| \ll 1)$

s.t. $-Id \le \bar{\varepsilon}\nabla^2\psi \le Id$ (Majorization in eigenvalues)

Hence for any ε, $|\varepsilon| \le \bar{\varepsilon}$,

the map: $x \mapsto \frac{|x|^2}{2} + \varepsilon \psi(x)$ is convex

(indeed, $\nabla^2\left(\frac{|x|^2}{2} + \varepsilon\psi(x)\right) = \underset{\substack{\uparrow \\ \text{matrix}}}{Id} + \varepsilon \nabla^2 \psi \geq 0$)

Hence its gradient is

$$T(x) = \nabla\left(\frac{1}{2}|x|^2 + \varepsilon \psi(x)\right)$$

$$= x + \varepsilon \nabla \psi(x)$$

is an optimal map, given $T_\# \mu = \nu$.

Applications:

Polar factorization of vector fields on $\mathbb{R}^d$.

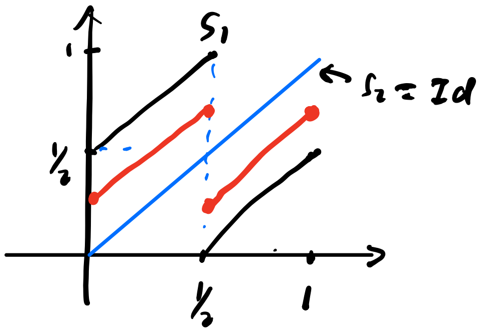
Let $\Omega \subset \mathbb{R}^d$ be a bounded domain,

μ_Ω = normalized Lebesgue measure on Ω

define

$$S(\Omega) := \{ \text{Borel map } s : \Omega \to \Omega \mid s_\# \mu_\Omega = \mu_\Omega \}.$$

s is a (Lebesgue) measure preserving map.



$\boxed{S(\Omega) \text{ is non-convex}}$

$\frac{S_1 + S_2}{2}$ is not measure preserving

Thm (Polar factorization)

Let $S \in L^2(\mu_\Omega; \mathbb{R}^d)$ be s.t. $\nu := S\#\mu_\Omega$ is regular.

Then $\exists ! \; s \in S(\Omega)$ and $\nabla\varphi$ with φ-convex, s.t.

$S = \nabla\varphi \circ s$. Also

$$s = \underset{\tilde{s} \in S(\Omega)}{\arg\min} \int |S - \tilde{s}|^2 \, d\mu_\Omega$$

$\boxed{\text{This is a very interesting theorem!}}$

Pf: By assumption, both μ_Ω and ν are regular,

and $\mu_\Omega, \nu \in \mathcal{P}_2(\mathbb{R}^d)$.

$\boxed{\text{Claim:}}$

$$\inf_{\tilde{s} \in S(\Omega)} \int |S - \tilde{s}|^2 \, d\mu_\Omega = \min_{\gamma \in \Pi(\mu_\Omega, \nu)} \int |x - y|^2 \, d\gamma(x, y).$$

measure preserving

Indeed, for each $\tilde{S} \in \mathcal{S}(\Omega)$,

the plan: $\gamma_{\tilde{S}} = (\tilde{S}, S)_\# \mu_\Omega \in \Pi(\mu_\Omega, \nu)$.

This gives that

$$\inf_{\tilde{S} \in \mathcal{S}(\Omega)} \int |S - \tilde{S}|^2 \, d\mu_\Omega \geq \min_{\gamma \in \Pi(\mu_\Omega, \nu)} \int |x-y|^2 \, d\gamma(x, y).$$

Now let $\bar{\gamma}$ be the unique optimal plan and apply Brenier's theorem twice to get that

$$\bar{\gamma} = (Id, \nabla\varphi)_\# \mu_\Omega = (\nabla\tilde{\varphi}, Id)_\# \nu,$$

for some convex functions φ, $\tilde{\varphi}$, which satisfies

$$\nabla\varphi \circ \nabla\tilde{\varphi} = Id, \quad \nu\text{-a.e.}$$

$$\nabla\tilde{\varphi} \circ \nabla\varphi = Id \quad \mu_\Omega\text{-a.e.}$$

Define $s = \nabla\tilde{\varphi} \circ S$, then $s_\# \mu_\Omega = (\nabla\tilde{\varphi})_\# \underbrace{S_\# \mu_\Omega}_{= \nu} = \mu_\Omega$.

i.e. $s \in \mathcal{S}(\Omega)$

Also $\boxed{S = \nabla\varphi \circ s}$, that proves the existence of the polar factorization. Indeed,

$$\int |x-y|^2 \, d\gamma_S(x,y) = \int |S - S|^2 d\mu_\Omega$$

$$\gamma_S = (S, S)_\# \mu_\Omega \quad = \int |\nabla\tilde{\varphi} \circ S - S|^2 d\mu_\Omega$$

$$= \int |\nabla\tilde{\varphi} - Id|^2 d\underbrace{S_\# \mu_\Omega}_{= \nu}$$

$$= \min_{\gamma \in \Pi(\mu,\nu)} \int |x-y|^2 d\gamma(x,y).$$

$$\int_\Omega |S - s|^2 d\mu \le \int_\Omega |S - \tilde{s}|^2 d\mu$$

This proves the claim.

Exe:

Question: Is the minimizer of the problem unique?

$$\left(\inf_{\tilde{S} \in S(\Omega)} \int_\Omega |S - \tilde{S}|^2 d\mu_\Omega \right)$$ and why?

To conclude, we show uniqueness of the polar factorization. Assume $S = (\nabla\bar{\varphi}) \circ \bar{S}$ is another factorization and

$$(\nabla\varphi)_\# \mu_\Omega = (\nabla\bar{\varphi} \circ \bar{S})_\# \mu_\Omega = \nu.$$

$\nabla\bar{\varphi}$ is a transport map from μ_Ω to ν, and it is in the gradient form of a convex function, $\nabla\bar{\varphi}$ is hence an optimal map.

Hence $\nabla \bar{\varphi} = \nabla \varphi$ $\blacksquare$

Rk: The classical Helmoltz decomposition of vector fields can be seen as a linearized version of the Polar factorization result.

Formal: Assume that Ω and all the objects are considered are smooth.

Let $u : \Omega \to \mathbb{R}^d$ vector field.

We apply the Polar factorization to the map

$$S_\varepsilon := Id + \varepsilon u \quad \text{with } |\varepsilon| \ll 1.$$

Then $S_\varepsilon = \nabla \varphi_\varepsilon \circ s_\varepsilon$

measure-preserving

perturbation of identity

Say $\nabla \varphi_\varepsilon = Id + \varepsilon v + o(\varepsilon)$,

$s_\varepsilon = Id + \varepsilon w + o(\varepsilon)$.

What information is carried on v, w

from the Polar factorization?

- At the level of r,

$$\nabla \varphi_\varepsilon(x) = x + \varepsilon v(x) + o(\varepsilon).$$

$$\nabla \underset{\substack{\uparrow \\ \text{curl}}}{\times} (\nabla \varphi_\varepsilon) = 0 \quad \Longrightarrow \quad \nabla \times v = 0$$

$$\Downarrow$$

$$v = \nabla p$$

- S_ε is measure preserving

$$\Longrightarrow \quad \nabla \cdot (w \chi_\Omega) = 0. \quad \text{in } \mathcal{D}'.$$

$$0 = \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \int f \, d(S_\varepsilon)_{\#}\mu_\Omega = \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \int f \circ S_\varepsilon \, d\mu_\Omega$$

$$= \int \nabla f \cdot w \, d\mu_\Omega..$$

Then from the identity

$$(\nabla \varphi_\varepsilon) \circ S_\varepsilon = Id + \varepsilon(\underbrace{\nabla p + w}_{u}) + o(\varepsilon)$$

we conclude $\quad \nabla p + w. = u.$

Lecture 7 | Remarks and Examples

§1.2 Distance cost function

Choose the cost function $c(x,y) = d(x,y)$ on $X = Y$ as the metric on X,

then more structure in the Kantorovich duality principle.

Thm (Kantorovich-Rubinstein theorem)

cost function $c = d$ ← a metric

Assume $X = Y =$ Polish

d : l.s.c. on X. (d is only l.s.c.)

Define $$T_d(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} d(x,y)\, d\pi(x,y).$$

$\boxed{(X, d)}$

(φ, φ^c) ← c-concave

$\boxed{c = d}$

(φ, φ^c)

Let $\mathrm{Lip}(X) = \{ f : X \to \mathbb{R} \text{ Lipschitz} \}$, and

$$\|\varphi\|_{Lip} = \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d(x,y)}.$$

W_1 ← Kantorovich-Rubinstein distance

Then

$$T_d(\mu,\nu) = \sup \left\{ \int_X \varphi \, d(\mu - \nu) : \varphi \in L^1(d|\mu-\nu|), \ \|\varphi\|_{Lip} \le 1 \right\}.$$

the optimal cost with cost function $= d$

Pf: Let $d_n = \dfrac{d}{1 + d/n}$, $0 \le d_n \le n$

$d_n \le d \qquad d_n(x,y) \uparrow d(x,y)$ as $n \uparrow +\infty$
$(\forall (x,y))$

If φ is 1-Lipschitz for d_n, i.e.

$$|\varphi(x) - \varphi(y)| \le d_n(x,y), \quad \boxed{\le d(x,y)}$$

then φ is 1-Lipschitz for d!

We only need to prove theorem for d_n.

Hence, we can just assume $d \in L^\infty$.

In this case, all Lipschitz functions are in L^∞,

and hence in $L^1(d\mu)$, $L^1(d\nu)$.

Hence, it suffices to check

$$\sup_{(\varphi,\psi)\in\Phi_d} J(\varphi,\psi) = \sup\left\{\int_X \varphi(d\mu - d\nu) \,;\, \|\varphi\|_{Lip} \le 1\right\}.$$

By Rüschendorf's "improving" trick,

$$\underset{(\varphi,\psi)}{\sup} \underbrace{J(\varphi,\psi)}_{\substack{\| \\ \langle\varphi,\mu\rangle + \langle\psi,\nu\rangle}} = \sup_{\varphi\in L^1(d\mu)} J(\varphi^{dd}, \varphi^d),$$

where $(\varphi,\psi) \hookrightarrow (\varphi, \varphi^d) \to (\varphi^{dd}, \varphi^d)$

$$\varphi^d(y) = \inf_{x\in X}\left(d(x,y) - \varphi(x)\right) \quad \leftarrow \text{1-Lipschitz}$$

$$|d(x,y) - d(x,w)| \le d(y,w)$$

$$\varphi^{dd}(x) \equiv \inf_{y\in X}\left(d(x,y) - \varphi^d(y)\right)$$

infimum of 1-Lip functions,

$$\varphi^d(y) \geq d(x_0,y) - \varphi(x_0) \geq -C$$

1-Lipschitz

$$\underbrace{-\varphi^d(x)}_{\text{1 Lipschitz}} \le \inf_y\left[d(x,y) - \varphi^d(y)\right] \underset{y=x}{\le} -\varphi^d(x)$$

$$\left(\varphi^d(y) - \varphi^d(x) \le d(x,y)\right)$$

That means $\varphi^{dd} = -\varphi^d$, and.

$$\sup_{\Phi_c} J(\varphi,\psi) \le \sup_{\varphi\in L^1(d\mu)} J(\varphi^{dd},\varphi^d)$$

$$= \sup_{\varphi\in L^1(d\mu)} J(-\varphi^d,\varphi^d)$$

$$\le \sup_{\|\varphi\|_{Lip}\le 1} J(\varphi,-\varphi) \le \sup_{\Phi_c} J(\varphi,\psi).$$

So there is equality everywhere, and the result follows. ∎

<u>Exercise 47</u>. (Total variation formula)

Apply the Theorem above to the cost function

$c(x,y) = 1_{x \neq y}$ (trivial distance function)

Then for $\mu, \nu \in P(X)$,

$$\inf_{\pi \in \Pi(\mu,\nu)} \pi\big[\{x \neq y\}\big] = \sup_{0 \le f \le 1} \int_X f \, d(\mu - \nu).$$

$$\left(\pi\big[\{x \neq y\}\big] = \int 1_{x \neq y} \, d\pi(x,y) \right)$$

Also note the decomposition

$$\mu - \nu = (\mu - \nu)_+ - (\mu - \nu)_-,$$

$(\mu - \nu)_+ \perp (\mu - \nu)_-$ singular to each other

$$\sup_{0 \le f \le 1} \int_X f \, (d\mu - d\nu) = (\mu - \nu)_+[X]$$
$$= (\mu - \nu)_-[X] = \frac{1}{2} \|\mu - \nu\|_{TV}.$$

Kantorovich - Rubinstein theorem (Skip the transshipment probl.)

implies that the total cost only depends on the difference $\mu - \nu$.

That is, when the cost function is a metric, $c(x,y) = d(x,y)$, Kantorovich's optimal transportation problem $\Leftrightarrow$ Kantorovich - Rubinstein transshipment problem:

$$\inf \left\{ I[\pi] : \underbrace{\pi[A \times X] - \pi[X \times A] = (\mu - \nu)(A)}_{\text{more general.}} \right\}.$$

$$\Uparrow$$

(Kantorovich's problem

$$\inf_{\pi \in \Pi(\mu,\nu)} I[\pi] \rightsquigarrow \pi[A \times X] = \mu[A], \quad \pi[X \times B] = \nu[B].$$

Revisit Chapter II: Geometry of O.T.

$c(x,y) = \frac{1}{2}|x-y|^2$ in $\mathbb{R}^n$

FT of O.T. $X = Y = \mathbb{R}^n$, $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^n)$

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

$\pi \in OPT(\mu,\nu)$

$\Updownarrow$

$supp(\pi) \subset \partial\varphi$, φ is a convex function.

Historically, Knott-Smith optimality criterion.

rediscovered by Brenier.

Consider the special case

REVIEW:

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

$$\inf I[\pi] = \frac{1}{2}\int_{\mathbb{R}^n \times \mathbb{R}^n} |x-y|^2 \, d\pi(x,y).$$

$$= M_2 - \sup \int x \cdot y \, d\pi(x,y)$$

inf

$\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d) \Rightarrow I[\pi] < \infty$ for $\pi \in \Pi(\mu,\nu)$

Knott-Smith

$$\frac{1}{2}|x-y|^2 \leq |x|^2 + |y|^2$$

Prop. 2.1: Existence of an optimal transference plan.

i.e. $OPT(\mu,\nu) \neq \phi$

The dual problem:

$(\varphi, \psi) \in \Phi_c$

i.e. $\varphi(x) + \psi(y) \le \frac{1}{2}|x-y|^2$

$\Updownarrow$ for $d\mu$-a.e. x
$\{ d\nu$-a.e. y.

$$x \cdot y \le \underbrace{\left[\frac{|x|^2}{2} - \varphi(x)\right]}_{= \tilde{\varphi}(x)} + \underbrace{\left[\frac{|y|^2}{2} - \psi(y)\right]}_{\tilde{\psi}(y)}.$$

$$\tilde{\varphi}(x) + \tilde{\psi}(y) \ge x \cdot y$$

$$\Phi_c \rightsquigarrow \tilde{\Phi}$$

$$\sup_{\pi \in \Pi(\mu,\nu)} \left\{ \int x \cdot y \, d\pi(x,y) \right\} \overset{\text{duality}}{=\!=} \inf_{\varphi \oplus \psi \ge x \cdot y} \left\{ J(\varphi, \psi) \right\}.$$

$$\int \varphi \, d\mu + \int \psi \, d\nu.$$

$\varphi \oplus \psi \ge x \cdot y$

$\varphi(x) + \psi(y) \ge x \cdot y$

$$\psi(y) \ge \underbrace{\sup_x \left[x \cdot y - \varphi(x) \right]}_{= \varphi^*(y)}$$

(φ, ψ)
$\Downarrow$
(φ, φ^*) (Admissible as well)
$\Downarrow$
$(\varphi^{**}, \varphi^*)$

$$\varphi^*(y) \triangleq \sup_x \left[x\cdot y - \varphi(x) \right].$$

$$\boxed{\varphi(x) + \psi(y) \geq x\cdot y.}$$

redefine $\varphi(x) = +\infty$ on N_x
$\psi(y) = +\infty$ on N_y
with $\mu[N_x] = \nu[N_y] = 0$

s.t. ① holds everywhere.

and $\sup_x$ understood as a true supremum.

$$\varphi^*(y) = \sup_x \{ \underbrace{x\cdot y - \varphi(x)}_{\text{affine}} \}.$$

affine $\Rightarrow \begin{cases} \text{convex} \\ \text{continuous} \end{cases} \xRightarrow{\text{taking superemum}} \varphi^*$ convex l.s.c.

Convex analysis:

Regularity of a convex function in $\mathbb{R}^n$

"small set" : of Hausdorff dimension $\leq n-1$.

(these sets are of zero Leb. measure.)

or just think it as "Leb. negligible" set.

Def:
$\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. not identically $+\infty$. (proper)
convex: $\forall x, y \in \mathbb{R}^n$, $t \in [0,1]$,

$$\varphi(tx+(1-t)y) \leq t\varphi(x) + (1-t)\varphi(y).$$

(strictly convex)

$$\mathrm{Dom}(\varphi) = \{x \in \mathbb{R}^n \mid \varphi(x) < +\infty\}$$

$\uparrow$

convex $\quad \partial(\mathrm{Dom}(\varphi))$ is small.

- Differentiability:

(NOT recorded)

Homework / Midterms:

$\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$

[A proper convex function φ is automatically

$\varphi \not\equiv +\infty$

continuous and locally Lipschitz on $\mathrm{Int}(\mathrm{Dom}(\varphi))$.]

finite

$\Downarrow$ Rademacher's theorem: ($\partial\varphi$ $\Rightarrow L^\infty \Rightarrow$ locally Lipschitz)

$\nabla\varphi$ is a.e. well-defined, locally bounded.

$(\nabla\varphi \in W^{1,\infty})$ $\quad$ (Evans – – + Fine Properties of functions.

- For x, s.t. φ is differentiable at x.

$$\forall z \in \mathbb{R}^n, \quad \varphi(z) \geq \varphi(x) + \nabla\varphi(x)\cdot(z-x),$$

The graph of φ lies above its tangent hyperplane at point x.

As a consequence, whenever φ is diff at both x and z, then

$$\langle \nabla\varphi(x) - \nabla\varphi(z), x - z \rangle \geq 0$$

$$(\Leftarrow \quad \varphi(z) \geq \varphi(x) + \nabla\varphi(x)\cdot(z-x).$$

$$\varphi(x) \geq \varphi(z) + \nabla\varphi(z)\cdot(x-z)$$

$$0 \geq (\nabla\varphi(x) - \nabla\varphi(z))\cdot(z-x)$$

$$\text{or} \quad \langle \nabla\varphi(x) - \nabla\varphi(z), x-z \rangle \geq 0$$

(in 1D case, $x \geq z$, $\varphi'(x) \geq \varphi'(z)$.
okay)

- Sub-differentiability (次可微)

  sub-differential.

  $y \in \partial\varphi(x)$

  $\Leftrightarrow [\forall z \in \mathbb{R}^n, \ \varphi(z) \geq \varphi(x) + \langle y, z-x \rangle]$

  $\partial\varphi \longleftrightarrow \text{Graph}(\partial\varphi)$.

$\partial\varphi \subset \mathbb{R}^n \times \mathbb{R}^n$.

Prop.

- $\forall x \in \mathrm{Int}(\mathrm{Dom}(\varphi))$, $\partial\varphi(x) \neq \phi$.
- " φ is differentiable at $x \iff \partial\varphi(x) = \{\nabla\varphi(x)\}$.

- If φ is l.s.c., then the sub-differential map $\partial\varphi$ is always "continuous" on the whole of $\mathbb{R}^n$:

$$\left.\begin{array}{l} x_k \to x \\ \partial\varphi(x_k) \ni y_k \to y \end{array}\right\} \Rightarrow y \in \partial\varphi(x).$$

$$\varphi(z) \geq \liminf_{k\to\infty}\left(\varphi(x_k) + \langle y_k, z - x_k\rangle\right)$$

$$\geq \varphi(x) + \langle y, z-x\rangle.$$

This is NOT mention before

- The sub-differential mapping generates the normal cone to the sublevel sets of φ:

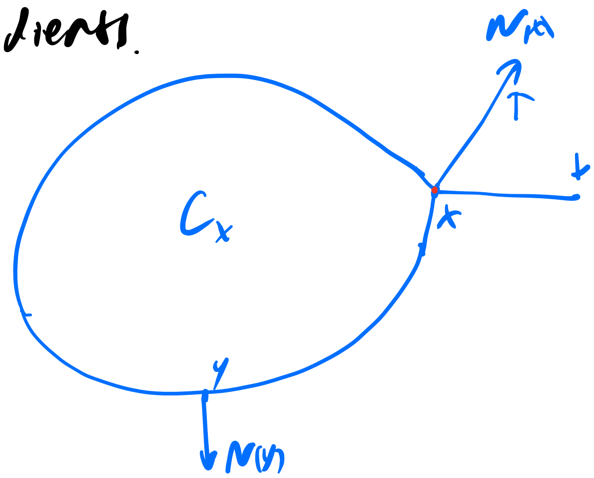
$$C_x = \{z \in \mathbb{R}^n \mid \varphi(z) \leq \varphi(x)\}.$$

$\forall\, y \in \partial\varphi(x)$, $y \in$ Normal Cone N_x

$$N_x = \{ n_x \in \mathbb{R}^n \mid \forall z \in C_x,\ \langle n_x, z-x \rangle \le 0 \}$$

$$\underbrace{\varphi(x) \ge \varphi(z)}_{z \in C_x} \ge \underbrace{\varphi(x) + \langle y,}_{y \in \partial\varphi(x)} z-x \rangle \Rightarrow \langle y, z-x \rangle \le 0 \quad \forall z \in C_x.$$

This generalizes the usual properties of the gradients.

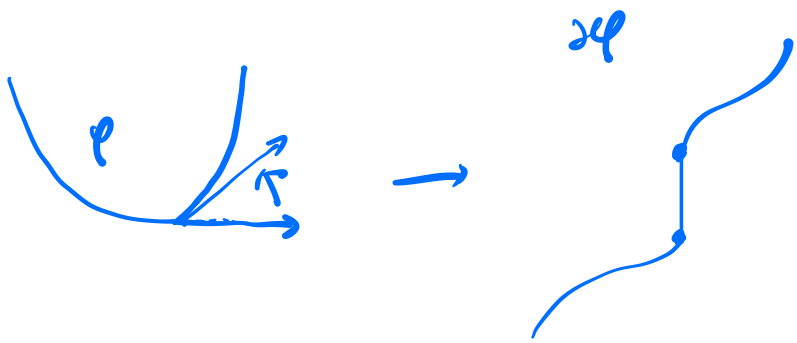


- On $\mathrm{Int}(\mathrm{Dom}(\varphi))$,

$$\partial\varphi(x) = \overline{\mathrm{Conv}}\left(\lim_{x_k \to x} \nabla\varphi(x_k)\right)$$

$\downarrow$

凸包



4) <u>Monotonicity</u>

set-valued.

★ $\partial\varphi : x \mapsto \partial\varphi(x)$ convex

is a monotone mapping

$\partial\varphi$: sub-differential
If φ is differentiable at x,
then $\partial\varphi(x) = \{\nabla\varphi(x)\}$

$\forall y_1 \in \partial\varphi(x_1), \ \forall y_2 \in \partial\varphi(x_2), \ \langle y_2 - y_1, x_2 - x_1 \rangle \geq 0$

$$\Leftarrow \left(\begin{aligned} &\varphi(x_2) \geq \varphi(x_1) + \langle y_1, x_2 - x_1 \rangle. \\ &\varphi(x_1) \geq \varphi(x_2) + \langle y_2, x_1 - x_2 \rangle \end{aligned} \right)$$

(Used to prove the points of non-diff form a small set.)

$\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}, \ \varphi \not\equiv +\infty.$

<u>convex conjugate:</u> $\varphi^*(y) = \sup_{x \in \mathbb{R}^n} (x \cdot y - \varphi(x))$

φ^*: proper, convex, l.s.c.

$\forall x, y, \quad \boxed{x \cdot y \leq \varphi(x) + \varphi^*(y)}$

always true

Characterization of sub-differential:

φ proper, l.s.c. convex on $\mathbb{R}^n$

Then $\forall\ x, y \in \mathbb{R}^n$.

$$\underline{x\cdot y = \varphi(x) + \varphi^*(y) \Leftrightarrow y \in \partial\varphi(x)} \Leftrightarrow x \in \partial\varphi^*(y)$$

$$x\cdot y = \varphi(x) + \varphi^*(y) \Leftrightarrow x\cdot y \geq \varphi(x) + \underbrace{\varphi^*(y)}_{\text{u}}$$

$$\Leftrightarrow x\cdot y \geq \varphi(x) + y\cdot z - \varphi(z), \quad \forall z$$

$$\Leftrightarrow \varphi(z) \geq \varphi(x) + \langle y, z-x\rangle, \quad \forall z.$$

$$\Leftrightarrow y \in \partial\varphi(x).$$

By symmetry $x \in \partial\varphi^*(y)$

6). Regularization

inf-convolution: $\boxed{\text{下卷积}}$??

φ, ψ. two convex proper

$$(\varphi \,\square\, \psi)(z) = \inf_{x+x'=z} (\varphi(x) + \psi(x'))$$

$$(\varphi \square \psi)^* = \varphi^* + \psi^*$$

$$(\varphi \square \psi)^*(y) = \sup_z \left(z \cdot y - (\varphi \square \psi)(z) \right)$$

$$= \sup_z \sup_{x+x'=z} \left(z \cdot y - \varphi(x) - \psi(x') \right) = \varphi^* + \psi^* .$$

7) <u>Duality and l.s.c.</u>

Let
$\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ proper.

Then TFAE: $\begin{cases} \text{i) } \varphi \text{ is l.s.c and convex;} \\ \text{ii) } \varphi = \psi^*, \text{ for } \psi \text{ proper;} \\ \text{iii) } \varphi^{**} = \varphi , \end{cases}$

- If φ is strictly convex in the nbhd of some $x \in \mathbb{R}^n$, then φ^* is differentiable on $\partial\varphi(x)$, and $\nabla\varphi^*(y) = x$ for all $y \in \partial\varphi(x)$
- If φ is differentiable and strictly convex, then so is φ^*, and $\nabla\varphi$ is one-to-one,

$$(\nabla\varphi)^{-1} = \nabla\varphi^*$$

If φ is superlinear, i.e.

$$\lim_{|x|\to\infty} \frac{\varphi(x)}{|x|} = +\infty,$$

then $\nabla\varphi(\mathbb{R}^n) = \mathbb{R}^n$.

$\nabla\varphi$ is a bijection $\mathbb{R}^n \to \mathbb{R}^n$, and

$\nabla\varphi^*$ is ~~its~~ inverse.

Rk: There is a duality correspondence between

<u>strictly convexity of φ</u> and smoothness of φ^*

$$\nabla\varphi^* \circ \nabla\varphi = Id$$

$$D^2\varphi^*(\nabla\varphi) \cdot D^2\varphi = Id$$

$$\Rightarrow \quad D^2\varphi^*(\nabla\varphi) = [D^2\varphi(x)]^{-1}$$

(Duality between strict convexity and smoothness.)

$D_A^2\varphi(x_0)$ 2nd differentiability

φ : λ- uniformly convex $(\lambda > 0)$

if $D^2_{D'}\varphi \geq \lambda I_n$ on $\mathbb{R}^n$;

semi-convex with $C>0$, if

$$D^2_{D'}\varphi \geq -C I_n \text{ on } \mathbb{R}^n.$$

($\varphi(x) - \lambda \frac{|x|^2}{2}$ — convex $\iff \varphi$: λ-uniformly convex

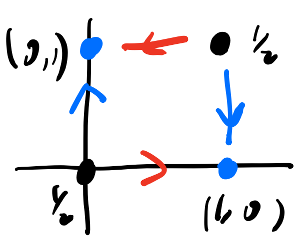
$\varphi(x) + C \frac{|x|^2}{2}$ — convex $\iff \varphi$: semi-convex.

$\langle \nabla\varphi(x) - \nabla\varphi(y), x-y \rangle \geq \lambda |x-y|^2.$

$\langle \nabla\varphi(x) - \nabla\varphi(y), x-y \rangle \geq -C |x-y|^2.$)

$\boxed{\text{RK 2013}}$ Good examples

In general, no uniqueness for the Kantorovich problem.



$\mu = \frac{1}{2}(\delta_{(0,0)} + \delta_{(1,1)})$

$\nu = \frac{1}{2}(\delta_{(1,0)} + \delta_{(0,1)})$

infinitely many

Cover the following computations : Page 74

Recover the duality by working out the Euler-Lagrange eq. associated with the dual optimization problem.

$$\inf_{\pi} \langle c, \pi \rangle \overset{\text{duality}}{=} \sup_{\phi} \left(\langle \phi, \mu \rangle + \langle \phi^c, \nu \rangle \right)$$

character[illegible] ϕ^c

(Gangbo, Caffarelli (give credit to Varadhan).)

Lagrange

The method : (maybe useful to address other variational problems for which there is no such a good duality theory as for Monge-Kantorovich problem)

a perturbative analysis.

(I guess this method must be very old.)

Setting : Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$, $\operatorname{supp}(\mu) = X$.
$\operatorname{supp}(\nu) = Y$.

(φ, φ^*) optimizing pair of the dual minimization problem, φ. convex, l.s.c. on $\mathbb{R}^n$

Existence of the maximizing pair can be

done without refering to Kantorovich duality (see Thm 2.9).

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

$\varphi \to \frac{1}{2}|x|^2-\varphi$

$\inf_{\tilde{\Phi}} \langle \phi, \mu \rangle + \langle \psi, \nu \rangle$

$\tilde{\Phi} \Leftarrow \phi(x)+\psi(y) \geq x\cdot y$

Now, we perform some variations of this optimizer.

$\left(\phi(x)+\psi(y) \leq \frac{1}{2}|x-y|^2 \right)$

$0<t<<1$. $\quad h \in C_b(\mathbb{R}^n)$

$(\varphi^{**}, \varphi^*) \sim$ optimal pair

φ^*, φ^{**}-convex

introduce the pair of dual functions

do perturbation

$$\underbrace{((\varphi^*+th)^*}_{\pi}, \underbrace{\varphi^*+th}_{\pi}) \in \tilde{\Phi}$$

$\uparrow$ not necessarily convex

$((\varphi^*+th)^*, \varphi^*+th)$

Since (φ, φ^*) is optimal,

$$J((\varphi^*+th)^*, \varphi^*+th) \geq J(\varphi, \varphi^*),$$

$$\langle (\varphi^*+th)^*, \mu \rangle + \langle \varphi^*+th, \nu \rangle$$

i.e.

$$\geq \langle \varphi, \mu \rangle + \langle \varphi^*, \nu \rangle$$

$$\left\langle \frac{(\varphi^*+th)^*-\varphi}{t}, \mu \right\rangle + \langle h, \nu \rangle \geq 0 \Leftarrow$$

$\searrow t\to 0$.

Pass to the limit as $t \downarrow 0$.

For simplicity, assume that X & Y are compact, and denote by y_t a point where

$y \mapsto x \cdot y - (\varphi^* + th)(y)$ achieves its maximum

$\left((\varphi^* + th)^*(x) = \sup_y \left(x \cdot y - (\varphi^* + th)(y) \right) \right)$

Of course.

y_t = function of x, and $y_0 = \nabla \varphi(x)$

One can check that $y_t \to y_0$ as $t \downarrow 0$. (Why?)

Then $\forall x \in \mathrm{Int}(X)$ and $t > 0$, $t \ll 1$,

$$-h(y_0) \le \frac{(\varphi^* + th)^*(x) - \varphi(x)}{t} \le -h(y_t)$$

$$\left(\Leftarrow \quad (\varphi^* + th)^*(x) = x \cdot y_t - (\varphi^* + th)(y_t) \right.$$

$$= x \cdot y_t - \varphi^*(y_t) - th(y_t)$$

$\Rightarrow$ RHS $\le -h(y_t)$

$\varphi = \varphi^{**}$ $\quad$ y_0 $\quad$ $\dfrac{(\varphi^* + th)^*(x) - \varphi^{**}(x)}{t}$

$$(\varphi^* + th)^*(x) \le x \cdot y_0 - (\varphi^* + th)(y_0)$$

$$\varphi^{**}(x) = x \cdot y_0 - \varphi^*(y_0).$$

$$\frac{(\varphi^* + t h)^*(x) - \varphi^{**}(x)}{t} \geq -h(y_0)$$

By LDCT, one can then pass $t \downarrow 0$

in $$0 \leq \int_Y h \, d\nu + \int_X \underbrace{\left[\frac{(\varphi^* + t h)^* - \varphi}{t} \right]}_{\downarrow \; -h(\nabla\varphi(x))} d\mu$$

i.e. $$\boxed{0 \leq \int_Y h \, d\nu - \int_X h(\nabla\varphi(x)) \, d\mu}$$

Replacing h by $-h$, we conclude that

$$\boxed{\int_X h \circ \nabla\varphi \, d\mu = \int_Y h \, d\nu} \quad \forall h \in C_b(\mathbb{R}^d)$$

$$\hookrightarrow \boxed{(\nabla\varphi)_\# \mu = \nu.} \quad T = \nabla\varphi.$$

(This computation may be useful for other optimization problems: to derive a necessary condition.)

THE REAL LINE CASE

$\hookrightarrow$ Section 2.2 in textbook.

$supp(\pi) \subset \partial^c \varphi$

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

On $\mathbb{R}^1$, $\nabla\varphi = \varphi'(x)$ = nondecreasing function

φ-convex

$T = \nabla\varphi$, φ-convex

any sub-gradients are "complete nondecreasing graph", or maximal monotone subsets of $\mathbb{R}^2$.

Def:

$\Gamma \subset \mathbb{R}^2$ is said to be monotone, if

$$\left.\begin{matrix}(x_1, y_1)\\(x_2, y_2)\end{matrix}\right\} \in \Gamma \Rightarrow (x_1 - x_2)\cdot(y_1 - y_2) \geq 0$$

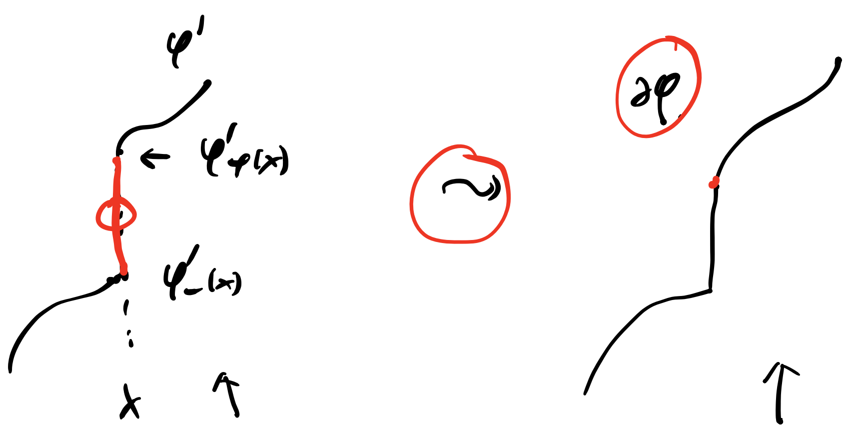
(i.e. either $\begin{cases} x_1 \leq x_2 \\ y_1 \leq y_2 \end{cases}$ or $\begin{cases} x_1 \geq x_2 \\ y_1 \geq y_2 \end{cases}$.)

几何上：

$\boxed{\text{Complete Nondecreasing Graph}}$

= usual graph of $\uparrow$ function

+ some vertical lines added to make the graph continuous



$\boxed{\text{Derivative}}$ $\boxed{\text{Sub-differential}}$

$\boxed{\varphi\text{-convex in 1D}}$

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

$\mathrm{supp}\,P(\pi) \subset \partial\varphi$

Cumulative Distribution Function 累积分布函数 (CDF)

Any $\mu \in \mathcal{P}(\mathbb{R})$ can be represented by its distribution function, or, its CDF $\boxed{\text{CDF}}$

Durrett

$$F(x) = \int_{-\infty}^{x} d\mu = \mu\left[(-\infty, x]\right].$$

闭

$$\bigcup_{n} (-\infty, x - \tfrac{1}{n}] = (-\infty, x).$$

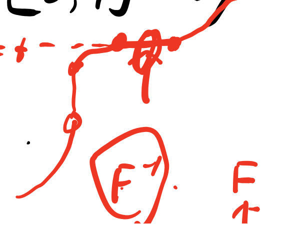
- F is right-continuous. $\left(\bigcap_{n=1}^{\infty} (-\infty, x + \frac{1}{n}] = (-\infty, x] \right)$ 左极右连
- $F \nearrow$ (单增)
- $F(-\infty) = 0$, $F(+\infty) = 1$. $\boxed{F \nearrow}$

$= \lim_{x \to -\infty} F(x)$ $= \lim_{x \to +\infty} F(x)$

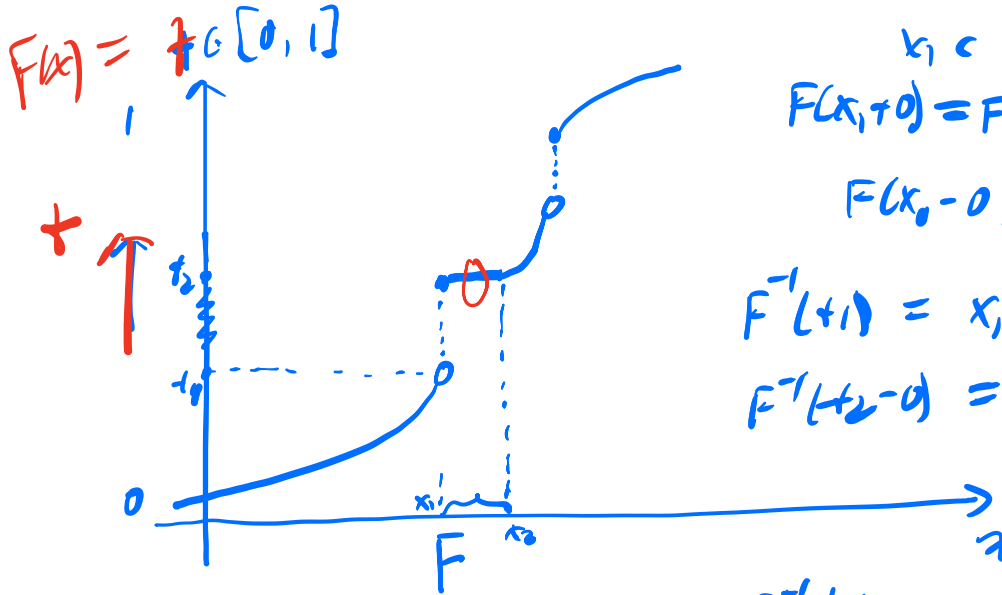
★Generalized inverse of F on $[0,1]$ by

广义逆

$$F^{-1}(t) = \inf\{ x \in \mathbb{R} \mid F(x) > t \}.$$



- F^{-1} is also right-continuous.



$x_1 < x_2$

$F(x_1+0) = F(x_2) = t_2$

$F(x_1-0) = t_1 < t_2$

$F^{-1}(t_1) = x_1$

$F^{-1}(t_2 - 0) = x_1$

$\Leftarrow \begin{cases} F^{-1}(t_2+0) = x_2. \\ F^{-1}(t_2) = x_2 \end{cases}$

跟定义方式有关

But here $F^{-1}(t) = \inf\{x \in \mathbb{R} \mid F(x) > t\}$

↑ right-continuous

- Also,

$\forall x \in \mathbb{R}$, $\boxed{F^{-1}(F(x)) \geq x}$

$\begin{cases} F^{-1} \circ F \geq Id \\ F \circ F^{-1} \geq Id \end{cases}$

$\underbrace{F(x)}_{t}$

$F^{-1}(t) = \inf\{y \mid F(y) > F(x)\}$

$= x.$ or $x + \underbrace{\text{jump}}_{>0}.$

$\forall t \in [0,1]$, $\boxed{F(F^{-1}(t)) \geq t.}$

(say $y = F^{-1}(t) = \inf\{x \mid F(x) > t\}$.

$x_n \downarrow y$, $F(x_n) > t$.

$F(y) = \lim F(x_n) \geq t$.

↑ right-continuous.)

For probability measures on $\mathbb{R} \times \mathbb{R}$.

$\Updownarrow$ $\qquad\qquad$ $\mu((-\infty, x]) = F(x)$. CDF in 1D

joint 2D $\boxed{\text{CDF}}$

$$H(x_0, y_0) = \int_{\substack{x \le x_0 \\ y \le y_0}} d\lambda = \lambda[R(x_0, y_0)]$$

$\left.\begin{matrix} x \le x_0 \\ y \le y_0 \end{matrix}\right\} \overset{\Delta}{\Rightarrow} R(x_0, y_0)$

$d\lambda = dH$

Check (see Durrett for example)

A function H on $\mathbb{R}^2$ which
$\left\{\begin{array}{l} \bullet \nearrow \; x, y \\ \bullet \text{ right-continuous in } x, y. \end{array}\right.$

$\bullet \lim_{\substack{x \to -\infty \\ y \to -\infty}} H(x, y) = 0$

$\bullet \lim_{\substack{x \to +\infty \\ y \to +\infty}} H(x, y) = 1$

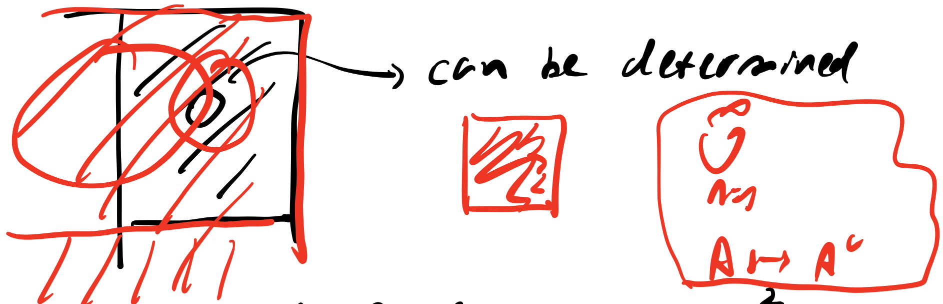
$(-\infty, x_0] \times (-\infty, y_0]$

H $\Rightarrow$ $\boxed{\pi \in P(\mathbb{R}^2)}$

gives rise to a unique probability measure on

$\mathbb{R}^2$ Direct exercise $H \rightleftarrows \pi$

(Note:



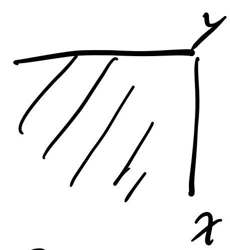
rectangles generates all Borel sets in $\mathbb{R}^2$.

<u>Thm</u> (Optimal transportation theorem $\{ c(x,y) = c(|x-y|)$

for $c(x,y) = \frac{1}{2}|x-y|^2$ on $\mathbb{R}$.)   c-convex

Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ with respective CDFs F and G. Let π be the probability measure on $\mathbb{R}^2$ with joint 2D CDF. $\pi \leftrightarrow H$, $d\pi = dH$

$$\star \quad \boxed{H(x,y) = \min\{F(x), G(y)\}},$$



then $\pi \ (d\pi = dH) \in \Pi(\mu,\nu)$

and $\pi \in OPT(\mu,\nu)$ with $c(x,y) = \frac{1}{2}|x-y|^2$.

$\pi = (F^{-1} \times G^{-1})_{\#} \mathrm{Leb}_{[0,1]}$

Moreover, the optimal transport cost is

$$T_2(\mu,\nu) = \frac{1}{2}\int_0^1 |F^{-1}(t) - G^{-1}(t)|^2\, dt.$$

$$\begin{aligned} T_2(\mu,\nu) &= W_2^2(\mu,\nu) \\ &\frac{1}{2}\int |x-y|^2\, d\pi(x,y) \\ &= \frac{1}{2}\int |x-y|^2\, d\left(F^{-1}\times G^{-1}\right)_\# \mathrm{Leb}|_{[0,1]} \\ &= \frac{1}{2}\int_0^1 |F^{-1}(t) - G^{-1}(t)|^2\, dt \end{aligned}$$

可以 = 计算结果.

Remarks:

i) **Hoeffding – Fréchet theorem:**

For $H: \mathbb{R}^2 \to \mathbb{R}_+$, $H \nearrow$, right-continuous in each arguments, define a probability measure π on $\mathbb{R}^2$ with given marginals μ and ν iff.

$$\forall (x,y) \in \mathbb{R}^2, \quad F(x) + G(y) - 1 \leq H(x,y) \leq \min\{F(x), G(y)\}$$

where F and G are CDFs of μ and ν resp.

Checking:

$$H(x,y) = \min\{F(x), G(y)\}$$

- $H \geq 0$; • $H \nearrow$; • Right-continuous ✓

- $H(-\infty, -\infty) = 0 \qquad H(+\infty, +\infty) = 1.$

- $$\int_{[-\infty,x]\times\mathbb{R}} d\pi(x,y) = \int_{[-\infty,x]\times\mathbb{R}} dH(x,y)$$
$$= H(x,+\infty) = F(x).$$

okay

similarly

$$\int_{\mathbb{R}\times[-\infty,y]} d\pi(x,y) = H(+\infty,y) = G(y)$$

Also checking the conditions:

$$\underline{\underline{F(x) + G(y) - 1}} \le H(x,y) \le \underbrace{\min[F(x), G(y)]}_{\text{okay}}$$

Let $H_L(x,y) = F(x) + G(y) - 1 \ge 0$.

$$H_L(x,+\infty) = F(x)$$
$$H_L(+\infty,y) = G(y)$$

; marginalization is correct.

ii) This π defined as

$$d\pi = dH,$$
$$H(x,y) = \min(F(x), G(y))$$

is optimal. as long as $c(x,y) = c(x-y)$.

c - convex. $c \ge 0$.

and the optimal cost is

$$\mathcal{T}_c(\mu,\nu) = \int_0^1 c(F^{-1}(t) - G^{-1}(t))\,dt$$

but we still need c is convex.

iii) for $c = |x-y|$, ✓

$$T_1(\mu,\nu) = \int_0^1 |F^{-1}(t) - G^{-1}(t)|\,dt$$

exercise

$$= \underbrace{\int_{\mathbb{R}} |F(x) - G(x)|\,dx}_{L^1 \text{ between CDFs.}}$$

(Fubini)

(Integrate over x or integrate over t.)

iv) If μ does not give mass to points,

then $T = G^{-1} \circ F$ transports μ to ν,

and

$$\int_{-\infty}^{x} d\mu \stackrel{?}{=} \int_{-\infty}^{T(x)=y} d\nu$$

$H(x,y) = \min\{F(x), G(y)\}$

General fact: the solution to the transportation prob.

is given by the monotone rearrangement of μ onto ν.

(One proceeds to transfer the sand into the hole starting from the left.)

Note that discontinuity points of G correspond to atoms for ν.

When ν has an atom, $G^{-1} \circ F$ will then be constant on some interval

(when encountering an atom in the filling process, one must keep putting mass in this hole for some time.)

v). Assume that $d\mu(x) = f(x)\,dx$, $d\nu(x) = g(x)\,dx$, $f, g \in C^0 \cap L^1$, $g > 0$ pointwise.

Then $T = G^{-1} \circ F$ is C^2, differentiating

$$\int_{-\infty}^{x} d\mu = \int_{-\infty}^{T(x)} d\nu,$$

i.e. $\int_{-\infty}^{x} f(y)\,dy = \int_{-\infty}^{T(x)} g(y)\,dy$ Taking derivative one obtains

$$\boxed{f(x) = g(T(x))\, T'(x).}$$

$T = \varphi'$

φ - convex

$f(x) = g(\varphi'(x))\,\varphi''(x)$

1D MA eq.

(Real Monge - Ampère Eq. in 1D)

$f(x) = g(\nabla\varphi(x))\,\det(\nabla^2\varphi(x))$

1D case $\overset{\text{Generalize}}{\Longrightarrow}$ Distributions and cost function are radially symmetric in multi-Dimen.

Proof of O.T. theorem for $c = \frac{1}{2}|x-y|^2$ on $\mathbb{R}^1$

F : CDF. $\quad F(x-)$ Left limit (always exists)

μ

$F(x+) = F(x)$ (since right continuous).

Step I : $\quad d\pi = dH(x,y)$ with $H(x,y) = \min\{F(x), G(y)\}$

Claim : $\operatorname{Supp}(\pi) \subset \{(x,y) \in \mathbb{R}^2 \mid F(x-) \le G(y)$ and and $G(y-) \le F(x)\}$

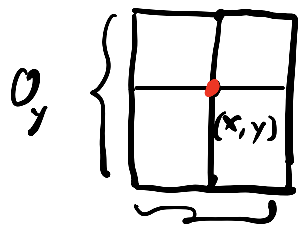
- $F(x-) \le G(y)$

If not, $F(x-) > G(y)$. $\quad (F(x) \geq F(x-) > G(y))$

For $x' \in O_x$,

$y' \in O_y$,

$F(x') > G(y')$



$F(x) > G(y)$

O_x ~ chosen nbhd of x

Hence $\forall (x', y') \in O_x \times O_y$, $F(x') > G(y')$

$$H(x', y') = \min\{F(x'), G(y')\} = G(y')$$

$H(x', y')$ in $R = O_x \times O_y$ only depends on y.

$\hookrightarrow$ $d\pi = dH$ assigns zero mass to this rectangle R, i.e. $(x, y) \notin \mathrm{supp}(\pi)$.

Step II:

Claim: $\mathrm{supp}(\pi)$ is monotone (c-monotone) in the sense $|x-y|^2$ [illegible] x y that if (x_1, y_1), $(x_2, y_2) \in \mathrm{supp}(\pi)$, then $\boxed{(x_1 - x_2)(y_1 - y_2) \geq 0}$

WLOG, assume that $x_1 > x_2$, we need to show $y_1 \geq y_2$. 需证

$\boxed{x_1 > x_2}$

Applying the claim in Step I:

$$G(y_1) \geq F(x_1 -) \geq F(x_2) \geq G(y_2 -)$$

If $G(y_1) > G(y_2-)$ then of course $y_1 \geq y_2$, then the proof is done!

Now assume that $G(y_1) = G(y_2-)$, then

$$\underline{G(y_1) = F(x_1-) = F(x_2) = G(y_2-)}$$

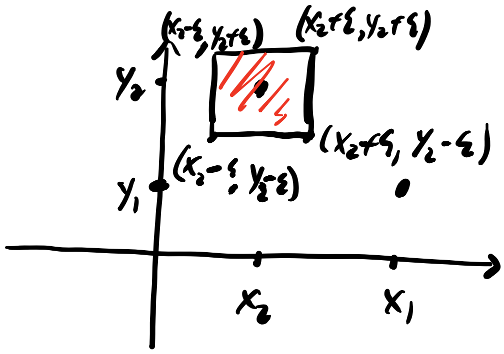
with $x_1 > x_2$.

Proof by contradiction. Assume $y_2 > y_1$

Choose $\varepsilon > 0$ small enough

consider the rectangle

with coordinates



$(x_2 \pm \varepsilon, y_2 \pm \varepsilon)$, denote this rectangle as R^ε

$$\pi(R^\varepsilon) = H(x_2+\varepsilon, y_2+\varepsilon) + H(x_2-\varepsilon, y_2-\varepsilon) - H(x_2-\varepsilon, y_2+\varepsilon) - H(x_2+\varepsilon, y_2-\varepsilon)$$

$$= \min\{\underline{\overset{F(x_2)}{F(x_2+\varepsilon)}}, G(y_2+\varepsilon)\} + \min\{F(x_2-\varepsilon), \underline{\overset{G(y_1)}{G(y_2-\varepsilon)}}\} - \min\{F(x_2-\varepsilon), G(y_2+\varepsilon)\} - \min\{F(x_2+\varepsilon), G(y_2-\varepsilon)\}$$

$\equiv 0$ for ε small enough

$\Rightarrow (x_2, y_2) \notin \mathrm{supp}(\pi)$ Contradiction!

Hence $y_1 \geq y_2$.

Step II: We have showed that supp(π) $\subset$ a monotone subset of $\mathbb{R}^2$, hence complete monotone

$$\mathrm{supp}(\pi) \subset \partial \varphi, \text{ for } \varphi\text{-convex l.s.c.}$$

By F.T. of O.T., $\pi \in OPT(\mu, \nu)$

Checking:

We claim: $\pi = \underbrace{(F^{-1} \times G^{-1})_\#}\ell_{[0,1]}$ (*)

Lebesg m.

Indeed, it suffices to show that the identity holds on an arbitrary rectangle of the form $R(x,y) = (-\infty, x] \times (-\infty, y]$.

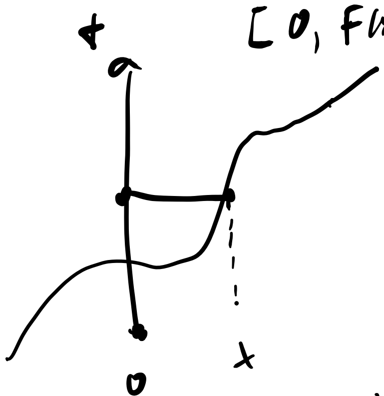
Need to check: $\pi[R(x,y)] = \left| (F^{-1}\times G^{-1})^{-1}\big((-\infty, x]\times(-\infty, y]\big) \right|$

def $\|$ $\quad = \left| (F^{-1}(t), G^{-1}(t)) \in R(x,y) \right|$

$\min\{F(x), G(y)\} = = \left| \{F^{-1}(t) \le x\} \cap \{G^{-1}(t) \le y\} \right|$

$\left|\{t \mid F^{-1}(t) \le x\}\right| = F(x)$, $\quad \left|\{t \mid G^{-1}(t) \le y\}\right| = G(y)$

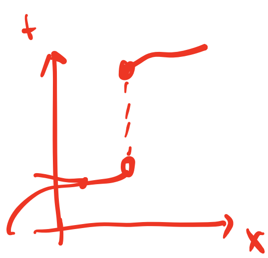
$\|$ $[0, F(x))$ or $[0, F(x)]$. $\qquad [0, G(y))$ or $[0, G(y)]$.



i.e. $\pi = (F^{-1} \times G^{-1})_\# \mathcal{L}$

with

$d\pi(x,y) = dH(x,y)$, with



$H(x,y) = \min\{F(x), G(y)\}$.

Step II: For any u - measurable

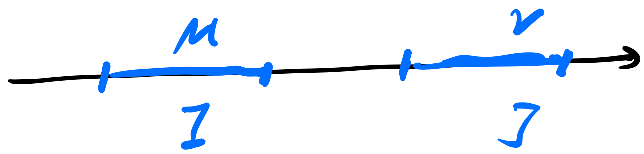
$$\int_{\mathbb{R}^2} \overset{C}{u}(x,y)\, d\pi(x,y) = \int_{\mathbb{R}^2} u(x,y)\, d(F^{-1}\times G^{-1})_\# \mathcal{L}_{[0,1]}$$

$$= \int_0^1 c\left(F^{-1}(t),\, G^{-1}(t)\right) dt.$$

$\blacksquare$

Exercise:

$I \cap J = \phi.$



I, J: two disjoint intervals.

$\mu, \nu \in \mathcal{P}(\mathbb{R})$

$\mathrm{supp}(\mu) \subset I,$

$\mathrm{supp}(\nu) \subset J,$

$I \cap J \neq \phi$

Consider $c(x,y) = |x-y|$

then any $\pi \in \Pi(\mu,\nu)$ is OPTIMAL.

Hint: $\varphi(x) = Ix$ achieves equality

Kantorovich duality

Brenier's Polar Factorization Theorem continued

$\nabla\varphi$ $(\partial\varphi)$ for φ: convex.

$(\nabla\varphi)_{\#}\mu = \nu \Rightarrow T = \nabla\varphi$. optimal map between μ and ν

Brenier's polar factorization theorem:

any "non degenerate" vector-valued mapping can be rearranged into the gradient of a convex function. $S \in L^2(\Omega; \mathbb{R}^n)$ S mapping. L^2 w.r.t. Lebesgue

$\Omega \subset \mathbb{R}^n$

$S = \nabla\psi \circ s$

$\nabla\psi$ 是 S 的一个重排

- Why is it related to O.T. ?
- How it was motivated by problems in fluid mechanics ?

Applications: ① Polar factorizations for matrices

Chapter 3

(Helmholtz decomposition)

② Hodge decomposition of vector field. ✓

$X = \mathbb{R}^n$.

- Vector valued mapping $\mathbb{R}^n \to \mathbb{R}^n$.
- vector fields:

mapping from $\mathbb{R}^n$ to $T\mathbb{R}^n$ (tangent bundle)

Def 3.1

Rearrangement:

measure spaces

Let $\underset{\pi}{m} : (\underline{W}, \lambda) \xrightarrow{m} (\underline{X}, \mu)$, measurable. $\quad \xrightarrow{F} \mathbb{R}$

Another function $\underset{\pi}{\tilde{m}} : (W, \lambda) \xrightarrow{\tilde{m}} (X, \mu)$ is $\quad \xrightarrow{F} \mathbb{R}$

said to be a rearrangement of m, if

$\forall\, F : X \to \mathbb{R}$ measurable, $\underline{F \circ m} \in L^1(d\lambda)$, then

$\underline{F \circ \tilde{m}} \in L^1$ and

$$\underline{\int_W (F \circ m)\, d\lambda} = \underline{\int_W (F \circ \tilde{m})\, d\lambda}.$$

$\tilde{m}$ is a rearrangement of m

Sobolev $\quad$ Lieb

$$\left((W, \lambda) \underset{\tilde{m}}{\overset{m}{\rightrightarrows}} (X, \mu) \xrightarrow{F} \mathbb{R}. \right.$$

(• if $\lambda[W] < \infty$, one can require only $F \in L^\infty$;

• One can also require that for any $F \geq 0$.)

Def 3.1 means that: one cannot tell the difference between m and $\tilde{m}$ by looking only at their values".

Same maximum

minimum etc. ...

mean

If $X = \mathbb{R}^n$, $F(x) = |x|^p$, we obtain that.

$$\|m\|_{L^p} = \|\tilde{m}\|_{L^p}, \quad \forall p.$$

i.e. Lebesgue norms (L^p) are invariant under rearrangement. $W^{k,p}$ ✗

But $\|\nabla m\|_{L^p}$ are not invariant,

even for smooth m, $\|\nabla \tilde{m}\|_{L^p}$ can be anything.

$$S = \nabla \psi \circ S$$

$$S \sim \nabla \psi$$

Measure-preserving maps:

Let (W, λ) be a given measure space. 重排之间差一个保测映射.

A measurable function $S: W \to W$ is said to be measure-preserving if $S_\# \lambda = \lambda$. 保测(λ)

Or: $\forall A \subset W$ measurable, $\underbrace{\lambda[S^{-1}[A]]}_{S_\# \lambda[A]} = \lambda[A]$.

Example: $\Omega \subset \mathbb{R}^n$ $\quad (\Omega, \lambda)$ $\quad$ matrix

open $\qquad$ Lebesgue $\qquad$ ∇S

C^1-diffeomorphism $S: \Omega \circlearrowleft$ is measure preserving

$$\iff \boxed{|\det(\nabla S)| = 1}$$

$+1, -1$

$S: \Omega \xrightarrow{C^1} \Omega$, $\quad \Omega \subset \mathbb{R}^n$

Those S form a group of measure preserving diffeomorphisms. (group "$\circ$" composition)

A important subgroup is the group of all diffeomorphism S with $\det(\nabla S) \equiv 1$.

Checking: $\quad S_{\#}\lambda = \lambda \qquad \forall \phi$ smooth

$$\int_\Omega \phi(x)\, d S_{\#}\lambda(x) = \int_\Omega \phi(x)\, d\lambda(x)$$

$$= \int_\Omega \phi(S(x))\, dx = \int_\Omega \phi(x)\, dx$$

$$= \int_\Omega \phi(S(y))\, \underbrace{|\det \nabla S(y)|}_{=1}\, dy$$

Denote $S(W) = \{ S: (W, \lambda) \circlearrowleft,\ S_{\#}\lambda = \lambda \}$ 所有保测映射

$SD(\Omega) = \{ s : \Omega \circlearrowleft$ diffeomorphism $\mid |\det(\nabla s)| = 1 \}$

open

(Ω, λ)

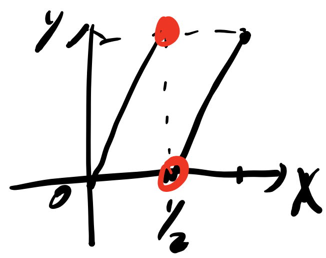
Lebesgue

$G(\Omega) = \{ s \in SD(\Omega) \mid \det(\nabla s) \equiv 1 \} \subseteq SD(\Omega)$

$SD(\Omega) \overset{\text{small part}}{\subsetneq} S(\Omega)$, for instance. $\Omega = (0,1)$

$SD((0,1)) = \{ Id, -Id \}$ $\quad D((0,1)) = \{ Id \}$

but there are many elements which are not diffeomorphisms $S((0,1))$



$x \mapsto 2x \pmod 1$

Exercise: $S(\Omega) \subset L^2(\Omega)$, $S(\Omega)$ is not convex

$\frac{s_1+s_2}{2} \notin S(\Omega)$ even $s_1 \in S(\Omega)$, $s_2 \in S(\Omega)$

Prop. 3.7: Measure-preserving maps and rearrangement.

$S \sim \circ \psi, \Leftrightarrow S = \circ \psi \circ S$

Let (W, λ) be a measure space.

If $s \in S(W)$ and $\tilde{m} = m \circ s$, then $\tilde{m}$ is a rearrangement of m;

"Conversely", if $\tilde{m}$ is 1-1 rearrangement of m, then $\tilde{m}^{-1} \circ m \in S(W)$.

Pf: (1) Assume $\tilde{m} = m \circ s$ with $s \in S(W)$.

Then, for all measurable $F : W \to \mathbb{R}_+$,

$$\int_W F \circ \tilde{m} \, d\lambda = \int_W (F \circ m) \circ s \, d\lambda$$

$$= \int_W F \circ m \, d s_\# \lambda$$

$$= \int_W (F \circ m) \, d\lambda,$$

so $\tilde{m}$ is a rearrangement of m.

(2) Let $\tilde{m}$ is 1-1 rearrangement of m.

Define $s = \tilde{m}^{-1} \circ m$ and consider any measurable non-negative function F on W

$$\int_W F \, d s_\# \lambda = \int_W F \circ s \, d\lambda = \int_W (F \circ \tilde{m}^{-1}) \circ m \, d\lambda$$

$$= \int (F \circ \tilde{m}^{-1}) \circ \tilde{m} \, d\lambda = \int F \, d\lambda$$

i.e. $s_\# \lambda = \lambda$.

13

Now we restrict to the case $W = \mathbb{R}^n$.

Question: ? $\exists$ a class $\mathcal{R}$ of functions with nice properties, s.t. any measurable function $m: W \to X$ admits a rearrangement in $\mathcal{R}$.

$S \overset{EL^2}{=} \nabla\varphi \circ s$ $\quad S: \Omega \to \mathbb{R}^n$, $\Omega \supset \mathbb{R}^n$

- $W = \mathbb{R}^n$, $X = \mathbb{R}_+$, $\quad \mathcal{R} = \{ f \mid f = F(|x - x_0|),\ x_0 \in \mathbb{R}^n,\ F \geq 0,\ F \nearrow \}$

(radially symmetric monotone rearrangements)

(This is about scalar functions.)

- Now $W = \mathbb{R}^n$, $X = \mathbb{R}^n$ $\quad \mathcal{R} \sim \nabla\varphi$, φ-convex.

Again Restate Brenier's famous theorem

Thm (Brenier) $\qquad$ Lebesgue measure

Let $\Omega \subseteq \mathbb{R}^n$ open, $|\Omega| > 0$. Let $h: \Omega \to \mathbb{R}^n$ be an L^2-vector-valued mapping with $\quad \lambda \overset{0.T}{\sim} h_\# \lambda$

the nondegeneracy condition : $\nu =$ regular

(*) for small set N in $\mathbb{R}^n$, $|h^{-1}(N)| = 0$ $h_{\#}\lambda(N) = 0$

($\Leftrightarrow$ $h_{\#}\lambda$ is regular) $\dim_{\mathcal{H}} N \le n-1$ small set

$\leftarrow$ Lebesgue

Then $\exists !$ rearrangement $\nabla\psi$ of h in the class of L^2-gradients of convex functions and a unique measure preserving $s \in S(\Omega)$, s.t.

$$h = \nabla\psi \circ s.$$

Moreover, s is the unique L^2 projection of h onto $S(\Omega)$.

RK: (i) L^2-norm is $L^2(d\lambda)$ $\leftarrow$ Lebesgue

$\nabla\psi$ = the restriction of $\nabla\psi$ to Ω

$\uparrow$ gradient of a convex function on $\mathbb{R}^n$.

$$s = \arg\min_{\tilde{s} \in S(\Omega)} \int_\Omega |h - \tilde{s}|^2 dx$$

ii) $h:\Omega\to\mathbb{R}^n$ (vector-valued mapping)

should not be understood as a tangent vector field, but as a plain mapping.

$\nabla\varphi$: regarded as a mapping.

iii) $h_{\#}\lambda$ is regular, not a necessary condition for existence

but for uniqueness.

iv) Brenier's theorem has an intrinsic formulation in the sense of Riemanian Geometry:

Let M be a compact Riemannian manifold, Let λ be the normalized volume on M, and let $h: M\to M$ be a measurable map s.t. $h_{\#}\lambda$ is A.C. w.r.t. λ. Then $\exists!$ pair $(\nabla\varphi, s)$ s.t.

$h = \nabla\varphi\circ s$

$h(x) = \exp_{s(x)}(-\nabla\varphi(s(x)))$, where s is measure preserving, and φ is $d^2/2$ - concave. Moreover,

S is the unique solution of the minimization problem

$$\min\left\{\int_M d(h(x), \sigma(x))^2\,dx \ : \ \sigma_\# \lambda = \lambda\right\}.$$

(we skip part iv)

Brenier's orginal motivations:

from fluid mechanics.

(Projection operator onto the set of measure-preserving maps)

§7.2.1. Incompressible Euler eq.

$\downarrow$

1st PDE ever written down

Models an incompressible, inviscid fluid in a bounded, smooth open set $\Omega = \mathbb{R}^n$ $(n=2, n=3)$.

(or the whole space)

The unknown:

the $\boxed{\text{velocity field}}$ of fluid

$v = v(t, x) : \mathbb{R}_t \times \Omega \to \mathbb{R}^n$

Euler Eq. reads: (tangent space to Ω)

$\mathbb{R}^3$ $n=3$ $v=\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$ $\in \mathbb{R}^n$.

$$\begin{cases} \partial_t v + \boxed{v \cdot \nabla v} + \nabla p = 0 \\ \nabla \cdot v = 0. \end{cases}$$

div v

$p = p(t,x) \in \mathbb{R}$ pressure

see ∇p as a Lagrangian multiplier

Here $\nabla \cdot v = \sum_i \frac{\partial v_i}{\partial x_i}$ (divergence)

$v=\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$

$v_i = v_i(t,x)$

$x \in \Omega \subseteq \mathbb{R}^3$

$v \cdot \nabla v = (v \cdot \nabla) v \in \mathbb{R}^n$.

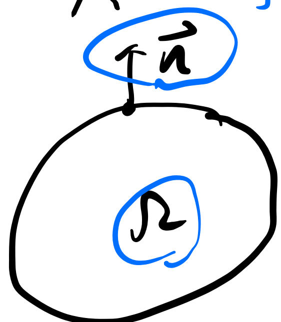
convection term

$(v_j \partial_j v_i)_i$

$(v_1, v_2, v_3) \cdot (\partial_{x_1}, \partial_{x_2}, \partial_{x_3}) = \sum_{j=1}^{n} v_j \partial_{x_j}$

$v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$

Boundary condition:



$v \cdot \vec{n} = 0$ on $\partial\Omega$ （→ 外法向）

or v is tangent to $\partial\Omega$.

$\Rightarrow$ (Lebesgue) measure preserving

$\nabla \cdot v = 0$ (divergence free)

$\downarrow$

incompressible

Cauchy Problem: (1st step)

(initial value)

Existence and uniqueness.

$$\begin{cases} \left.\begin{aligned} &\partial_t v + (v \cdot \nabla) v + \nabla p = 0 \\ &\nabla \cdot v = 0 \end{aligned}\right\} \text{ on } \Omega \\ v|_{t=0} = v_0 \leftarrow \text{initial value} \\ v \cdot \vec{n} = 0 \quad \text{on } \partial\Omega \end{cases}$$

$\underline{\underline{n=2}}$ Youdovich's theorem (Rosenweig's provide approximation using Serfaty's method)

under: $\nabla \wedge v_0$ or curl v_0 (vorticity) or $\nabla^{\perp}\cdot v_0 = \omega_0 \in L^{\infty}$.

$n=3$. $\boxed{\text{OPEN Problem}}$ Also $\boxed{\text{3D Navier-Stokes}}$

$\boxed{\text{A priori estimate}}$: (Energy Estimates)

A natural space $L^2(\Omega;\mathbb{R}^n)$

Energy conservation for smooth solutions:

Let $v \in C^{\infty}$. Then

$$\frac{d}{dt}\int_\Omega |v(t,x)|^2\,dx = 0,$$

i.e. $\|v(t,\cdot)\|^2_{L^2(\Omega;\mathbb{R}^n)}$ is preserved with time.

$\uparrow$ Kinetic Energy

Sketch of Pf: Assume $v \in C^{\infty}$. (for a priori estimate)

$$\frac{1}{2}\frac{d}{dt}\int_\Omega |v|^2 = \int_\Omega v\cdot\partial_t v = \int_\Omega v\cdot(-v\cdot\nabla v) - \underbrace{\int_\Omega v\cdot\nabla p}_{=0}$$

IBP 分部积分

$$\int_\Omega v\cdot\nabla p\,dx = -\int_\Omega \underbrace{(\nabla\cdot v)}_{0}\,p + \cancel{边界项} \quad 0 \quad \because \vec{v}\cdot\vec{n}=0$$

$$= 0$$

(Recall the IBPs formula:

$$\int_\Omega v\,\partial_{x_i} u\,dx = -\int_\Omega u\,\partial_{x_i} v\,dx + \int_{\partial\Omega} u v \nu^i\,dS \quad)$$

unit outer normal vector.

$v_i \partial_j v_i = \frac{1}{2}\partial_j |v_i|^2$

$$\int_\Omega v\cdot(v\cdot\nabla v) = \sum_{i,j}\int_\Omega v_i v_j \partial_j v_i = \sum_{i,j}\int_\Omega v_j\, \partial_j\left(\tfrac{1}{2}|v_i|^2\right)$$

$$= -\sum_{i,j}\int (\partial_j v_j)\cdot \tfrac{1}{2}|v_i|^2 = 0$$

$$\text{"} -\frac{1}{2}\int_\Omega (\nabla\cdot v)\,|v|^2 \quad \text{"} \; 0 .$$

Conservation of Energy

RK: $n=3$, solution to Euler Eq. should not in general have enough regularity for the conclusion above to hold true, so that energy could be decreasing.

"Onsager's conjecture".

$C_x^{\frac{1}{3}-}$

De Lellis, Isett.

§7.2.2. Lagrangian formulation

$$\boxed{\frac{d}{dt}\int |v(t,x)|^2 = 0}$$

For $v = v(t,x)$: unknown is a time-dependent velocity field.

$v = v(t,x)$

$\dot{x} = u(t,x)$

↓ Eulerian formulation $\boxed{x = m(t, x_0)}$

Another equivalent way of description in fluid mechanics: the Lagrangian point of view:
focus on the trajectories of particles

("Lagrangian" point of view ← introduced by Euler;
"Eulerian" - - ← introduced by Bernoulli and D'Alembert.)

- Eulerian: $v(t,x)$

  Fix point of space x,
  measures the velocity $v = v(t,x)$

- Lagrangian: puts a label on each particle
  and then study the trajectory of each labelled particle.

Say $x = m(t, x_0)$ ← initial position as a label.

x denotes the position of a particle that was located at position x_0 at time 0.

$\forall t$, the map: $x_0 \mapsto m(t, x_0)$

flow map

usually assumed to be 1-1.

Switch between these two descriptions:

$$\begin{cases} \dfrac{d}{dt} m(t, x_0) = v(t, m(t, x_0)) \\ m(0, x_0) = x_0 \end{cases}$$

Eulerian expression of the Lagrangian acceleration:

$$\begin{cases} \partial_t v + v \cdot \nabla v + \nabla p = 0 \\ \nabla \cdot v = 0 \end{cases}$$

$$\frac{d}{dt} m(t, x_0) = v(t, m(t, x_0))$$

$$\underbrace{\frac{d^2}{dt^2} m(t, x_0)}_{\text{加速度}} = \underbrace{\left[\frac{\partial v}{\partial t} + v \cdot \nabla v\right]}_{\text{acceleration}} (t, m(t, x_0))$$

$\dfrac{Dv}{dt}$ material derivative

- $m(t, x_1) = m(t, x_2) \Rightarrow x_1 = x_2$

  $m(t, \cdot)$: one-to-one; 单 t, ($x \mapsto m(t, x)$ 一一对应)

$\rho \equiv 1$ also it is natural to assume m $\Omega \ni \Phi^t$

be surjective. $\operatorname{div} v = \nabla \cdot v = 0$

i.e. $(m(t,\cdot))_{t\geq 0}$: a family of diffeomorphism

$$\boxed{\nabla \cdot v = 0 \iff \det\left(\frac{\partial m}{\partial x_0}\right) = 1}$$

$v = v(t,x)$ $\quad x_0 \mapsto m(t,x_0)$ $\quad m(t,\cdot) \in D(\Omega)$

Pf: RHS above holds true for $t=0$

since $m(0,\cdot)$ is the identity map.

Then applying the identity

$$\frac{\partial}{\partial t}\log \det\left[\frac{\partial m}{\partial x_0}\right] = (\nabla \cdot v)(t, m(t,x_0))$$

$\hookrightarrow$ then done. $\frac{d}{dt}\det(\Phi(t)) = ?$

(recall the differential of the determinant

This should leave as an exercise: $\Phi(t) = (\phi_{ij}(t))_{i,j}$

$$\frac{d}{dt}\det(\Phi(t))$$

$$= \frac{d}{dt}\sum_{\sigma\in S_n}(-1)^{|\sigma|}\phi_{1,\sigma(1)}\phi_{2,\sigma(2)}\cdots\phi_{n,\sigma(n)}$$

Another way: $\quad + \nabla \cdot v = 0$

$$\left\{\begin{array}{l}\partial_t \rho + \operatorname{div}(\rho v) = 0 \\ \partial_t \rho + v\cdot\nabla\rho = 0 \\ \boxed{d\rho(t,X_t)/dt \equiv 0}\end{array}\right.$$

$$= \sum_{\sigma} (-1)^{|\sigma|} \dot\phi_{1,\sigma(1)} \phi \cdots \phi_{n,\sigma(n)} + \sum \cdots +$$

$$= \det \begin{pmatrix} \dot\Phi_1 \\ \Phi_2 \\ \vdots \\ \Phi_n \end{pmatrix} + \cdots + \det \begin{pmatrix} \Phi_1 \\ \vdots \\ \dot\Phi_n \end{pmatrix}$$

If

$\Phi(t) = I$

$$\det \begin{pmatrix} \dot\phi_{11} & \dot\phi_{12} & \cdots & \phi_{1n} \\ 0 & 1 & & 0 \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix} = \dot\phi_{11}$$

$$\det \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ \times \times \cdots & & \dot\phi_{ii}(t) & \cdots \times \times & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}$$

By analysis above:

$$\frac{d}{dt} \det(\Phi(t)) = \operatorname{Tr}(\dot\Phi(t)) \quad \text{if} \quad \Phi(t) = I$$

$$\det(A\,\Phi(t)) = \det(A) \det(\Phi(t)),$$

with $A\,\Phi(t) = I$.

$$\underline{\det(A) \frac{d}{dt} \det(\Phi(t))} = \frac{d}{dt} \det(A\,\Phi(t))$$

$$= \underline{\operatorname{Tr}(A\,\dot\Phi(t))}$$

$$\Rightarrow \quad \frac{d}{dt} \det(\Phi(t)) = (\det(A))^{-1} \operatorname{Tr}(A\,\dot\Phi(t))$$

Taking $A = (\Phi(t))^{-1}$

$$\boxed{\frac{d}{dt}\log\det(\Phi(t)) = \mathrm{Tr}\left([\Phi(t)]^{-1}\dot{\Phi}(t)\right)}$$

Applying in our case

$x_0 \to m(t, x_0)$

$\frac{d}{dt} m(t, x_0)$

$= v(t, m(t, x_0))$

$$\frac{d}{dt}\log\det\left[\frac{\partial m(t, x_0)}{\partial x_0}\right]$$

$$= \mathrm{tr}\left(\Phi^{-1}(t)\cdot\frac{\partial v}{\partial m}\cdot\frac{\partial m}{\partial x_0}\right)$$

$x_0 \to v(t, m(t, x_0))$

$$= \mathrm{tr}(ABC) = \mathrm{tr}(CAB) = \mathrm{tr}(B)$$

$$= \nabla\cdot v.$$

$$\begin{cases} \Phi(t) = \dfrac{\partial m(t, x_0)}{\partial x_0} \\ \dot{\Phi}(t) = \dfrac{d}{dt}\dfrac{\partial m(t, x_0)}{\partial x_0} = \dfrac{\partial}{\partial x_0}\left(\dfrac{d}{dt} m(t, x_0)\right) \\ \qquad = \dfrac{\partial}{\partial x_0} \underbrace{v(t, m(t, x_0))} \\ \dfrac{\partial v}{\partial x_0} = \boxed{\dfrac{\partial v}{\partial m}}\dfrac{\partial m}{\partial x_0} \end{cases}$$

Leave as an exercise

After all previous computations:

In Lagrangian formulation:

Euler equation becomes an evolution eq.

for a map: $t \mapsto m(t, \cdot)$

with each $(t \mapsto m(t, \cdot)) \in G(\Omega)$

diffeomorphism with unit determinant.

Use g for the trajectory map.

In particular, g is (Lebesgue) measure preserving (restricted to Ω).

Physical interpretation: the volume of a set of particles is kept constant under time evolution, which is precisely the incompressibility.

$$v(t, g(t, x_0)) = \frac{d}{dt} g(t, x_0)$$

or $v = \frac{\partial g}{\partial t} \circ g^{-1}$.

Then Euler Eq. translates into an eq. on the trajectory field

$$t \mapsto g(t, \cdot) \text{ of } \mathbb{R}_+ \to G(\Omega)$$

$$\frac{d^2}{dt^2} g(t, x_0) = -\nabla p(t, g(t, x_0))$$

Some motivations from fluid mechanics continued.

Recall that $x = m(t, x_0)$

denote now as $x = g(t, x_0)$

Then $\frac{d}{dt} g(t, x_0) = v(t, g(t, x_0))$,

or $v = \frac{\partial g}{\partial t} \circ g^{-1}$,

i.e. $v(t, x) = \boxed{\partial_t g(t, g^{-1}(t, x))}$

Euler eq. translates into an eq. on the trajectory field $t \mapsto g(t, \cdot)$ of $\mathbb{R}_t$ into $G(\Omega)$.

(C^2-diff with $\det(Ds) = 1$)

s.t. $$\frac{d^2}{dt^2} g(t, x_0) = -\nabla p(t, g(t, x_0))$$ ☆

$\partial_t g$

Arnold's interpretation

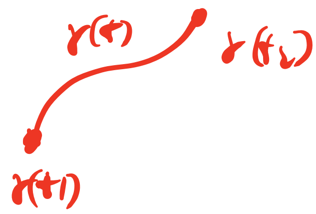
Formal interpretation of the Euler Eq.

The Euler Eq. is the equation of geodesics on $G(\Omega)$, endowed with the Riemannian Structure inherited from the Euclidean space $L^2(\Omega; \mathbb{R}^n)$.

s: diffeomorphism and $\det(Ds) = 1$

Formal Discussion:

a. Geodesic on R.M. M is a path $\gamma(t)$ which minimizes the distance



Action $= \left(\int_{t_1}^{t_2} |\dot g(t)|^2 \, dt \right)^{1/2}$

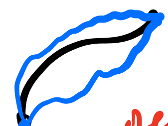
among all curves $g : [t_1, t_2] \to M$ with constraints

$$g(t_1) = \gamma(t_1), \quad g(t_2) = \gamma(t_2)$$

(At least for $|t_1 - t_2| \ll 1$)

(Calculus of Variations)

Geodesic $\Leftrightarrow$ $\boxed{\ddot\gamma(t) \perp T_{\gamma(t)}M}$



perturbe minimizer γ ✓ resh

$$P(g) = \int_{t_1}^{t_2} |\dot g(t)|^2 \, dt$$

$$\frac{d}{d\varepsilon} \int_{t_1}^{t_2} (\dot\gamma + \varepsilon \dot h)^2 \, dt$$

$$= \int_{t_1}^{t_2} 2 \cdot \dot\gamma \, \dot h \, dt$$

$$= -2\int_{t_1}^{t_2} \ddot\gamma \, h \, dt = 0$$

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} P(\gamma + \varepsilon h) = 0$$

$\boxed{h(t_0) = h(t_1) = 0}$ ✓

$\Downarrow$

$\boxed{\ddot\gamma(t) = 0}$

acceleration should vanishes.

Consider the Riemannian structure on $G(\Omega)$ inherited from $L^2(\Omega)$, i.e.

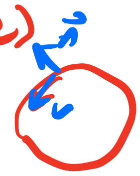
Explanation

acceleration $\dfrac{d^2 g}{dt^2} \perp T_{g(t)} G(\Omega)$ in $L^2(\Omega; \mathbb{R}^n)$

" $-\nabla p(t, g(t, x_0))$ "

Compute the tangent space.

For a path $g(t)$, starting from $g_0 \in G(\Omega)$,

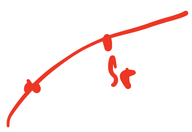


"v" ← velocity field

stays in $G(\Omega)$ iff

$$\begin{cases} \dfrac{\partial g}{\partial t} \cdot \vec{n} = 0 \text{ on } \partial\Omega \\ \nabla \cdot \left[\dfrac{\partial g}{\partial t} \circ g^{-1}\right] = 0 \end{cases}$$

$\nabla \cdot v = 0$



Thus

$$T_g G = \{ \text{Vector field } h \mid \nabla \cdot [\underbrace{h \circ g^{-1}}_{w_0}] = 0 \}$$

$$= \{ h \mid h = w_0 \circ g, \ w_0 \in D_0 \}$$

$w_0 \in D_0$

D_0: all divergence-free vector field.

Since g is measure-preserving,

$$(T_g G)^\perp = \{ \text{vector fields } q_0 \circ g \mid q_0 \in D_0^\perp \}$$

$D_0^\perp$ is the orthogonal subspace to D_0 in L^2

$\langle q_0, w_0 \rangle_{L^2} = 0$.

Helmholtz decomposition:

$$v = w + \nabla p$$

where $\nabla \cdot w = 0$

$\forall\, w \in D_0$

$v = w + \nabla p$

$\langle \nabla p, w \rangle$

$= \int_\Omega \nabla p \cdot w$

$= -\int_\Omega p \, \nabla \cdot w$

$= 0$

Hence $D_0^\perp = \{ -\nabla p \mid p : \Omega \to \mathbb{R} \}$.

So the equation for geodesics reads

$$\frac{d^2}{dt^2} g(t) = -\nabla p(t, g(t)) \in \mathrm{Tan}_{g(t)} G(\Omega).$$

Exactly the incompressible Euler in Langrangian.

$$\left(\int_{t_1}^{t_2} |\dot{g}(t)|^2 \, dt \right)^{1/2} \leftarrow \text{Action of the trajectory } g(t,x).$$

Formal but useful point of view.

More applications:

$A = SO$ Matrix Factorization

半正定对称阵

(Provides a better intuition of the polar factorization)

Brenier's theorem is a natural generalization of a well-known theorem of monotone rearrangement on the line:

Thm 3.18 (Monotone rearrangement theorem)

Let $h : [0,1] \to \mathbb{R}$ be an L^p function $(p \geq 1)$

Then $\exists !$ nondecreasing rearrangement $h^\#$ of h, $(h^\# = \varphi')$. Moreover, $\exists$ a Lebesgue measure preserving map $s : [0,1] \to [0,1]$, s.t. $h = \underset{\varphi'}{\underset{\shortparallel}{h^\#}} \circ s$. (A particular case of Brenier's factorization theorem)

Brenier's theorem also unify several other known facts

(A) The polar factorization of real matrices:

Any matrix $M \in M_n(\mathbb{R})$ can be written as

$M = SO$, where S is symmetric non-negative, and O is an orthogonal (对称，半正定) matrix. ($OO^T = O^TO = I$. 正交)

(i.e. $S \in S_n^+(\mathbb{R})$, $O \in O_n(\mathbb{R})$)

Pf: $M_n(\mathbb{R})$ $A_{n\times n} \rightsquigarrow$ linear mapping.

$M_n : B(0,1) \to \mathbb{R}^n$

$x \mapsto M_n \cdot x$

$\hookrightarrow$ embedded isometrically

W_1

$c(x,y) = d(x,y)$

$L^2(B(0,1), \mathbb{R}^n)$, ($B(0,1)$ unit ball in $\mathbb{R}^n$)

by $M \longmapsto [x \mapsto Mx]$.

Here $M_n(\mathbb{R})$ is endowed with the Hilbert-Schmidt norm $\|\cdot\|_{HS}$, defined by

$$\|M\|_{HS}^2 = tr(M^T M) = \boxed{\sum_{i,j=1}^n m_{ij}^2}$$

with $M = (m_{ij})_{1\le i,j\le n}$.

Then $O_n(\mathbb{R}) \subset S(B(0,1))$ measure preserving

(distance preserving)

while symmetric matrices $S \in S_+(\mathbb{R}^n)$ convex

$= \nabla(\text{quadratic functions})$

$\left(S \in S_n^+(\mathbb{R}),\ x \mapsto \frac{1}{2}\langle x, Sx\rangle = f(x) \right.$

$x \in B(0,1)$

$\left. \nabla f(x) = Sx \right)$

Thm: Let $M \in M_n(\mathbb{R})$.

Then $\exists\, O \in O_n(\mathbb{R})$ and $S \in S_n^+(\mathbb{R})$, s.t.

$$M = SO. \quad (S = \nabla\varphi \circ m)$$

Moreover, the admissable matrices O in the decomposition are the orthogonal projections of M onto $O_n(\mathbb{R})$. (L^2-projection)

i.e. $MO^{-1} \in S_n^+(\mathbb{R})$ ✓

$\Leftrightarrow \left[\forall \tilde{O} \in O_n(\mathbb{R}),\ \|M - O\|_{HS} \le \|M - \tilde{O}\|_{HS} \right].$ ✓

Pf:

Step I: $\|M - O\|_{HS} \le \|M - \tilde{O}\|_{HS}$

$\mathrm{tr}\left((M-O)^T(M-O)\right) \le \mathrm{tr}\left((M-\tilde{O})^T(M-\tilde{O})\right)$

$\|$

$$\Downarrow$$

$$\cancel{M^TM} - M^TO - O^TM - \cancel{O^TO}$$

$$-M^T\tilde{O} - \tilde{O}^TM - \cancel{\tilde{O}^T\tilde{O}}$$

$tr(O^TO) = n.$

$tr(AB)$

$= tr$

$$\Longleftrightarrow \quad \boxed{tr(M^TO) \geq tr(M^T\tilde{O})}$$

$$\text{i.e. } M : O \geq M : \tilde{O}$$

$O^T = O^{-1}$. $O \in O(\mathbb{R}^n)$

Note also

$$tr(M^TO) = tr(O^TM) = tr(MO^T) = tr(\underbrace{MO^{-1}}_{= S})$$

$$tr(M^T\tilde{O}) = tr(\tilde{O}^TM) = tr(O\,\underbrace{\tilde{O}^TM}_{S}\,O^{-1})$$

$$= tr((MO^{-1})(O\tilde{O}^T))$$

$\leftarrow$ any in $O_n(\mathbb{R})$

Since $O\tilde{O}^T$ is an arbitrary element of $O_n(\mathbb{R})$,

it is equivalent to prove, $\forall S \in M_n(\mathbb{R})$,

对称半正定 $\quad \lambda_1 + \cdots + \lambda_n$

$$(P) \quad S \in S_n^+(\mathbb{R}) \Longleftrightarrow [\forall O \in O_n(\mathbb{R}),\ tr\,S \geq tr(SO)].$$

Step 2: Prove P. "$\Rightarrow$" this direction

S: diagonalizable matrix

$$S = O_1 \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} O_1^T = O_1 \Lambda O_1^T$$

all those $\lambda_i \geq 0$

$$tr\,S = \sum_{i=1}^{n} \lambda_i = tr\,\Lambda$$

$$\operatorname{tr}(SO) = \operatorname{tr}(O_1 \Lambda \underline{\underline{O_1^T O}})$$

$$= \operatorname{tr}(\Lambda \underbrace{O_1^T O O_1}_{\in\, O_n(\mathbb{R})}) \le \operatorname{tr}(\Lambda)$$

Step III:

For "$\in$" :

choose $O = I_n + \varepsilon A + O(\varepsilon^2)$, where A is an arbitrary anti-symmetric matrix.

(Check: $\operatorname{Tan}_{I_n} O_n(\mathbb{R}) \cong \underset{\text{space of anti-symmetric matrices.}}{A_n(\mathbb{R})}$

$$(I_n + \varepsilon A)^T (I_n + \varepsilon A) = I_n$$

$$(I_n + \varepsilon A^T)(I_n + \varepsilon A) = I_n$$

$$I_n + \varepsilon \underbrace{(A + A^T)}_{=0} + \cancel{\varepsilon^2} = I_n \;)$$

$\forall O$

By 要记 S 对称矩阵

$$\operatorname{tr}(S) \ge \operatorname{tr}(SO) = \operatorname{tr}(S(I_n + \varepsilon A)) + o(\varepsilon^2)$$

i.e. $0 \ge \varepsilon \operatorname{tr}(SA) + o(\varepsilon^2)$ $\quad f'(\varepsilon) = 0$

Letting $\varepsilon \to 0$, $\operatorname{tr}(S\check{A}) = 0$, holds for any A anti-symmetric.

$f(0) = \max$
$\Downarrow$
$f'(0) = 0$

$$a_{ij} \qquad s_{ij} = \frac{a_{ij} + a_{ji}}{2} = s_{ji}$$

$$a_{ji} \qquad \tilde{a}_{ij} = \frac{a_{ij} - a_{ji}}{2} = -\tilde{a}_{ji}$$

$M_n(\mathbb{R}) = S_n(\mathbb{R}) + A_n(\mathbb{R})$.

$$\operatorname{tr}(A^T A) \geq 0$$
$$= A : A = \sum a_{ij}^2 \geq 0$$

$$\operatorname{tr}(SA) = S^T : A = \left(S_{sy}^T + S_{Anti}^T\right) : A = 0 \quad \forall A.$$

$$\Rightarrow \left(\cancel{S_{sy}^T} + S_{Anti}^T\right) : S_{Anti} = 0$$

$$\sum_{ij} a_{ij} b_{ij}$$

$$S_{sym}^T : S_{Anti} = 0$$

$$= \frac{1}{2}\left\{\sum_{ij} a_{ij} b_{ij} + \sum_{ij} a_{ij} b_{ij}\right\}$$

$$= \frac{1}{2}\left\{\sum_{ij} a_{ij} b_{ij} - \sum_{ij} a_{ij} b_{ij}\right\} = 0$$

$$= S_{Anti}^T : S_{Anti} = 0$$

$$\Rightarrow S_{Anti}^T = 0$$

$\Rightarrow S$ is symmetric.

Further, $\operatorname{tr}(S) \geq \operatorname{tr}(SO) \quad \forall\, O \in O_n(\mathbb{R})$ and S - 对称

$\Rightarrow S \in S_n^+(\mathbb{R})$ non-negative definite.

b) The helmholtz decomposition of vector fields

Any L^2-vector field $\underline{w}$ in a (reasonably smooth) open set $\Omega \subset \mathbb{R}^n$ can be decomposed uniquely as

$$w = v + \nabla p,$$

where $\nabla \cdot v = 0$, tangent to $\partial\Omega$

and p is a real valued function (or distribution) on Ω.

RK: Brenier's theorem: nonlinear version of $\underbrace{\text{Helmholtz's decomposition.}}$

(differential version, of Brenier's theorem.)
or linear ...

Let v be a vector field on Ω,

then one formally consider a path $Z(\cdot)$ in $L^2(\Omega;\mathbb{R}^n)$ of the form

$$Z(\varepsilon) = Id + \varepsilon v + o(\varepsilon).$$

If $|\varepsilon| \ll 1$, and $v \in C_c^\infty$, then $z(\varepsilon)$ satisfies the non-degeneracy condition, so

$$z(\varepsilon) = \nabla\psi(\varepsilon) \circ s(\varepsilon).$$

It is natural to look for

$$\psi(\varepsilon) = \frac{|x|^2}{2} + \varepsilon p + o(\varepsilon), \quad \text{(convex functions)}$$

and

$$s(\varepsilon) = Id + \varepsilon w + o(\varepsilon)$$

$\nabla \cdot w = 0$ & $w \cdot \vec{n} = 0$ on $\partial\Omega$

$$z(\varepsilon) = \underbrace{Id + \varepsilon\nabla p}_{} + \varepsilon w + o(\varepsilon)$$

$$= Id + \varepsilon v + o(\varepsilon)$$

i.e. $\boxed{v = \nabla p + w}$

Calculus

Chapter 4

Caffarelli, Figalli 2018 Bourbaki Review.

An overview of Monge-Ampère Eq.

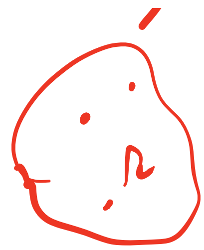
Key questions: The regularity of Optimal Transport Map.

$-\Delta u = f$ Elliptic Equations

Fully nonlinear Eq.

Informal Presentations:

Monge-Ampère Eq.

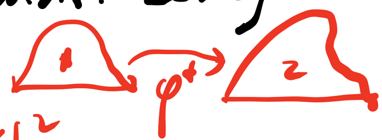


• unbalanced O.T.
$d\mu_{p} d\nu(\rho\nu) \to$

Take $d\mu(x) = f(x)\,dx$ ← Source measure

T ↓ $d\nu(y) = g(y)\,dy$ ↖ target

Two Probability Measures. A.C. w.r.t. Lebesgue.



By Brenier's theorem, $\exists!\ T = \nabla\varphi$ $\quad c(x,y) = \frac{1}{2}|x-y|^2$

$\exists$ ($d\mu$-a.e.) unique $\nabla\varphi$ (gradient of a convex function),
(also Leb.-a.e.)

s.t. $\forall$ test function $\eta \in C_b(\mathbb{R}^n)$, $\quad (\nabla\varphi)_\#\mu = \nu$

$$\int_{\mathbb{R}^n} \eta(y)\, \underbrace{g(y)\, dy} = \int_{\mathbb{R}^n} \eta(\nabla\varphi(x)) f(x)\, dx.$$

$$\int \eta(y)\, d\nu(y)$$

$$\int \eta(y)\, d(\nabla\varphi)_\#\mu(y) = \int \eta(\nabla\varphi(x))\, d\mu(x)$$

Assume that $\nabla\varphi$ is smooth (say, C^2)
and one-to-one ($\Leftarrow$ if φ is strictly convex).

Then $\displaystyle\int_{\mathbb{R}^n} \eta(y)\, g(y)\, dy = \int_{\mathbb{R}^n} \eta(\nabla\varphi(x)) f(x)\, dx$

$\|$ $y = \nabla\varphi(x)$ Change of variables $\quad \| \quad \forall \eta$

$$\int_{\mathbb{R}^n} \eta(\nabla\varphi(x))\, \underbrace{g(\nabla\varphi(x))\, \det(\nabla^2\varphi(x))}\, dx .$$

Hess$(\varphi)(x)$

$\nabla\varphi \in C^2 \Rightarrow \varphi \in C^2$

But a priori $\varphi \in$ Convex.

Since η is arbitrary, one obtains

source, target

$$\text{(OT-MA)} \quad f(x) = g(\nabla\varphi(x)) \det(\nabla^2\varphi(x)).$$

1. Derive a eq. formally
2. Study this PDE rigorously

If g is positive everywhere, then

$$g > 0 \qquad \det(\nabla^2\varphi(x)) = \frac{f(x)}{g(\nabla\varphi(x))}$$

$T = \nabla\varphi$

General Monge-Ampère Eq. reads:

$$\text{(4.5)} \quad \det(\nabla^2\varphi(x)) = F(x, \varphi(x), \nabla\varphi(x)).$$

一般形式

↳ VERY OLD TOPIC

Famous example of (OT-MA) is the equation of prescribed Gauss curvature:

given

$$\det D^2\varphi(x) = \kappa(x)\left(1 + |\nabla\varphi(x)|^2\right)^{\frac{n+2}{2}}, \quad x \in \mathbb{R}^n.$$

Graph(φ)?

Meaning: Graph(φ) has scalar curvature $\underline{K(x)}$

at point $\underline{(x, \varphi(x))}$.

LINEARIZATION

$\det(\nabla^2 \varphi(x)) = 1$ in Ω

M-A Eq. $\xrightarrow[\text{with}]{\text{connected}}$ Laplace-type eq.

?

$\downarrow$ or

linear 2nd order elliptic eq. of the form

$$\sum_{i,j} a_{ij} \frac{\partial^2 \varphi}{\partial x_i \partial x_j} + \sum_i a_i \frac{\partial \varphi}{\partial x_i} + c\varphi = h.$$

scalar

where

$c = c(x) \in \mathbb{R}$.

$(a_{ij}(x))_{1 \le i,j \le n}$: Positive Definite

$(a_i(x))_{1 \le i \le n}$ Vector-valued function

Most particular case.

Laplace eq: $\Delta \varphi = h$

Link with General M-A Eq:

$\det \nabla^2 \varphi$

$$= \prod_{i=1}^{n} \lambda_i, \quad \lambda_i \text{ - eigenvalues of } D^2\varphi.$$

while $\Delta\varphi = \sum_i \lambda_i$

Laplacian = linearized versions of M.-A.

Computations:

Assume that f is strictly positive;

and $\varphi(x) \approx x$, and accordingly $g \approx f$.

Make the ansatz : $\nabla\varphi(x) = x + \varepsilon\nabla\psi + O(\varepsilon^2)$

$$\begin{cases} \varphi(x) = \varphi_\varepsilon(x) = \frac{|x|^2}{2} + \varepsilon\psi + O(\varepsilon^2), \\ g = g_\varepsilon = (1 + \varepsilon h + O(\varepsilon^2)) f, \end{cases}$$

$\Downarrow$ Plugging in : $f(x) = g(\nabla\varphi(x)) \det(\nabla^2\varphi(x))$

$\hookrightarrow$ $\nabla\varphi(x) = x + \varepsilon\nabla\psi + O(\varepsilon^2)$, $\nabla^2\varphi(x) = Id + \varepsilon\nabla^2\psi$.

$g(\nabla\varphi(x)) = f(x+\varepsilon\nabla\psi)(1+\varepsilon h(x))$

$$\frac{1}{1+\varepsilon h} \approx 1 - \varepsilon h$$

$$\frac{f(x)}{f(x+\varepsilon\nabla\psi)}(1-\varepsilon h(x)) = \det(Id+\varepsilon\nabla^2\psi)$$

$$\|$$

$$\frac{f(x)}{f(x)+\varepsilon\cdot\nabla f(x)\cdot\nabla\psi}(1-\varepsilon h(x)) = \det(Id+\varepsilon\nabla^2\psi)$$

$$\frac{1}{1+\varepsilon\,\nabla\log f\cdot\nabla\psi}(1-\varepsilon h(x)) \qquad \varepsilon \qquad \cancel{\varepsilon^2}$$

$$= (1-\varepsilon\,\nabla\log f\cdot\nabla\psi)(1-\varepsilon h(x)) = 1+\varepsilon\,\Delta\psi.$$

$$\boxed{-\nabla\log f\cdot\nabla\psi - h(x) = \Delta\psi} \qquad f\equiv 1_{\text{Ann}}$$

i.e. $\boxed{-\Delta\psi - \nabla\log f\cdot\nabla\psi = h.}$

score $\Downarrow$ $\boxed{-\Delta\psi = h}$

Standard theory of elliptic Eqs

$\Rightarrow$ ψ is smooth given that h and f are

$\psi\in C^{2,}$ $\quad -\Delta\psi = f_\pi$ $\quad f\in C^{0,\alpha}$ and f is >0.

But **NONLINEAR** Problem M-A. Eq.

is much more tricky.

Fully Nonlinear Elliptic Eqs.

Def: Let $G : D \subset \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n \times S_n(\mathbb{R}) \to \mathbb{R}$ be

$(x, \varphi, \nabla\varphi, \nabla^2\varphi)$

a continuous function,

convex w.r.t. the last (matrix) variable.

The eq. $G(x, \varphi, \nabla\varphi, \nabla^2\varphi) = 0$

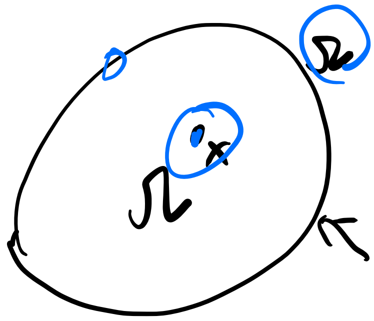
is said to be elliptic if,

for all choices of x, r, p, X, Y,

$$\underbrace{Y \geq X}_{\text{in matrix sense.}} \Rightarrow G(x, r, p, Y) - G(x, r, p, X) \geq 0$$

Uniformly elliptic if $\exists\ \lambda, \Lambda > 0$, s.t.

$$Y \geq X, \Rightarrow \Lambda \operatorname{tr}(Y - X)$$
$$\geq G(x, r, p, Y) - G(x, r, p, X)$$
$$\geq \lambda \operatorname{tr}(Y - X).$$



$\to$ smooth $\partial\Omega \in$?

$\Omega \subseteq \mathbb{R}^n$. $\partial\Omega$ - smooth

simplest boundary condition

Dirichlet $\varphi|_{\partial\Omega} \equiv 0$.

Review paper by A. Figgali.

Function Spaces : $W^{k,p}(\Omega)$, $C^{k,\alpha}(\Omega)$

Sobolev Holder

"Morally": • $\det D^2 u \in W^{k,p} \Rightarrow u \in W^{k+2,p}$

Caffarelli $\{$ • $\det D^2 u \in C^{k,\alpha} \Rightarrow u \in C^{k+2,\alpha}$.

but much intricate than the Laplace eq.

Some difficulties of M. A. eqs.

For simplicity, consider

(Sim- M-A) $\det D^2\varphi = 1$.

Invariant under the action o

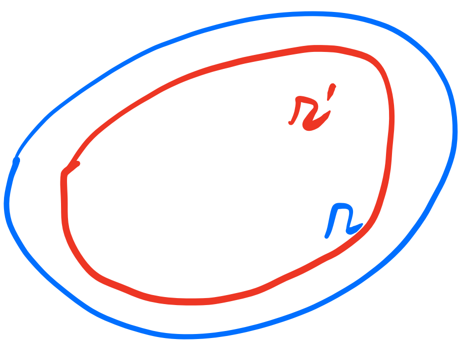
- rotations;
- well-chosen dilations;
- any affine transformation with unit det.

e.g. if $\det(D^2\varphi) \equiv 1$ on $\mathbb{R}^2$,

then $\tilde{\varphi} = \varphi(\varepsilon x, \frac{y}{\varepsilon})$ also solves it.

- For Eq. $\det(\nabla^2\varphi) = 1$,

NO interior a priori estimates like those for Laplace eqs.



$\Omega' \subset\subset \Omega$.

open.

$\overline{\Omega'} \subset \Omega$.

For Laplace eq. $\Delta u = 0$ in $\Omega \subset \mathbb{R}^n$

one has: $\|u\|_{C^k(\Omega')} \le C_k \|u\|_{L^1(\Omega)}$. $\forall k$.

$C_k \sim k, n, d(\partial\Omega, \partial\Omega')$

See Gilbarg–Trudinger:

Elliptic PDEs of 2nd Order.

- Non-uniform convexity in $n \geq 3$.

Pogorelov's example: non-uniformly convex function.

$$\varphi(x) = (1+x_n^2)\left(\sum_{k=1}^{n-1} x_k^2\right)^{1-1/n}$$

Satisfies: $\det(D^2\varphi) \in C^\infty$ $n \geq 3$.

but $\varphi \in C^{1,1-\frac{2}{n}}$

Caffarelli's example: $\det(D^2\varphi)$ even analytic

$$\varphi(x) = \left(\sum_{k=1}^{n-1} x_k^2\right)^{1/2} + (1+x_n^2)\left(\sum_{k=1}^{n-1} x_k^2\right)^{\frac{\varphi}{n}}$$

s.t., $\det D^2\varphi$ is positive Lipschitz if $n=3$,

, — — — is — — analytic if $n=4$,

but $\boxed{\varphi \text{ is not better than Lipschitz !}}$

$\det(D^2\varphi) = f_k$ $f \in W^{k,p} \Rightarrow \varphi \in W^{k+2,p}$

Various Notions of weak solutions

To derive (44) $\det(D^2\varphi(x)) = \dfrac{f(x)}{g(\nabla\varphi(x))}$,

we assumed that $\nabla\varphi \in C^1$ or $\varphi \in C^2$.

$\nabla\varphi$: 1-1

But we don't know whether this is true.

Since φ is convex, a priori φ is C^0

and $W^{1,\infty}_{loc}$ on $\mathrm{int}(\mathrm{Dom}(\varphi))$, but not necessarily

C^2.

Study: $\det(D^2\varphi) = 2$ without a priori $\varphi \in C^2$.

$D^2\varphi$ 未必存在

How to define $\det(D^2\varphi)$ if φ is not C^2

$\Rightarrow |\partial\varphi(x)|$

How to make sense of general

$$\det(D^2\varphi(x)) = F(x, \varphi(x), \nabla\varphi(x)).$$

THERE WAYS

$\det(D^2\varphi(x)) \equiv 2$

1) Aleksandrov solutions:

$\underbrace{\det_H D^2\varphi}$: Hessian measure associated to φ.

$\Downarrow$ defined as

Why does this definition make sense?

A Borel measure st: $\forall$ measurable set $E \subset \mathbb{R}^n$,

$\{0\}$ = 1.

$$\star \quad (\det_H D^2\varphi)[E] = |\partial\varphi(E)|,$$

$\varphi = |x|$ in $\mathbb{R}^1$.

where $\quad \partial\varphi(E) = \bigcup_{x\in E} \partial\varphi(x).$

$$(\det_H D^2\varphi)\Big|_0 = \delta_0$$

We call that φ is an Aleksandrov solution of (General M-A) if the Hessian measure $\det_H D^2\varphi$ is A.C. w.r.t. Lebesgue and its density $= F(x, \varphi(x), \nabla\varphi(x))$, defined a.e.

(Or the measure $\det_H D^2\varphi$ has no singular part, and (4.5) holds a.e. with

$$\det D^2\varphi = \det D_A^2\varphi.$$

(Aleksandrov 2nd derivative)

ii) Viscosity Solutions:

Theory for Hamilton-Jacobi. (Evans) Eq.

φ is a viscosity solution of

$\det D^2\varphi(x) = F(x, \varphi(x), \nabla\varphi(x))$, $x \in \Omega$ open.

$(\varphi-\psi)|_{x_0} \geq (\varphi-\psi)|_{B(x_0,\delta)}$

if

1) whenever ψ is convex C^2 test function s.t. $\varphi - \psi$ has a strict local maximum at x_0, then

$\nabla^2(\varphi-\psi)|_{x_0} \leq 0$ (matrix) $\Rightarrow \det \nabla^2\psi|_{x_0} \geq \det \nabla^2\varphi|_{x_0} = F(x_0, \varphi(x_0), \nabla\varphi(x_0))$

$$\det D^2\psi(x_0) \geq F(x_0, \varphi(x_0), \nabla\psi(x_0)).$$

2) whenever ψ is convex C^2 test function s.t. $\varphi - \psi$ has a strict local minimum at x_0, then

$$\det D^2\psi(x_0) \leq F(x_0, \varphi(x_0), \nabla\psi(x_0)).$$

well-defined. since $\psi \in C^2$.

iii) Brenier solutions

Only for eq. $\det(D^2\varphi(x)) = \dfrac{f(x)}{g(\nabla\varphi(x))}$ (OT-M.A.)

$(\nabla\varphi)_\# \mu = \nu \Rightarrow$ then it is Brenier's solution

φ is a Brenier solution to (OT.-M.A.) if

$(\nabla\varphi)_{\#}\mu = \nu$ given $\begin{cases} d\mu(x) = f(x)\,dx \\ d\nu(y) = g(y)\,dy. \end{cases}$

Next Class]

Chapter 5. $\boxed{\text{Displacement convexity}}$

Displacement interpolation / convexity

§5.1 Displacement interpolation

§5.1.1. Time-dependent Monge-Kantorovich problem

Previously, $c = c(x, y)$: function of the initial and the final locations

cost

NOT DEPEND ON THE PATH

Benamou and Brenier : $c(x,y) = \frac{1}{2}|x-y|^2$

mass transportation : distance problem $W_2(\mu, \nu)$

compare

time-dependent minimization : geodesic problem

(OPTIMAL PATH between μ and ν)

Monge's Formulation :

(1) $$\inf \left\{ \int_X c(x, T(x))\, d\mu(x) \;;\; T_\# \mu = \nu \right\}$$

$\forall x \longmapsto (T_t(x))_{0\le t\le 1}$ (or $(T_t x)$).

$C[(T_t x)]$: displacement cost

Require: the path $t \mapsto T_t x$ C^0 and piecewise C^1

for $d\mu$-a.e. x.

Solve the time dependent minimization problem

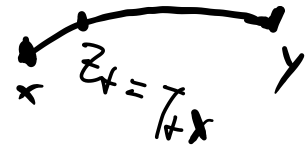
(2) $$\inf\left\{ \int_X \underbrace{C[(T_t x)_{0\le t\le 1}]}_{\Uparrow} \, d\mu(x) : T_0 = Id, \; T_{1\#}\mu = \nu \right\}$$

How to define this?

Problem (1) & (2) compatible if they predict $\boxed{\text{the same total cost}}$ and the same displacement map : i.e. each optimal (T_t) in (2) gives rise to an optimal T in (1), via $T = T_1$.

Indeed, we need

$$\forall x, y, \quad c(x,y) = \inf\left\{ C[(z_t)_{0\le t\le 1}] : z_0 = x, \; z_1 = y \right\}.$$



If the underlying space is a RM, and (or with a differentiable structure)

$$\dot{z}_t = \frac{dz_t}{dt}$$

then usually,

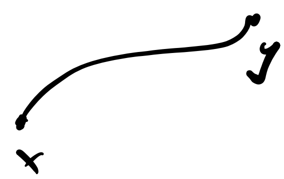
$$C[(z_t)] = \int_0^1 c(\dot{z}_t)\, dt$$

$\boxed{c(z)}$ ⤳ differential cost

<u>Example:</u>

i) $C[(z_t)] = \int_0^1 |\dot{z}_t|^2 \, dt$ in $\mathbb{R}^n$

$$c(x,y) = \inf \left\{ C[(z_t)_{0 \le t \le 1}] \;\middle|\; \begin{matrix} z_0 = x \\ z_1 = y \end{matrix} \right\}$$



$z_t^\varepsilon = z_t + \varepsilon h_t$ with $h_0 = x$, $h_1 = y$

Let $\psi(\varepsilon) = \int_0^1 |\dot{z}_t + \varepsilon \dot{h}_t|^2 \, dt$

$$= \int_0^1 |\dot{z}_t|^2 + 2\varepsilon \int_0^1 \dot{z}_t \dot{h}_t \, dt + \varepsilon^2 \int_0^1 |\dot{h}_t|^2 \, dt$$

$$\psi'(\varepsilon)\Big|_{\varepsilon=0} = 2 \underbrace{\int_0^1 \dot{z}_t \dot{h}_t \, dt}_{} = 0 \quad \left(\text{since } \psi(0) \text{ minimal}\right)$$

$$= -2\int^1 h(t)\, z''(t)\, dt = 0$$

$\forall h$

$$\Rightarrow \quad z''(t) \equiv 0 \quad \forall\, t \in [0,1].$$

$$z_0 = x \qquad z_1 = y \qquad z(t) = (1-t)x + ty$$

A line segment

$$\dot{z}(t) = y - x$$

$$|\dot{z}(t)|^2 = |y-x|^2$$

i.e. $\boxed{c(x,y) = |y-x|^2}$

ii) $C[(z_t)] = \int_0^1 |\dot{z}_t|^p \, dt, \quad p \geq 1 \quad$ in $\mathbb{R}^n$

$\hookrightarrow \quad c(x,y) = |x-y|^p.$

(Check: $\displaystyle\int_0^1 |\dot{z}_t|^p \, dt \geq |x-y|^p$

Simply be Jensen:

$$\left| \int_0^1 \dot{z}_t \, dt \right|^p \leq \int_0^1 |\dot{z}_t|^p \, dt$$

$$\varphi\left(\frac{x_1+x_2}{2}\right) \leq \frac{1}{2}\left(\varphi(x_1) + \varphi(x_2)\right)$$

$$\varphi(\mathbb{E}X) \leq \mathbb{E}\,\varphi(X)$$

<u>Example 3:</u>

$$C[(z_t)] = \int_0^1 \|\dot{z}_t\|^p \, dt$$ on a smooth complete Riemannian manifold M.

$$\Downarrow$$

$$c(x,y) = d(x,y)^p$$

(Again $\displaystyle\int_0^1 \|\dot{z}_t\|^p \, dt \geq \left\| \int_0^1 \|\dot{z}_t\| \, dt \right\|^p = d(x,y)^p.$)

More generally,

PROP 5.2 (Extremal trajectories for convex costs are straight lines).

If c is convex function on $\mathbb{R}^n$, then

$$\inf\left\{\int_0^1 c(\dot{z}_t)\,dt : z_0 = x,\ z_1 = y\right\}$$

$$= c(y-x)$$

Up to change of time horizon:
for $T>0$,

$$\inf\left\{\int_0^T c(\dot{z}_t)\,dt \mid z_0 = x,\ z_T = y\right\}$$

$[0,T] \ni t = T\tilde{t}$

$\Rightarrow \tilde{t} = \frac{t}{T} \in [0,1]$

$$= \int_0^1 T c\left(\frac{1}{T}\frac{dz}{d\tilde{t}}\right) d\tilde{t}$$

$$\frac{dz_t}{dt} = \frac{dz_t}{d\tilde{t}} \cdot \frac{d\tilde{t}}{dt}\ \ 1/T$$

$c(z) \longmapsto \boxed{T c\left(\frac{1}{T}\cdot\right)}$ $\leftarrow$ This is also convex of convex)

Moreover, if c is strictly convex,

then the infimum is achieved uniquely

by $z_t = (1-t)x + ty = x + t(y-x)$

(For variant, $x + \frac{t}{T}(y-x)$)

Pf is also by Jensen's inequality.

Important Remarks:

(i) If $c(z) = |z|^p$ $(p \geq 1)$ on $\mathbb{R}^n$,

then
- $\inf\left\{\int_0^1 c(\dot{z}_t)\,dt \,\middle|\, z_0 = x,\ z_1 = y\right\} = c(y-x)$

while
- $\inf\left\{\int_0^T c(\dot{z}_t)\,dt \,\middle|\, z_0 = x,\ z_T = y\right\} = T\,c\left(\frac{y-x}{T}\right)$

$$T\,c\left(\frac{z}{T}\right) = T\cdot\left(\frac{|z|}{T}\right)^p = \underline{\frac{1}{|T|^{p-1}}}\,|z|^p$$

$\Downarrow$

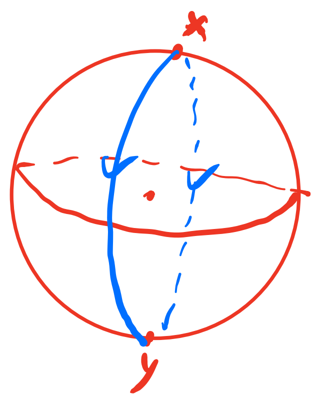
The infima are the same up to a multiplicative factor which depends only on T.

(ii) PROP 5.2 for strictly convex differential cost c, the only optimal trajectories are straight lines, parametrized with constant velocity.

$$\left(z_t = x + t(y-x) \qquad \dot{z}_t = y - x \leftarrow \boxed{\text{constant}} \right)$$

Also, the only optimal trajectories for the differential cost $c(\dot{z}) = \|\dot{z}\|^p$ on a manifold $(p>1)$ are the minimizing geodesics with arc length parametrization.

(May be non-unique)



- Strictly convex case:

$$\int_0^1 \|\dot{z}_t\|^p \, dt \geq \left| \int_0^1 \|\dot{z}_t\| dt \right|^p = d(x,y)^p$$

"=" $\Rightarrow$ $\|\dot{z}_t\| = $ constant

- $p=1$

$$\int_0^1 \|\dot{z}_t\| \, dt \overset{t = h(\tau)}{=\!=} \int_0^1 \underbrace{\|\dot{z}_t\| \, h'(\tau)}_{\int_0^1 \|\dot{z}_\tau\| d\tau} d\tau$$

$$\frac{dz}{dt} = \frac{dz}{d\tau}\frac{d\tau}{dt} \Rightarrow \frac{dz}{d\tau} = \dot{z}\, h'$$

$p=1$ is Degenerate in the sense of time-reparametrization.

If we require that

$$c(x,y) = \inf\left\{ C[(z_t)_{0\le t\le 1}] : \begin{array}{l} z_0 = x \\ z_1 = y \end{array} \right\}$$

Then for $d\mu$-a.e. x, $(T_t x)_{0\le t\le 1}$ is optimal,

i.e. $$c(x, T(x)) = C\left[(T_t x)_{0\le t\le 1} \right]$$

i.e. up to a negligible set of initial locations, each trajectory should be OPTIMAL.

Examples: • If $c(x,y) = c(x-y)$ in $\mathbb{R}^n$, with c strictly convex, $c(0)=0$ then a.e. trajectories are straight lines.

• If $c(x,y) = d(x,y)^p$ $(p\ge 1)$ on M, then a.e. trajectories have to be minimizing geodesics.

Time-independent minimization problem

$\Updownarrow$

Time-dependent --- .

Thm 5.5 (Time-dependent optimal transportation theorem).

Consider the cost function $c(x,y) = c(x-y)$ in $\mathbb{R}^n$, with c strictly convex, $c(0) = 0$.

Let $\mu, \nu \in P_{AC}(\mathbb{R}^n)$ and $C[(z_t)] = \int_0^1 c(\dot{z}_t)\, dt$.

Let $\nabla\psi$ be the ($d\mu$-a.e.) unique gradient of a c-concave function ψ s.t.

$$[Id - \nabla c^*(\nabla\psi)]\#\mu = \nu.$$

Then the solution of the time-dependent minimization problem

$$\inf\left\{ \int_X C[(T_t x)_{0\le t\le 1}]\, d\mu(x) : \begin{array}{l} T_0 = Id \\ (T_1)_\#\mu = \nu \end{array} \right\}.$$

is given by

$$T_t(x) = x - t\nabla c^*(\nabla \psi(x)),$$

$$0 \le t \le 1.$$

SOME FACTS

See Page 93: Connecting c-superdifferential and differential.

Let $c(x,y) = c(x-y)$, where c is strictly convex and ∇c invertiable (then $(\nabla c)^{-1} = \nabla c^*$, c^* is the usual Legendre transform of c).

Let φ be a c-concave function, and φ is differentiable at x, then

$$\partial^c \varphi(x) = \{x - \nabla c^*(\nabla \varphi(x))\}.$$

Prove this!

For c is convex, $y \in \partial \varphi(x)$, one has

$$c(x) + c^*(y) = x \cdot y$$

and $c(z) + c^*(y) \geq z \cdot y \quad \forall z$.

Hence the function

$$h(z) = c(z) + c^*(y) - z \cdot y$$

satisfies $h(z) \geq h(x) = 0$

Taking the gradient gives us that

$$\nabla_x c(x) - y = 0$$

Similarly: $\nabla_y c^*(y) - x = 0$

Hence $\nabla c(x) = y$ and $\nabla c^*(y) = x$

$$\boxed{\nabla c^* \circ \nabla c = Id.}$$

For Kantorovich potential (φ, φ^c), one can assume that $\varphi \in W^{1,\infty}_{loc}$ and hence φ is differentiable a.e. w.r.t. Lebesgue.

Now for $(x, y) \in supp(\gamma)$, $\gamma \in OPT(\mu, \nu)$, one has

$$\varphi(x) + \varphi^c(y) = c(x,y) \quad \forall (x,y) \in \mathrm{supp}(\gamma)$$

and $$\varphi(z) + \varphi^c(y) \le c(x,z) \quad \forall z.$$

Hence $h_y(z) = \varphi(z) + \varphi^c(y) - c(x,z)$

reaches an extremal at $z = x$

i.e. $\nabla h_y(x) = 0.$

i.e. $\nabla \varphi(x) - \nabla c(x-y) = 0$

$$x - y = (\nabla c)^{-1} \nabla \varphi(x)$$

$$y = x - (\nabla c)^{-1} \nabla \varphi(x)$$

i.e. $$\boxed{y = T(x) = x - \nabla c^*(\nabla \varphi(x))}$$

Proof of Thm I.5

Let $T(x) = x - \nabla c^*(\nabla \varphi(x))$ s.t.

$$T_\# \mu = \nu.$$

and T is the optimal map

$T_0 = Id$

$T_1 = T.$

$$T_t(x) = (1-t)\,T_0(x) + t\,T_1(x)$$
$$= (1-t)x + t\left(x - \nabla c^*(\nabla\psi(x))\right)$$
$$= x - t\,\nabla c^*(\nabla\psi(x))$$

②

Transportation on R.M. M with cost $(d(x,y))^2/2$.. The the optimal transport map

$$T(x) = \exp_x(-\nabla\psi(x))$$

(See theorem 2.4.7 : McCann's theorem)

Expect: the solution to the time-dependent minimization problem is given by the geodesic path:

$$T_t(x) = \exp_x(-t\,\nabla\psi(x)).$$

(Need this geodesic is minimizing)

A cost function is homogeneous if it is of the form $c(x,y) = |x-y|^p$ in $\mathbb{R}^n$

or $c(x,y) = d(x,y)^p$

(on a smooth complete manifold)

Only consider $p \geq 1$.

Thm 5.6

Monge- Kantorovich problem

i) μ, ν do not give mass to small sets,

$c(x-y) = |x-y|^p$ in $\mathbb{R}^n$ $(p>1)$

and the optimal map is of the form

$$T(x) = x - Dc^*(D\psi(x))$$

ii) μ, ν — A.C. and compactly supported in a smooth, complete RM. M, $c(x,y) = d^2(x,y)/2$.

and the optimal map takes the form

$$T(x) = \exp_x(-\nabla\varphi(x)).$$

And $\forall t \in [0,1]$,

$$\text{define } T_t(x) = \begin{cases} x - t\,\nabla c^*(\nabla\varphi(x)) & \text{in } \mathbb{R}^n \text{ case} \\ \exp_x(-t\nabla\varphi(x)) & \text{in RM. case.} \end{cases}$$

Proof: Case ii)

Recall $d^2/2$ - concave functions are functions of the form

$$\varphi(x) = \inf_{y \in M}\left[\tfrac{1}{2}d(y,x)^2 + \eta(y)\right],$$

$$\eta : M \to \mathbb{R} \cup \{-\infty\}.$$

What we need to show is that

$t\varphi$ is also $d^2/2$-concave when $0 \le t \le 1$.

We treat the particular case $\varphi(x) = \dfrac{d^2(z,x)}{2}$

To show: $\forall \lambda \in [0,1]$, one can write

$$\lambda \frac{d(z,x)^2}{2} = \inf_{y\in M} \left[\frac{d(y,x)^2}{2} + \eta(y) \right]$$

This is just a particular case of a well-known identity

$$\inf_y \left[\frac{d(x,y)^2}{a} + \frac{d(y,z)^2}{b} \right] = \frac{d(x,z)^2}{a+b},$$

$a, b > 0$

Hence choose a, b, st. $\lambda = \frac{a}{a+b}$,

$$\lambda d(x,z)^2 = \frac{a}{a+b} d(x,z)^2 = \inf_y \left[\underline{d(x,y)^2} + \underbrace{\frac{a}{b} d(y,z)^2}_{\eta(y).} \right]$$

okay.

$$\lambda \psi(x) = \inf_{y\in M} \left[\frac{\lambda d(y,x)^2}{2} + \lambda \eta(y) \right] \overset{\text{inf.}}{\Rightarrow} \frac{d^2}{2} \text{ - concave}$$

$\frac{d^2}{2}$ - concave. w.r.t. x.

$$\inf_y \left[\inf_z \left[\frac{1}{2} d^2(x,z) + \tilde{c}\, d(z,y)^2 \right] + \lambda \eta(y) \right]$$

$$= \inf_z \left[\frac{1}{2} d^2(x,z) + \inf_y \tilde{\eta}_z(y) \right]$$

Exercise: Prove formula (5.10), i.e.

$$\inf_y \left[\frac{d(x,y)^2}{\frac{a}{a+b}} + \frac{d(y,z)^2}{\frac{b}{a+b}} \right] = \frac{d(x,z)^2}{\cancel{a+b}}$$

λ

$y = (1-\lambda)x + \lambda z.$

$\lambda = \frac{a}{a+b}.$

($\Leftarrow$

$$\phi(y) \triangleq \frac{|x-y|^2}{a} + \frac{|y-z|^2}{b}$$

$$\nabla\phi(y) = \frac{2}{a}(y-x) + \frac{2}{b}(y-z) = 0$$

$$\Rightarrow \left(\frac{1}{a}+\frac{1}{b}\right)y = \frac{1}{a}x + \frac{1}{b}z.$$

$$(a+b)\,y = bx + az$$

$$\Rightarrow y = \frac{az+bx}{a+b}$$

$\phi(y_*) =$

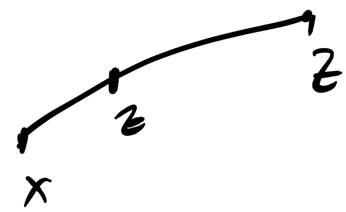
$$\frac{1}{a}\left|x - \frac{b}{a+b}x - \frac{a}{a+b}z\right|^2 + \frac{1}{b}\left|z - \frac{a}{a+b}z - \frac{b}{a+b}x\right|^2$$

$$= \frac{1}{a}\left(\frac{a}{a+b}\right)^2|x-z|^2 + \frac{1}{b}\cdot\left(\frac{b}{a+b}\right)^2|x-z|^2.$$

$$= \frac{1}{a+b}|x-z|^2.$$)

More generally:

$$\inf_y \left[\frac{d(x,y)^2}{a} + \frac{d(y,z)^2}{b} \right] = \frac{d(x,z)^2}{a+b}$$



$$\frac{1}{\frac{a}{a+b}} d(x,y)^2 + \frac{1}{\frac{b}{a+b}} d(y,z)^2$$

$$\inf_y \left[\frac{d(x,y)^2}{\lambda} + \frac{d(y,z)^2}{1-\lambda} \right] = d(x,z)^2.$$

if $\quad d(x,y) = \lambda\, d(x,z).$

$\quad\quad d(y,z) = (1-\lambda)\, d(y,z)$ $\quad$ interpolation in geodesics.

then $\quad \dfrac{d(x,y)^2}{\lambda} + \dfrac{d(y,z)^2}{1-\lambda} = d^2(x,z)$

$\forall y$

$$\frac{d(x,y)^2}{\lambda} + \frac{d(y,z)^2}{1-\lambda} \geq d^2(x,z)$$

$$d(x,z)^2 \leq \left(d(x,y) + d(y,z)\right)^2 \leq \frac{d(x,y)^2}{\lambda} + \frac{d(y,z)^2}{1-\lambda}.$$

$$= d^2(x,y) + d^2(y,z) + 2\, d(x,y)\, d(y,z) \leq \frac{1}{\lambda} d(x,y)^2 + \frac{1}{1-\lambda} d(y,z)^2$$

$$2\, d(x,y)\, d(y,z) \le \left(\frac{1}{\lambda}-1\right) d(x,y)^2 + \left(\frac{1}{1-\lambda}-1\right) d(y,z)^2$$

$$= \frac{1-\lambda}{\lambda} d(x,y)^2 + \frac{\lambda}{1-\lambda} d(y,z)^2 .$$

$$= \text{okay}$$

$\lambda \in (0,1)$

$$2ab \le \varepsilon a^2 + \frac{b^2}{\varepsilon}$$

(5.10) $$\inf_y \left[\frac{d(x,y)^2}{a} + \frac{d(y,z)^2}{b} \right] = \frac{d(x,z)^2}{a+b}$$

$\Uparrow$

Hopf – Lax formula

Link to:

The invariance property of Hamilton-Jacobi:

if u solves

$$\partial_t u + c^*(\nabla u) = 0$$

where $c(z) = |z|^p$,

then $\lambda u(\lambda^{p'-1} t, \cdot)$ also solves the same eq.

$\frac{1}{p'} + \frac{1}{p} = 1$

5.1.3. McCann's interpolation

Particular important case: $c(x,y) = |x-y|^2$ in $\mathbb{R}^n$

Solution of the time-dependent minimization problem coincides with McCann's interpolation

or Displacement interpolation.

Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$, and μ & ν do not charge small sets.

Then by Brenier's theorem,

$\exists$ ($d\mu$ - a.e. unique) gradient of a convex function $\nabla\varphi$, st $\nabla\varphi_{\#}\mu = \nu$.

Define:

$$\rho_t = [\mu, \nu]_t = \left[(1-t)\, Id + t\nabla\varphi\right]_{\#}\mu.$$

$$[\mu,\nu]_0 = \rho_0 = \mu, \quad \rho_1 = \nu = [\mu,\nu]_1$$

The family of probability measures $(\rho_t)_{0\le t\le 1}$ interpolates between μ and ν. in a remarkable way.

$$(1-t)\,Id + t\nabla\varphi = \nabla\Big(\underbrace{(1-t)\frac{|x|^2}{2} + t\varphi}_{\text{also convex.}}\Big)$$

By Brenier's theorem

$$T_t = (1-t)\,Id + t\nabla\varphi, \quad (T_t)_\# \mu = \rho_t$$

$\uparrow$ is the optimal map transporting μ to ρ_t.

And

$$W_2^2(\mu, \rho_t) = \int_{\mathbb{R}^n} \left| x - [(1-t)x + t\nabla\varphi(x)] \right|^2 d\mu(x)$$

$$= t^2 \int_{\mathbb{R}^n} |x - \nabla\varphi(x)|^2 d\mu(x)$$

$$= t^2 W_2^2(\mu, \nu) = t^2 W_2^2(\rho_0, \rho_1)$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad \shortparallel$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad W_2^2(\rho_0, \rho_t)$$

Or:

$$W_2(\mu, \rho_t) = t\,W_2(\mu, \nu)$$

$\hookrightarrow$ Quadratic Wasserstein Distance.

Other Properties of Displacement interpolations

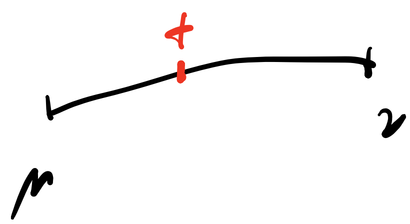
PROP. 5.9. One also has

i) $[\mu,\nu]_t = [\nu,\mu]_{1-t}$;

ii) $[[\mu,\nu]_t, [\mu,\nu]_{t'}]_s = [\mu,\nu]_{(1-s)t+st'}$;

iii) if μ or $\nu \in \mathcal{P}_{ac}(\mathbb{R}^n)$, then so is $[\mu,\nu]_t$, $\forall\, t \in (0,1)$.

Pf:



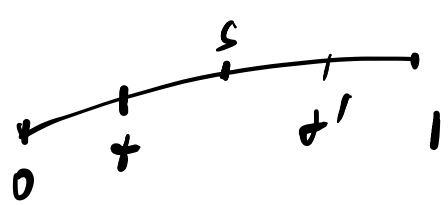
i) $$[\mu,\nu]_t = \big((1-t)\,Id + t\nabla\varphi\big)_\# \mu$$

$$= \big((1-t)\,Id + t\nabla\varphi\big)_\# (\nabla\varphi^*{}_\# \nu)$$

$$= \big[\big((1-t)\,Id + t\nabla\varphi\big)\circ \nabla\varphi^*\big]_\# \nu$$

$$= \big((1-t)\nabla\varphi^* + t\,Id\big)_\# \nu$$

ii



$$T_t\, x = (1-t)\,x + t\,\nabla\varphi(x)$$

$$T_{t'}\, x = (1-t')\,x + t'\,\nabla\varphi(x)$$

$$(1-s)\,T_t + s\,T_{t'} = \big(1-[(1-s)t+st']\big)\,Id + [(1-s)t+st']\,\nabla\varphi$$

iii) Let us consider the case that $\mu\in\mathcal{P}_{AC}(\mathbb{R}^d)$.

Let $(1-t)x + t\nabla\varphi(x) = \nabla\Big(\underbrace{t\,\varphi(x) + (1-t)\frac{|x|^2}{2}}_{\varphi_t(x)}\Big)$

$$\langle \nabla\varphi_t(x) - \nabla\varphi_t(y),\ x-y\rangle$$

$$= (1-t)\,|x-y|^2 + \underbrace{t\,\langle \nabla\varphi(x) - \nabla\varphi(y),\ x-y\rangle}_{\geq 0}$$

$$\geq (1-t)\,|x-y|^2$$

By Cauchy-Schwarz,

$$(*)\quad |\nabla\varphi_t(x) - \nabla\varphi_t(y)| \geq (1-t)\,|x-y|$$

Since φ_t is uniformly convex,

φ_t^* is everywhere differentiable,

and $(*) \Rightarrow \nabla\varphi_t^* = (\nabla\varphi_t)^{-1}$

is Lipschitz with Lipschitz norm $\le \frac{1}{L_t}$.

In particular, when $|A| = 0$, then

$$|\nabla\varphi_t^*(A)| = 0.$$

Then

$$\begin{aligned}\rho_t[A] &= (\nabla\varphi_t)_{\#}\mu[A] \\ &= \mu[(\nabla\varphi_t)^{-1}(A)] \\ &= \mu[\nabla\varphi_t^*(A)] = 0.\end{aligned}$$

§ 5.2 Displacement Convexity

Two types interpolations:

McCann or Displacement interpolation $\begin{cases} \rho_t = (T_t)_\# \mu & T_t = (1-t)Id + t\nabla\varphi(x) \\ T_\# \mu = \nu. & T - \text{optimal} \end{cases}$

(interpolation for maps)

Trivial "Linear" interpolation:

$$\rho_t = (1-t)\mu + t\nu.$$

(interpolation in probability space)

$\rho \in \mathcal{P}(\mathbb{R}^n)$

if $\rho \ll$ Leb, then we shall identify it with its Lebesgue density, and write

$$d\rho(x) = \rho(x)dx.$$

Assume $\mu, \nu \in \mathcal{P}_{ac}(\mathbb{R}^n)$ and

$$\{ \quad \rho_t = [\mu, \nu]_t = ((1-t)Id + t\nabla\varphi)_\# \mu, \quad 0 \le t \le 1$$

$(\nabla\varphi)_\# \mu = \nu \quad \varphi$ - convex

For Quadratic cost

More generally, for $c = c(x,y) = c(x-y)$

c : strictly convex.

$T_\# \mu = \nu$, T optimal

$$T = Id - \nabla c^*(\nabla\psi(x))$$

$$T_t = (1-t) Id + t(Id - \nabla c^*(\nabla\psi(x)))$$

$$= Id - t\nabla c^*(\nabla\psi),$$

where ψ is a c-concave Kantorovich potential.

Def (Displacement convexity)

(i) A subset $\mathcal{P} \subset P_{ac}(\mathbb{R}^n)$ is said to be displacement convex if it is stable under displacement interpolation: $\forall \mu, \nu \in \mathcal{P}$, $\forall t \in [0,1]$,

$$\rho_t = [\mu,\nu]_t \in \mathcal{P}.$$

ii) Let F be a functional defined on a displacement convex subset $\mathcal{P} \subset \mathcal{P}_{ac}(\mathbb{R}^n)$, with values in $\mathbb{R} \cup \{+\infty\}$. It is said to be **displacement convex on $\mathcal{P}$** if:

Given $\rho_0 = \mu$, $\rho_1 = \nu \in \mathcal{P}$ and $(\rho_t)_{0 \le t \le 1}$ is their displacement interpolant, then $t \mapsto F(\rho_t)$ is convex on $[0,1]$.

Note: classical def. of convexity is the same, except the displacement interpolation is replaced by linear interpolation $\tilde{\rho}_t = (1-t)\mu + t\nu$.

Extension to more general prob. measures.

$\mathcal{P}_2(\mathbb{R}^n)$: TWO MOMENT FINITE

Def 5.12 (General version of Displacement Convexity)

Let $\sigma_t : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ be.

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

(i) A subset $\mathcal{P} \subset \mathcal{P}_2(\mathbb{R}^n)$ is displacement convex, if for all $\rho_0, \rho_1 \in \mathcal{P}$, for all π optimal in the Monge-Kantorovich problem with marginals ρ_0, ρ_1 and $c(x,y) = \frac{1}{2}|x-y|^2$, and $\forall t \in [0,1]$,

$$\rho_t := (\sigma_t)_\# \pi \in \mathcal{P}.$$

(ii) Then a functional F on $\mathcal{P}$ is said to be Displac. Convex, if $\forall \rho_0, \rho_1 \in \mathcal{P}$,

$$t \longmapsto F(\rho_t) \quad \text{is convex in } [0,1].$$

(iii) For any strictly convex cost function $c(x-y)$ one can define the concept of c-displacement convexity in a similiar way, provided that

$$T_c(\mu, \nu) < +\infty, \quad \forall \mu, \nu \in \mathcal{P}.$$

choosing $\pi \in OPT(\mu, \nu)$.

and define $\rho_t = \sigma_t \# \pi$

(Not very useful for general c-displace convex.)

Variants of Defn 5.12.

F is strictly displacement convex on $\mathcal{P}$ if

$\forall \rho_0, \rho_1 \in \mathcal{P}$, $\rho_0 \neq \rho_1$

$\Rightarrow [\, t \mapsto F(\rho_t)$ is strictly convex on $[0,1]\,]$

The functional F is said to be

λ-uniformly displacement convex on $\mathcal{P}$ for some $\lambda > 0$ if for all $\rho_0, \rho_1 \in \mathcal{P}$,

$$\frac{d^2}{dt^2} F(\rho_t) \geq \lambda W_2^2(\rho_0, \rho_1)$$

It is said to be semi-displacement-convex on $\mathcal{P}$, with $C \geq 0$, if $\forall \rho_0, \rho_1 \in \mathcal{P}$,

$$\frac{d^2}{dt^2} F(\rho_t) \geq -C W_2^2(\rho_0, \rho_1),$$
$$\forall \ t \in (0,1)$$

Recall

$$W_2^2(\rho_0, \rho_1) = \inf_{\pi \in \Pi(\rho_0, \rho_1)} \int_{\mathbb{R}^n \times \mathbb{R}^n} |x-y|^2 \, d\pi(x,y)$$

TYPICAL FUNCTIONALS

- Internal Energy

includes $U(x) = x \log x$.

$$\mathcal{U}(\rho) = \underbrace{\int_{\mathbb{R}^n} \rho(x) \log \rho(x) \, dx}_{\text{entropy}}$$

$$\mathcal{U}(\rho) = \int_{\mathbb{R}^n} \underbrace{U(\rho(x))}_{= \text{ density of internal energy}} dx,$$

where $U: \mathbb{R}_+ \to \mathbb{R} \cup \{+\infty\}$ measurable

- Potential Energy:

$$\mathcal{V}(\rho) = \int_{\mathbb{R}^n} \overset{\text{potential function.}}{V} \, d\rho,$$

where $V: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ measurable

- Interaction Energy

$$\mathcal{W}(\rho) = \frac{1}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n} W(x-y)\, d\rho(x)\, d\rho(y),$$

$\to$ interaction potential

$W : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$; measurable

The domains of $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{W}$ depends on the behavior of U, V and W.

- $\mathcal{U}$ – well-defined on $P_{ac}(\mathbb{R}^n)$, as long as $U \geq 0$.
- $\mathcal{V}$ (resp. $\mathcal{W}$) is defined on $P(\mathbb{R}^n)$, with values in $\mathbb{R} \cup \{+\infty\}$. as long as V (resp. W) is bounded below by some real number.

Remark

Some examples:

- $U(\rho) = \rho \log \rho$ $\quad U(0) = 0$ But $U(\rho) \ngeq 0$

$$\mathcal{U}(\rho) = \int_{\mathbb{R}^n} \rho \log \rho \, dx$$ Entropy

or Boltzmann's H functional.

- $\mathcal{V}(\rho) = \int_{\mathbb{R}^n} V(x)\, \rho(x)\, dx$

$$V(x) = \frac{1}{2}|x|^2.$$

or $\mathcal{V}(\rho) = \int_{\mathbb{R}^n} \frac{1}{2}|v|^2 \rho(v)\, dv \leftarrow$ Kinetic Energy

- $W(x,y) = \begin{cases} \dfrac{1}{|x-y|} & \text{for } n=3 \\ -\log|x-y| & \text{for } n=2 \end{cases}$

Coulomb potential.

Theorem (McCann) (Criteria for displacement convexity)

Let $\mathcal{P}$ be a displacement convex subset of either (for i)) or $P_2(\mathbb{R}^d)$ (for ii) and iii)). Then

i) If $\mathcal{U}$ satisfies $\mathcal{U}(0) = 0$ and

$$\psi : r \mapsto r^n\, \mathcal{U}(r^{-n}) \text{ is convex}$$

nonincreasing on $(0, +\infty)$, then $\mathcal{U}$ is displacement convex on $\mathcal{P}$.

(Note that $\mathcal{U} : \mathbb{R}_+ \to \mathbb{R} \cup \{+\infty\}$)

$$r \to r^n \mathcal{U}(r^{-n}) = r^n \underbrace{r^{-n}} \log r^{-n} = \log r^{-n} = \underset{\underset{\psi(r)}{\shortparallel}}{\boxed{-n \log r}}$$

$$\begin{cases} \mathcal{U}(\rho) = \rho \log \rho \\ \Psi(r) = -n \log r \quad \Psi'(r) = -n \, 1/r \\ \qquad\qquad\qquad\qquad \Psi''(r) = n \, 1/r^2 \geq 0 \\ \qquad\qquad\qquad\qquad \Psi(r) \downarrow \text{ and convex} \end{cases}$$

$\Rightarrow \mathcal{U}(\rho) = \int_{\mathbb{R}^n} \rho \log \rho \, dx$ is displacement convex

iv) If V is convex (resp. strictly convex, λ-uniformly convex, semi-convex with C),

then $\mathcal{V}$ is displacement convex (resp. strictly displacement convex, λ-uniformly convex, semi-displacement convex with constant C) on P.

Conversely, if $\mathcal{V}(\rho)$ is displacement convex on $P_2(\mathbb{R}^n)$, then V is convex.

viii) If W is convex (resp. semi-convex with C), then W is displacement convex (resp. semi-displacement-convex with constant C).

(Slightly different statement for strictly convex / λ-uniform convex case

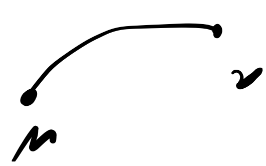
$\mathcal{P}_m \subset \mathcal{P}$ $\qquad \int x\,d\mu = m \Leftarrow$ mean is given)

Conversely, if W is even continuous and $n \geq 2$, and W is displacement convex on $\mathcal{P}_2(\mathbb{R}^d)$, then W is convex.

Rk: i) the mean = the center of mass

$$= \int x\,d\mu(x)$$

$\mathcal{P}_m \subset \mathcal{P} \cap \mathcal{P}_2(\mathbb{R}^n)$ is a displacement convex set of $\mathcal{P}$ is



$$\rho_t = \big((1-t)x + ty\big)_{\#}\gamma(x,y) \qquad \gamma \in OPT(\mu,\nu)$$

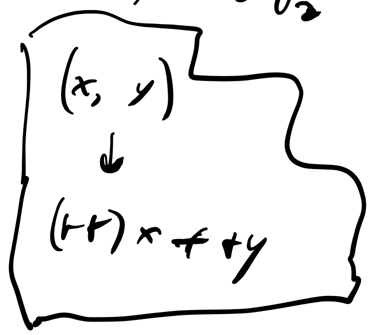
$\mu, \nu \in \mathcal{P}_2$

$$\int_{\mathbb{R}^d} z \, d\rho_t(z)$$

$$= \int_{\mathbb{R}^d} z \; d\big[(1-t)x + ty\big]_{\#}\gamma(x,y)$$

$$= \int_{\mathbb{R}^d} \big[(1-t)x + ty\big] \, d\gamma(x,y)$$

$$= (1-t)m + tm = m$$



ii) $\mathcal{U}(\rho) = \rho \log \rho$

$\rho \in \mathcal{P} = \mathcal{P}_{ac,2}(\mathbb{R}^n) \equiv \mathcal{P}_{ac}(\mathbb{R}^n) \cap \mathcal{P}_2(\mathbb{R}^n)$

then $\mathcal{U}(\rho) = \int_{\mathbb{R}^n} \rho \log \rho \, dx$ is well-defined with values in $\mathbb{R} \cup \{+\infty\}$

Why? Choosing $\varphi(x) = |x|^2$, then

$$\int_{\mathbb{R}^n} \rho\, |x|^2 \le \int_{\mathbb{R}^n} \rho \log \frac{\rho}{M} + \log \int_{\mathbb{R}^n} M \exp(-|x|^2)$$

Gaussian

M is a p.d.f.

$$\int \rho \log \rho \ge \underbrace{\int \rho\, |x|^2 + \int \rho \log m - \log \int m \exp(-|x|^2)}$$

$> -\infty$

iii) More general interaction energies could also be considered,

$$\int W(L(x_1, \cdots x_k))\, d\rho(x_1) \cdots d\rho(x_k),$$

where L is an arbitrary linear function.

OPEN Problems:

Besides the three examples stated there, can one find other useful examples of displacement convex functionals?

Some involving the gradients of ρ?

$$\mathcal{U}(\nabla \rho) = ?$$

$$E(\rho) = \frac{1}{m-1} \int \rho^m dx \qquad m > 1$$

Free energy for Porus medium Eq.

$$\partial_t \rho = \Delta \rho^m.$$

5.2.2. Internal energy:

i) Physical meaning of:

Eq (5.1.7): $\Psi : r \mapsto r^n U(r^{-n})$ convex nonincreasing on $(0,+\infty)$

Consider a homogeneous (or uniform) cloud of n-dim gas with mass M in a volume V,

$$\text{density} = \frac{M}{V}$$

Assume that the gas expands:

$$n \mapsto \lambda n$$
$$V \mapsto \lambda^n V$$
$$\text{density} = \frac{M}{\lambda^n V}$$

internal energy: $\underbrace{V\lambda^n} U(\lambda^{-n}\frac{M}{V})$

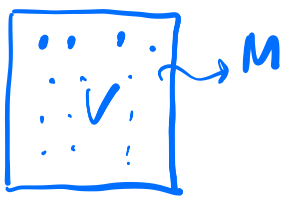
$$\sim r^n U(r^{-n})$$

ii) $U \in C^2$

Thermodynamical pressure:

$$P(\rho) = \rho\, \mathcal{U}'(\rho) - \mathcal{U}(\rho)$$

($\Leftarrow$



$\rho = \frac{M}{V}$ $\qquad P(\rho) = -\frac{d\mathcal{U}}{dV}$

(or $dW = -P\,dV$)

if $V = \infty$, or $\rho = 0$, $\mathcal{U}(0) = 0$.

$$\int_{V_0}^{\infty} \left(-\frac{d\mathcal{U}}{dV}\right) dV = \int_{V_0}^{\infty} -d\mathcal{U} = \overset{\mathcal{U}(\rho_0)V_0}{\overset{\|}{\mathcal{U}(V_0)}} - \overset{0}{\overset{\|}{\cancel{\mathcal{U}(0)}}}$$

$$= \int_{V_0}^{\infty} P\left(\frac{M}{V}\right) dV \quad \left(\rho = M/V \Rightarrow V = M/\rho\right.$$

$$= \int_{V_0}^{\infty} P(\rho)\, d\left(\frac{M}{\rho}\right) = M \int_0^{M/V_0} \frac{P(\rho)}{\rho^2}\, d\rho = \underline{\mathcal{U}(\rho_0)V_0}$$

I don't know why?

$$\mathcal{U}(\rho_0) = \rho_0 \int_0^{\rho_0} \frac{P(\rho)}{\rho^2}\, d\rho$$

It forces that $P(0) = 0$.

$$\frac{\mathcal{U}(\rho)}{\rho} = \int_0^{\rho} \frac{P(s)}{s^2}\, ds$$

$$\left(\frac{\mathcal{U}(\rho)}{\rho}\right)' = P(\rho)/\rho^2$$

$$= \frac{U'(\rho)\rho - U(\rho)}{\rho^2} = \frac{P(\rho)}{\rho^2}$$

$\hookrightarrow$ ie $P(\rho) = U'(\rho)\rho - U(\rho)$

iii) $\Psi(r) = r^n \mathcal{U}(r^{-n})$

$$\Psi'(r) = n r^{n-1} \mathcal{U}(r^{-n}) + r^n \underline{\mathcal{U}'(r^{-n})} (-n r^{-n-1})$$

$$= -n \frac{1}{r} U'(r^{-n}) + \frac{n}{r} r^n U(r^{-n})$$

$$P(r^{-n}) = r^{-n} U'(r^{-n}) - U(r^{-n})$$

$$\Psi'(r) = -n r^{n-1} P(r^{-n}) \quad \downarrow$$

$\Leftrightarrow$ $P \geq 0$ (non-negativity of P)

$$\Psi''(r) = n^2 r^{n-2}\left[r^{-n} P'(r^{-n}) - (1-1/n) P(r^{-n})\right] \geq 0$$

$\Leftrightarrow$ $\rho P'(\rho) \geq (1-1/n) P(\rho)$

$\Leftrightarrow$ $\rho \mapsto \dfrac{P(\rho)}{\rho^{1-1/n}}$ is non-decreasing

$$\left(h(\rho) = \frac{P(\rho)}{\rho^{1-1/n}} \right.$$

$$\frac{dh}{d\rho} = \frac{P'(\rho)\,\rho^{1-1/n} - P(\rho)\cdot(1-1/n)\,\rho^{-1/n}}{\rho^{2(1-1/n)}} \geq 0$$

$$\rho P'(\rho) \geq P(\rho)\cdot(1-1/n)$$

$$\Downarrow$$

iv) $P(0)=0.$ $\quad P(\rho) \geq 0.$ $\quad \Rightarrow \quad P'(\rho) \geq 0$

$P(\rho)\ \uparrow$

Also $P'(\rho) = \rho U''(\rho) + U'(\rho) - U'(\rho) = \underbrace{\rho U''(\rho)}$

$\Rightarrow$ $\boxed{U(\rho) \text{ has to be convex.}}$

Examples: (Typical energy densities)

satisfying (5.20): $\rho \longmapsto \dfrac{P(\rho)}{\rho^{1-1/n}}$ $\nearrow$

① $U(\rho) = \rho^\gamma$, $\gamma \geq 1$,

$$P(\rho) = \rho U'(\rho) - U(\rho) = \rho\cdot\gamma\rho^{\gamma-1} - \rho^\gamma$$
$$= (\gamma-1)\rho^\gamma$$

② $\mathcal{U}(\rho) = \rho \log \rho$

$P(\rho) = \rho\, \mathcal{U}'(\rho) - \mathcal{U}(\rho) = \rho\left(1 \cdot \log\rho + \rho \cdot \frac{1}{\rho}\right) - \rho\log\rho = \rho$

③ $\mathcal{U}(\rho) = -\rho^{\gamma} \quad \left(1 - \frac{1}{n} \leq \gamma \leq 1\right)$

$$P(\rho) = \rho\, \mathcal{U}'(\rho) - \mathcal{U}(\rho)$$
$$= \rho(-\gamma \rho^{\gamma-1}) + \rho^{\gamma} = (1-\gamma)\rho^{\gamma}$$

Famous case: $\mathcal{U}(\rho) = \rho^{5/3}$ in dim $n = 3$,

$\mathcal{U}(\rho) = \rho^{5/3}$: semi-classical approximation to the quantum kinetic energy of a 3D gas of fermions.

Pf of theorem 5.15.

part i) : The internal energy case

Assume that $\mathcal{U}(0) = 0$ and

$\Psi : r \mapsto r^n \mathcal{U}(r^{-n})$ is convex nonincreasing on $(0, +\infty)$

Let $\mu, \nu \in \mathcal{P}_{ac}(\mathbb{R}^n)$.

Consider McCann's interpolation:

$$\rho_t = ((1-t)Id + t\nabla\varphi)_\# \mu$$
$$= (Id - t(\underbrace{Id - \nabla\varphi}_{\theta}))_\# \mu$$
$$= (Id - t\theta)_\# \mu \quad \text{with } \varphi\text{-convex}$$

$\nabla\theta$ = Jacobian matrix of θ.

Applying theorem 4.8

$$\boxed{\mathcal{U}(\rho_t) = \int_{\mathbb{R}^n} U\left(\frac{\rho(x)}{\det(I_n - t\nabla\theta(x))}\right)\det(I_n - t\nabla\theta(x))\,dx}$$

We assume this first, we will com back to this later.

As a function of t, the integrand in RHS is given by the composition of two mappings

$$\begin{cases} t \mapsto \lambda = (\det(I_n - tS))^{1/n} \\ \lambda \mapsto U\left(\frac{r}{\lambda^n}\right)\lambda^n \end{cases}$$

$r = \rho(x)$

$S = \nabla\theta(x) \Leftarrow$ symmmetric

$\nabla\theta = I_n - \nabla^2\varphi \leq I_n$
clear.

Recall:

Lemma 5.21 (Concavity of $\det^{1/n}$)

Given a symmetric matrix $S \le I_n$, the function $t \mapsto \det(I_n - tS)^{1/n}$ is concave (strictly unless S is a multiple of the identity).

$$t \underset{\text{concave}}{\mapsto} \lambda = f(t) \longmapsto \Psi(\lambda) = \lambda^n U(\lambda^{-n} r)$$

$\downarrow$, convex

$$\Psi \circ f\left(\frac{t_1+t_2}{2}\right) \underset{\Psi \downarrow}{\le} \Psi\left(\frac{f(t_1)+f(t_2)}{2}\right)$$

f-concave

$$\ge \frac{1}{2}\left(f(t_1)+f(t_2)\right)$$

$$\underset{\Psi\text{ convex}}{\le} \frac{1}{2}\left(\Psi\circ f(t_1) + \Psi\circ f(t_2)\right)$$

$\Rightarrow$ $\mathcal{U}(\rho_t)$ is convex w.r.t. t.

Necessary Part of Part i) is left as exercise:

If $\mathcal{U}$ is displacement convex on $\mathcal{P}_{ac}(\mathbb{R}^n)$, then

Ψ is convex.

$(\Leftarrow \ \psi(\lambda) = \lambda^n U(\frac{\rho(x)}{\lambda^n}) \ \cdots \ )$

Lemma 5.23 (Arithmetic–Geometric inequality)

i) Let $(x_i)_{1 \le i \le n}$ and $(\lambda_i)_{1 \le i \le n}$ satisfying

$$x_i \geq 0, \quad \lambda_i \geq 0, \quad \sum_{i=1}^n \lambda_i = 1$$

Then
$$\sum_{i=1}^n \lambda_i x_i \geq \prod_{i=1}^n x_i^{\lambda_i}.$$

ii) Let A and B be two non-negative symmetric $n \times n$ matrices (i.e. $A, B \in S_n^+(\mathbb{R})$), and $\lambda \in [0,1]$. Then

$$\det(\lambda A + (1-\lambda) B)^{1/n} \geq \lambda \det(A)^{1/n} + (1-\lambda)(\det B)^{1/n}.$$

iii) Let $A, B \in S_n^+(\mathbb{R})$ and $\lambda \in [0,1]$.

Then
$$\det(\lambda A + (1-\lambda) B) \geq (\det(A))^{\lambda} (\det B)^{1-\lambda}.$$

Pf: i) $\sum_{i=1}^{n} \lambda_i x_i \geq \prod_{i=1}^{n} x_i^{\lambda_i}$

$$\Updownarrow$$

$$\log\left(\sum_{i=1}^{n} \lambda_i x_i\right) \geq \sum_{i=1}^{n} \lambda_i \cdot \log x_i$$

$(\log x)' = \frac{1}{x}$ $\quad (\log x)'' = -\frac{1}{x^2} < 0$ concave

ii) In view of $\det(\lambda A) = \lambda^n \det(A)$

$$\det(\underbrace{\lambda A}_{A} + \underbrace{(1-\lambda)B}_{B})^{1/n} \geq \det(\lambda A)^{1/n} + \det((1-\lambda)B)^{1/n}$$

It suffices to show

$$\det(A+B)^{1/n} \geq (\det A)^{1/n} + (\det B)^{1/n}.$$

WLOG, assume A is invertiable

$$A + B = A^{1/2}(I + \underbrace{A^{-1/2} B A^{-1/2}}_{C}) A^{1/2}$$

Only need to show

(*) $\det(I_n + C)^{1/n} \geq (\det I_n)^{1/n} + (\det C)^{1/n}$

where $C = A^{-1/2} B A^{-1/2}$.

Since now C is symmetric non-negative

Diagonalize C, i.e. $C = O \begin{pmatrix} c_1 & & 0 \\ & \ddots & \\ 0 & & c_n \end{pmatrix} O^T$,

O - orthogonal

$c_i \geq 0$.

Then ($\det C = \det(CI) = \det(CO^TO)$

$= \det(OCO^T)$

$= \det\left(\begin{pmatrix} c_1 & & 0 \\ & \ddots & \\ 0 & & c_n \end{pmatrix}\right)$

$= \prod_i c_i$

(*) above reduces to

$$\prod_{i=1}^n (1+c_i)^{1/n} \geq 1 + \left(\prod_{i=1}^n c_i\right)^{1/n}.$$

$$\Big(\Leftarrow \prod_i \left(\frac{1}{1+c_i}\right)^{1/n} + \prod_i \left(\frac{c_i}{1+c_i}\right)^{1/n}$$

$$\leq \frac{1}{n}\sum_i \frac{1}{1+c_i} + \frac{1}{n}\sum_i \frac{c_i}{1+c_i} = 1 \Big)$$

okay

3. $\det(\lambda A + (1-\lambda)B)^{1/n} \geq (\det A)^{\lambda/n} (\det B)^{\frac{1-\lambda}{n}}$

$\Uparrow$

ii) + $\lambda(\det A)^{1/n} + (1-\lambda)(\det B)^{1/n}$

Arith-Geo $\geq (\det A)^{\frac{\lambda}{n}} (\det B)^{\frac{1-\lambda}{n}}$. okay

5.2.4. Potential Energy

Part ii) of Thm 5.15.

$\theta = Id - D\varphi$,
φ - convex

<u>Pf</u>:

$$\mathcal{V}(\rho_t) = \int_{\mathbb{R}^n} V d\left[(Id - t\theta)_\# \mu\right]$$

$$= \int_{\mathbb{R}^n} V(x - t\theta(x))\, d\mu(x).$$

To show the convexity of $t \mapsto \mathcal{V}(\rho_t)$,

it suffices to impose the convexity of V.

If V is strictly convex, the convexity of $t \mapsto \mathcal{V}(\rho_t)$ can be degenerate only if $\theta(x) = 0$ for $d\mu$-a.e. x,

$\Rightarrow \rho_0 = \rho_1$.

$$\Big(\quad V\left(x - \tfrac{t_1 + t_2}{2}\theta(x)\right) = V\left(\tfrac{1}{2}\left[(x - t_1\theta(x)) + (x - t_2\theta(x))\right]\right)$$

$$\underset{=}{\leq} \tfrac{1}{2}\left\{V(x - t_1\theta(x)) + V(x - t_2\theta(x))\right\} \Big)$$

If V is λ-uniformly convex, then for all t_1, t_2, σ in $[0,1]$,

$$\sigma V(\rho_{t_1}) + (1-\sigma) V(\rho_{t_2}) - V(\rho_{\sigma t_1 + (1-\sigma) t_2})$$

$$= \int_{\mathbb{R}^n} d\mu(x) \Big\{ \sigma V(x - t_1\theta(x)) + (1-\sigma) V(x - t_2\theta(x))$$

$$- V\big(\sigma(x - t_1\theta(x)) + (1-\sigma)(x_2 - t_2\theta(x))\big) \Big\}$$

$$\geq \lambda \frac{\sigma(1-\sigma)}{2} \int_{\mathbb{R}^n} d\mu(x) \, \underbrace{|(x - t_1\theta(x)) - (x - t_2\theta(x))|^2}_{= |t_1 - t_2|^2 \theta(x)}$$

$$= \lambda \frac{\sigma(1-\sigma)}{2} (t_1 - t_2)^2 \underbrace{\int_{\mathbb{R}^n} |\theta(x)|^2 d\mu(x)}_{= W_2^2(\mu, \nu)}.$$

We see

$\Rightarrow$ $t \mapsto V(\rho_t)$ is uniformly displacement with λ.

Remark: λ- uniform displacement convexity

$$\Updownarrow$$

$$\Big\{ \frac{d^2}{dt^2} F(\rho_t) \geq \lambda W_2^2(\mu, \nu)$$

$$\rho_t = [\mu, \nu]_t$$

Connect the global condition of: $\forall \sigma \in [0,1]$

$$\sigma f(t_1) + (1-\sigma) f(t_2) - f(\sigma t_1 + (1-\sigma) t_2)$$

$(\sigma + 1-\sigma)$

$$\geq \lambda \frac{\sigma(1-\sigma)}{2} (t_1 - t_2)^2 C$$

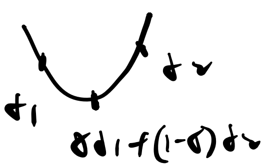
to local condition:

$$\frac{d^2}{dt^2} f(t) \geq \lambda C$$

Leave this as an exercise.

$$\forall t_1, t_2$$

$$\frac{d^2}{dt^2} f(t) \geq \lambda \Leftarrow h(\sigma) = \sigma f(t_1) + (1-\sigma) f(t_2) - f(\sigma t_1 + (1-\sigma) t_2) - \frac{1}{2}\sigma(1-\sigma)(t_1-t_2)^2 \lambda \geq 0.$$



$$\frac{2\sigma[f(t_1) - f(\sigma t_1 + (1-\sigma) t_2)]}{\sigma(1-\sigma)(t_1-t_2)^2} + \frac{2(1-\cancel{\sigma})(f(t_2) - f(\sigma t_1 + (1-\sigma) t_2))}{\sigma(1-\cancel{\sigma})(t_1-t_2)^2} \geq \lambda$$

$$\frac{\dfrac{f(t_1) - f(t_\sigma)}{t_1 - t_\sigma} - \dfrac{f(t_2) \cdot f(t_\sigma)}{t_2 - t_\sigma}}{t_1 - t_2} \geq \lambda/2$$

$$t_\sigma = \sigma t_1 + (1-\sigma) t_2$$

$$t_1 - t_\sigma = (1-\sigma)(t_1 - t_2)$$

$$f(t_1) = f(t_\sigma) + f'(t_\sigma)(t_1 - t_\sigma) + \frac{1}{2} f''(t_\sigma)(t_1 - t_\sigma)^2 + o(|t_1 - t_2|^2)$$

$$f(t_2) = f(t_\sigma) + f'(t_\sigma)(t_2 - t_\sigma) + \frac{1}{2} f''(t_\sigma)(t_2 - t_\sigma)^2 + o(|t_1 - t_2|^2)$$

$$\frac{\frac{1}{2} f''(t_\sigma)\left[(t_1 - t_\sigma) - (t_2 - t_\sigma)\right]}{t_1 - t_2} \geq \lambda/2$$

i.e. $|f''(t)| \geq \lambda$

Conversely,

assuming $\frac{d^2}{dt^2} f(t) \geq \lambda$

$\Rightarrow h(\sigma) \geq 0$

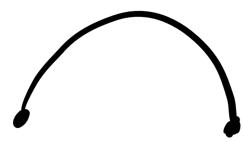
$\forall \sigma \in [0,1], \forall t_1, t_2.$

where

$$h(\sigma) = \sigma f(t_1) + (1-\sigma) f(t_2) - f(\sigma t_1 + (1-\sigma) t_2) - \frac{1}{2}\sigma(1-\sigma)(t_1 - t_2)^2 \lambda$$

$$h(0) = f(t_2) - f(t_2) + 0 = 0 = h(1)$$

$$h'(\sigma) = f(t_1) - f(t_2) - f'(\sigma t_1 + (1-\sigma) t_2)(t_1 - t_2) - \frac{\lambda}{2}|t_1 - t_2|^2 (1 - 2\sigma)$$



$$h''(\sigma) = - f''(\sigma t_1 + (1-\sigma) t_2)(t_1 - t_2)^2 + \lambda |t_1 - t_2|^2$$

$$= \left[- f''(t_\sigma) + \lambda\right] |t_1 - t_2|^2 \leq 0$$

Okay

Exercise: Conversely, if $\mathcal{V}$ is displacement convex on $\mathcal{P}_2(\mathbb{R}^n)$, then V itself is convex.

Part iii): Interaction Energy

Consider an interaction potential W, and

$$\mathcal{W}(\rho) = \frac{1}{2}\int_{\mathbb{R}^n \times \mathbb{R}^n} W(x-y)\, d\rho(x)\, d\rho(y)$$

(W can be replaced by its symmetric part

$$W^S(z) = \frac{1}{2}[W(z) + W(-z)]$$

without affecting the functional $\mathcal{W}$)

WLOG, we may assume that W is even

<u>Important Remark</u>:

The functional $\mathcal{W}$ is in general not convex in the usual sense, except for certain particular cases, namely when W is of positive type $(\hat{W} \geq 0)$.

For example:

W is an <u>inverse power law</u> $\frac{1}{|x|^s}$

or a <u>power law under certain restrictions</u>

- $W(z) = |z|^3$ with restriction that the mean is fixed

$\int x\,p(x)dx = c \quad \forall p \in \mathcal{P}$.

$\boxed{\text{Exercise}}$

The plain convexity of $\mathcal{W}$ is extremely sensitive to the particular form of the interaction potential, while the displacement convexity is a much more robust property.

$$W(z) = \begin{cases} \frac{1}{|z|^{n-2}} & n \geq 3 \\ -\log|z| & n = 2 \end{cases}$$ ← Coulomb potential

Lead to a functional $\mathcal{W}$ which is convex in the usual sense, but not displacement convex.

(It would be interesting to study this!)

ii) $\mathcal{W}$ invariant under translation:

if $\tau_a : x \mapsto x+a$, for some $a \in \mathbb{R}^n$,

then $\mathcal{W}(\tau_{a\#}\rho) = \mathcal{W}(\rho)$.

Pf: Recall again $\rho_t = T_{t\#}\mu$

$= (Id - t\theta)_{\#}\mu,$

with $\theta = Id - D\varphi$ φ - convex.

Hence

$$\mathcal{W}(\rho_t) = \frac{1}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n} W\big(x - t\theta(x) - (y - t\theta(y))\big)\, d\mu(x)\, d\mu(y)$$

$$= \frac{1}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n} W\big((x-y) + t(\theta(x) - \theta(y))\big)\, d\mu(x)\, d\mu(y)$$

W

t

Easily, if W is convex, it follows that

$$t \mapsto \mathcal{W}(\rho_t) \text{ also is.}$$

If W is strictly convex, then

$$W(\rho_{\frac{t_1+t_2}{2}}) \le \frac{1}{2}\left(W(\rho_{t_1}) + W(\rho_{t_2})\right)$$

$$\Leftrightarrow \left\langle W\left((x-y) + \frac{t_1+t_2}{2}(\theta(x)-\theta(y))\right), \rho^{\otimes 2} \right\rangle$$

$$\le \frac{1}{2}\left\langle W\left(x-y+t_1(\theta(x)-\theta(y))\right), \rho^{\otimes 2} \right\rangle$$

$=$

$$+ \frac{1}{2}\left\langle W\left(x-y+t_2(\theta(x)-\theta(y))\right), \rho^{\otimes 2} \right\rangle$$

$\Downarrow$ W strictly convex

$t_1 = t_2$ $\theta(x) \ne \theta(y)$ with $\rho^{\otimes 2}$ — measure 0.

i.e. $\theta(x) = a$ for $d\mu$-a.e. x.

i.e. $\theta(x) = x - \nabla\psi(x) = a \Rightarrow \underbrace{\nabla\psi(x)}_{T(x)} = x - a$

$d\nu(x) = d\mu(x+a)$.

Now, if further W is uniformly convex with constant λ, then

$$\sigma W(\rho_{t_1}) + (1-\sigma) W(\rho_{t_2}) - W(\rho_{\sigma t_1 + (1-\sigma) t_2})$$

$$\geq \frac{\sigma}{2} \int W([x-y] - t_1[\theta(x)-\theta(y)])\, d\mu^{\otimes 2}(x,y)$$

$$+ \frac{1-\sigma}{2} \int W([x-y] - t_2[\theta(x)-\theta(y)])\, d\mu^{\otimes 2}$$

$$- \frac{1}{2} \int W(\sigma([x-y] - t_1[\cdot\,]) + (1-\sigma)([x-y] - t_2[\;]))\, d\mu^{\otimes 2}$$

$$\geq \frac{1}{2} \int d\mu^{\otimes 2} \Big\{ \frac{\lambda}{2} \sigma(1-\sigma) \Big| \cancel{(x-y)} - t_1[\theta(x)-\theta(y)] - \big[\cancel{(x-y)} - t_2[\theta(x)-\theta(y)]\big] \Big|^2 \Big\}$$

$$= \frac{\lambda}{4} \sigma(1-\sigma) |t_1 - t_2|^2 \int d\mu^{\otimes 2} |\theta(x) - \theta(y)|^2$$

The condition of common center of mass

$$\int_{\mathbb{R}^n} x \, d\mu(x) = \int_{\mathbb{R}^n} y \, d\nu(y)$$

$$\Longrightarrow \int_{\mathbb{R}^n \times \mathbb{R}^n} |\theta(x) - \theta(y)|^2 \, d\mu(x)\, d\mu(y)$$

$$= 2 \int \underset{\substack{\| \\ |x - \nabla\psi(x)|^2}}{|\theta|^2} d\mu - 2 \Big(\underset{\substack{\| \\ 0}}{\int \theta(x)\, d\mu(x)} \Big)^2$$

$$= 2\, W_2^2(\mu,\nu)$$

Olcany

$$\int (x - \nabla\psi(x))\, d\mu = \int x\, d\mu - \int y\, d(\nabla\psi)_\# \mu = \int x\, d\mu - \int y\, d\nu = 0$$

<u>Remark:</u> By Jensen's inequality: $\forall$ c-convex c,

$$\int \theta(y)\, d\mu(y) = 0$$

$$\Rightarrow \int_{\mathbb{R}^n\times\mathbb{R}^n} c(\theta(x) - \theta(y))\, \overbrace{d\mu(x)\, d\mu(y)}^{\mu^{\otimes 2}}$$

$$\frac{\theta(x)+\theta(y)}{2} - \frac{\theta(x)}{2}$$

$$= \int_{\mathbb{R}^n} d\mu(x) \int_{\mathbb{R}^n} c(\theta(x) - \theta(y))\, d\mu(y)$$

$$\geq \int_{\mathbb{R}^n} d\mu(x) \left\{ c\left(\theta(x) - \int \theta(y)\, d\mu(y)\right)\right\}$$

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

But when $c = |\cdot|^2$, the inequality can be improved by a factor 2.

Exercise: Necessary condition for displacement convexity of $\mathcal{W}$.

Assume that W is even $(W(-z) = W(z))$ and continuous on $\mathbb{R}^n$, and that $n \geq 2$.

Show that the displacement convexity of $\mathcal{W}$ on $\mathcal{P}_2(\mathbb{R}^n)$ implies the convexity of W.

Above-tangent formulation:

General Fact:

if a function $\Phi : [0,1] \to \mathbb{R} \cup \{+\infty\}$ is λ-uniformly convex, or semi-convex with constant $-\lambda$,

then
$$\Phi(1) \geq \Phi(0) + \frac{d}{dt}\Big|_{t=0}^{+} \Phi(t) + \frac{\lambda}{2},$$

where
$$\frac{d}{dt}\Big|_{t=0}^{+} \Phi(t) = \limsup_{t \downarrow 0^+} \frac{\Phi(t) - \Phi(0)}{t}$$

(True limit since Φ is semi-convex)

Prop 5.29. (Above-tangent formulation of displacement convexity)

Let F be a functional with values in $\mathbb{R} \cup \{+\infty\}$, define on some displacement convex subset $\mathcal{P} \subset \mathcal{P}_{ac}(\mathbb{R}^n)$ or $\mathcal{P}_2(\mathbb{R}^n)$.

Assume that F is λ-uniformly displacement convex, for some $\lambda > 0$; or semi-displacement convex with const $-\lambda \geq 0$. ($\exists \lambda \in \mathbb{R}$, s.t. $\frac{d^2}{dt^2} F(\rho_t) \geq \lambda W_2(\rho_0, \rho_1)^2$ for all ρ_0, ρ_1)

Let ρ_0 and ρ_1 be in $\mathcal{P}$ and let $(\rho_t)_{0 \leq t \leq 1}$ be their displacement interpolation. Then,

$$F(\rho_t) \geq F(\rho_0) + \left.\frac{d}{dt}\right|_{t=0}^{+} F(\rho_t) + \frac{\lambda}{2} W_2^2(\rho_0, \rho_1)$$

To apply the above formula, let us compute $\left.\frac{d}{dt}\right|_{t=0}^{+} F(\rho_t)$.

Thm 5.30 (Practical computation of the tangent)

Let $\mathcal{U}: \mathbb{R}_+ \to \mathbb{R} \cup \{+\infty\}$, $V, W: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ } be measurable functions

s.t. $\mathcal{U}$ satisfies the convexity assumption:

$$\begin{cases} \mathcal{U}(0) = 0 \\ \Psi(r) = r^n \mathcal{U}(r^{-n}) \text{ is convex, nonincreasing on } (0, +\infty) \end{cases}$$

V, W are convex,

W is symmetric. $(W(x-y) = W(y-x))$.

Let $\rho_0, \rho_1 \in \mathcal{P}_{ac}(\mathbb{R}^n)$, s.t. $\mathcal{U}(\rho_0)$, $\mathcal{U}(\rho_1)$, $\rho_0 V$, $\rho_1 V \in L^1(\mathbb{R}^n)$

and $\rho_0(x)\rho_0(y)W(x-y)$, $\rho_1(x)\rho_1(y)W(x-y)$

$\in L^1(\mathbb{R}^n \times \mathbb{R}^n, dx\, dy)$.

Let $\nabla\varphi$ be the unique gradient of a convex function that pushes ρ_0 forward to ρ_1. Then

$$① \left.\frac{d}{dt}\right|_{t=0} \mathcal{U}(\rho_t) = \int_{\mathbb{R}^n} [\mathcal{U}(\rho_0) - \rho_0 \mathcal{U}'(\rho_0)](\Delta_A \varphi - n)$$

$$= -\int_{\mathbb{R}^n} P(\rho_0)(\Delta_A \varphi - n),$$

where Δ_A stands for the Laplace operator in the Aleksandrov sense;

② $\frac{d}{dt}\Big|_{t=0} \mathcal{V}(\rho_t) = \int_{\mathbb{R}^n} \rho_0(x) \nabla V(x) \cdot (\nabla\varphi(x) - x)\, dx$;

and

③ $\frac{d}{dt}\Big|_{t=0} \mathcal{W}(\rho_t)$

$$= \int_{\mathbb{R}^n \times \mathbb{R}^n} \rho_0(x) \rho_0(y)\, \nabla W(x-y) \cdot \left[(\nabla\varphi(x) - x) - (\nabla\varphi(y) - y) \right] dx\, dy.$$

If, on the other hand, V and W are not convex, but only semi-convex, and $W_2^2(\rho_0, \rho_1) < +\infty$, then

then

②': $\frac{d}{dt}\Big|_{t=0}^{+} \mathcal{V}(\rho_t) \geq \int_{\mathbb{R}^n} \rho_0(x) \nabla V(x) \cdot (\nabla\varphi(x) - x)\, dx$

③': $\frac{d}{dt}\Big|_{t=0}^{+} \mathcal{W}(\rho_t) \geq \int_{\mathbb{R}^n \times \mathbb{R}^n} \rho_0(x) \rho_0(y)\, \nabla W(x-y) \cdot \left[(\nabla\varphi(x) - x) - (\nabla\varphi(y) - y) \right] dx\, dy$

Pf:

i) Recall formula (8.11) in the textbook:

For $U : \mathbb{R}_+ \to \mathbb{R}_+ \cup \{+\infty\}$ $U(0) = 0$

U measurable

one has

$$\int_{\mathbb{R}^n} U(g(y))\, dy = \int_{\mathbb{R}^n} U\left(\frac{f(x)}{\det D_A^2 \varphi(x)}\right) \det D_A^2 \varphi(x)\, dx$$

($\mu, \nu \in P_{ac}(\mathbb{R}^n)$, $(\nabla\varphi)_\# \mu = \nu$, φ-convex

$d\mu = f(x)dx$, $d\nu = g(y)dy$ i.e. f. g. p.d.f. of μ and ν.

$\Omega = \mathrm{Int}(\mathrm{Dom}(\varphi))$

$\det D_A^2 \varphi \leftarrow$ determinant of the Hessian of φ,

in the Aleksandrov sense.

$\underbrace{\det D_A^2 \varphi}_{\geq 0} \in L^1_{loc}(\Omega)$. defined a.e.

$M \subset \Omega$: the set of points in Ω where $D_A^2 \varphi$ is defined.

and invertible (also Lebesgue points for $\det D_A^2 \varphi$)

i) $\mu(M) = 1$, $\nu[\partial\varphi(M)] = 1$.

ii) The measure $\det D_A^2 \varphi(x)\, dx$ coincides with A.C.

part of the Hessian $\det_H D^2 \varphi$, is concentrated on M

and satisfies the push-forward formula

$$\nabla\varphi_{\#}[\det D_A^2\varphi(x)\,dx] = 1_{\partial\varphi(\Omega)}\,dx.$$

iii) for a.e. $x \in \mathbb{R}^n$, we have Monge–Ampère eq.

$$\boxed{\det D_A^2\varphi(x)}\, g(\nabla\varphi(x)) = f(x).$$

What is $\det D_A^2\varphi$? See page 60.

Alekasandrov's theorem

(φ – convex $\Rightarrow$ φ is twice differentiable a.e. on $\mathrm{Int}(\mathrm{Dom}(\varphi))$.

Here twice differentiability means:

$\exists$ $D_0^2\varphi = [D^2\varphi]_{ac}$: matrix-valued function L^1_{loc}, ≥ 0 in matrix sense

s.t. $\forall$ a.e. $x \in \mathrm{Int}(\mathrm{Dom}(\varphi))$,

$$\varphi(x+h) = \varphi(x) + \nabla\varphi(x)\cdot h + \langle [D^2\varphi]_{ac}(x)\cdot h, h\rangle + o(|h|^2).$$

$[D^2\varphi]_{ac}$: A.C. part of the Distributional Hessian $D_{\mathcal{D}'}^2\varphi$. This distributional Hessian is the linear form defined on $\mathcal{D}(\Omega)$, $\Omega = \mathrm{Int}(\mathrm{Dom}(\varphi))$ by

$\uparrow$ $C_c^\infty(\Omega)$

$$\langle \underbrace{D_{D'}^2 \varphi}_{\geq 0}, \zeta \rangle = \underbrace{\int_\Omega \varphi \, D^2 \zeta}, \qquad \forall \zeta \in \mathcal{D}(\Omega).$$

non-negative matrix-valued distribution (measure).

$D_{D'}^2 \varphi$: Locally finite.

(each of its component has finite total variation on any compact subset of Ω.)

$D\varphi \in BV_{loc}(\Omega; \mathbb{R}^n)$

$$D^2\varphi^*(D\varphi(x)) \circ D^2\varphi(x) = Id \quad a.e.\ x.$$

Let x_0 be a point φ is twice differentiable in the Aleksandrov sense, then

$D^2\varphi(x_0)$ is invertible $\Leftrightarrow$ φ^* is twice differentiable at $D\varphi(x_0)$ in the Aleksandrov sense

<u>Volume distortion</u>

$$D_A^2 \varphi = [D_{D'}^2 \varphi]_{ac}$$

Whenever φ is twice differentiable at x_0 in the Aleksandrov sense, then

$$\frac{|\partial\varphi(B_r(x_0))|}{|B_r(x_0)|} \xrightarrow{r\to 0} \det\left(D_A^2\varphi(x_0)\right).$$

What we really need is the following:

$$\int_{\mathbb{R}^n} U(g(y))\,dy \overset{y=\nabla\varphi(x)}{=} \int_{\mathbb{R}^n} U(g(\nabla\varphi(x)))\cdot \det D_A^2\varphi(x)\,dx$$

$$= \int_{\mathbb{R}^n} U\left(\frac{f(x)}{\det D_A^2\varphi(x)}\right) \det D_A^2\varphi(x)\,dx$$

using Monge-Ampère Eq:

McCann: $\boxed{\det D_A^2\varphi(x)\, g(\nabla\varphi(x)) = f(x)}$

Going back to the proof:

i) $\dfrac{\mathcal{U}(\rho_t) - \mathcal{U}(\rho_0)}{t}$

$$\rho_t = (T_t)_\# \mu = (\underbrace{(1-t)I_n + t\nabla\varphi}_{= T_t})_\# \mu.$$

$$= \int_{\mathbb{R}^n} \frac{1}{t}\left\{ U\left(\frac{\rho_0(x)}{\det[(1-t)I_n + t D_A^2\varphi(x)]}\right)\cdot \det[(1-t)I_n + t D_A^2\varphi(x)]\right.$$

$$- U(\rho_0(x)) \Big\} \, dx,$$

which $= \displaystyle\int_{\mathbb{R}^n} \frac{1}{t} \Big\{ u(t,x) - u(0,x) \Big\} \, dx.$

From our assumptions: both $u(1,x) = U(\rho_1(x))$
$u(0,x) = U(\rho_0(x))$

are integrable

For a.e. x,

$t \mapsto u(t,x)$ is well-defined and convex

Hence

its slope $\frac{1}{t}(u(t,x) - u(0,x))$

is non-increasing as $t \downarrow 0^+$,

and converges monotonically to $u'(0,x)$.

Note that

$$u(t,x) = U\left(\frac{\rho_0(x)}{\det(I_n + t(D_\theta^2 \varphi(x) - I_n))} \right) \det(I_n + t(D_\theta^2 \varphi(x) - I_n))$$

$$u'(0,x) = -U'(\rho_0(x)) \, \rho_0(x) \left. \frac{d}{dt} \right|_{t=0} \det\Big(I_n + t(D_\theta^2 \varphi(x) - I_n)\Big)$$

$$+ \ \mathcal{U}(\rho_0(x)) \left.\frac{d}{dt}\right|_{t=0} \det\left(I_n + t(D_\theta^2 \varphi(x) - I_n)\right)$$

Recall the Jacobi formula:

$$\frac{d}{dt} \det(\Phi(t)) = \det(\Phi(t)) \operatorname{Tr}\left((\Phi(t))^{-1} \dot{\Phi}(t)\right)$$

in our case $\quad \Phi(0) = I_n \quad \Phi^{-1}(0) = I_n$

$$\dot{\Phi}(t) = D_\theta^2 \varphi(x) - I_n$$

$$\left.\frac{d}{dt}\right|_{t=0} \bullet = \operatorname{Tr}\left(D_\theta^2 \varphi(x) - I_n\right)$$

$$= \boxed{\Delta_\theta^2 \varphi(x) - n}$$

To sum it up,

$$u'(0,x) = \left[\mathcal{U}(\rho_0(x)) - \rho_0(x)\, \mathcal{U}'(\rho_0(x))\right]\left(\Delta_\theta \varphi(x) - n\right).$$

Then the conclusion follows by an application of the monotone convergence theorem.

ii) $$\frac{\mathcal{V}(\rho_t) - \mathcal{V}(\rho_0)}{t} = \int \rho_0(x) \Big[\underbrace{\frac{V((1-t)x + t\nabla\varphi(x)) - V(x)}{t}}_{} \Big],$$

$$\Big\downarrow \begin{array}{l} V \text{ convex} \\ t \to 0 \end{array}$$

$$\nabla V(x) \cdot (\nabla\varphi(x) - x)$$

okay if V is convex;

If V is semi-convex, then $\exists\, C \in \mathbb{R}$, s.t. for $t \in \frac{1}{2}$,

$$\frac{\mathcal{V}(\rho_t) - \mathcal{V}(\rho_0)}{t} \geq \int \rho_0(x) \cdot \nabla V(x) \cdot (\nabla\varphi(x) - x)\, dx$$
$$- \frac{C}{2} \int \rho_0(x) |\nabla\varphi(x) - x|^2 \, dx$$

(V semi-convex with const C,

i.e. $V(x) + \frac{C|x|^2}{2}$ is convex. $\Uparrow$

$$V(x+th) + \frac{C}{2}|x+th|^2 - \left(V(x) + \frac{C}{2}|x|^2\right)$$
$$\geq \left(\nabla V(x) + C \cdot x\right) \cdot th. \Big)$$

$$\Leftrightarrow V(x+th) - V(x) \geq \nabla V(x) \cdot th - \frac{C}{2} t^2 h^2$$

Or

$$\frac{V(x+th)-V(x)}{t} \geq \nabla V(x)\cdot h - \frac{c}{2}th^2$$

$$\Rightarrow \frac{d}{dt}\Big|_{t=0}^{+} \mathcal{V}(\rho_t) \geq \int \rho_0(x)\, \nabla V(x)\cdot(\nabla\varphi(x) - x).$$

iii) Leave as exercise

Riemannian manifolds:

<u>Ricci curvature</u> plays a crucial role.

$\mathcal{U}(\rho)$: further need: $\mathrm{Ric} \geq 0$.

$\mathcal{V}(\rho)$: $\quad \nabla^2 V + \underbrace{\mathrm{Ric}}_{\text{Ricci tensor}} \geq 0.$

<u>Applications: Uniqueness of ground state</u>

$\Downarrow$

Motivation for displacement convexity

The following result is due to McCann.

Thm 5.32 (Strict displacement convexity implies uniqueness of minimizer.)

Consider the energy functional, defined for A.C. probability measures on $\mathbb{R}^n$:

$$F(\rho) = \int_{\mathbb{R}^n} U(\rho(x))\,dx + \int_{\mathbb{R}^n} V(x)\,\rho(x)\,dx + \frac{1}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n} W(x-y)\,d\rho(x)\,d\rho(y),$$

Assume that

- $\Phi(r) = r^n U(r^{-n})$ $\downarrow$, convex;
- $\inf V > -\infty$;
- V and W are convex

Assume that V (resp. W) is strictly convex. Then, there is at most one minimizer (resp. at most one minimizer, up to translation) for $F(\rho)$, $\rho \in \mathcal{P}_{ac}(\mathbb{R}^n)$

Pf: Take $\rho_1, \rho_2 \in \mathcal{P}_{ac}(\mathbb{R}^n)$ be two distinct A.C. minimizers, and set $\rho = [\rho_1, \rho_2]_{1/2}$.

Hence F is strictly displacement convex, i.e. $t \mapsto F([\rho_1, \rho_2]_t)$ is strictly convex.

But strict convexity implies that

$$F(\rho) < \frac{1}{2}[F(\rho_1) + F(\rho_2)]$$

impossible. $\blacksquare$

Materials from Figalli and Glaudo

Optimal Transport $\leftrightarrow$ gradient flows $\leftrightarrow$ PDEs

Wasserstein distances

Gradient flow in Hilbert space

JKO scheme

---

p-Wasserstein distance and geodesics

Def: Let (X, d) locally cpt, separable metric space.

Given $1 \le p < \infty$, let

$$\mathcal{P}_p(X) := \left\{ \sigma \in \mathcal{P}(X) \,\middle|\, \int_X d(x, x_0)^p \, d\sigma(x) < +\infty \text{ for some } x_0 \in X \right\}.$$

$\Downarrow$

the set of probability measures with finite p-moment.

One has:

$$d(x, x_1)^p \overset{\text{triangle}}{\le} \left(d(x, x_0) + d(x_0, x_1) \right)^p$$
$$\le 2^{p-1} \left(d(x, x_0)^p + d(x_0, x_1)^p \right)$$

$$\left(\left(\frac{s_1 + s_2}{2} \right)^p \le \frac{1}{2} \left(s_1^p + s_2^p \right) \Leftarrow s \mapsto s^p \text{ convex} \right)$$

Def: $\mu, \nu \in \mathcal{P}_p(X)$, define their p-Wasserstein distance

as
$$W_p(\mu,\nu) := \left(\inf_{\gamma \in \Pi(\mu,\nu)} \int_{X \times X} d(x,y)^p \, d\gamma(x,y) \right)^{1/p} \underset{\substack{\text{since} \\ \mu, \nu \in \mathcal{P}_p(X)}}{<} +\infty$$

$$\Big(\Leftarrow \int_{X \times X} d(x,y)^p \, d\gamma$$

$$\leq 2^{p-1} \int_{X \times X} \left(d(x,x_0)^p + d(x_0,y)^p \right) d\gamma$$

$$= 2^{p-1} \Big(\int_X d(x,x_0)^p \, d\mu + \int_X d(y,x_0)^p \, d\nu \Big) \quad \Big)$$

$$< \infty$$

Thm: W_p is a distance on $\mathcal{P}_p(X)$.

($\cdot \geq 0$, $\cdot$ symmetric $\cdot$ triangle)

The connection between the Wasserstein topology and the weak-* topology

Thm: Let $1 \leq p < \infty$ and a base point $x_0 \in X$.

Let $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}_p(X)$ be a sequence of prob. measures,

let $\mu \in \mathcal{P}_p(X)$.

The following statement are equivalent:

a) $\mu_n \xrightarrow{*} \mu$ and $\int_X d(x_0, x)^p \, d\mu_n \to \int_X d(x_0, x)^p \, d\mu$

b) $W_p(\mu_n, \mu) \longrightarrow 0$

$\boxed{\text{Corollary 3.1.7}}$

Let X (ambient space) be compact, $p \geq 1$, $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}_p(X)$ a sequence of probability measures, and $\mu \in \mathcal{P}_p(X)$. Then.

$$\mu_n \xrightarrow{*} \mu \iff W_p(\mu_n, \mu) \longrightarrow 0$$

<u>Construction of geodesics</u>

$X = \mathbb{R}^d$, $\gamma \in \Pi(\mu, \nu)$ be an optimal coupling for W_p, $c(x,y) = |x-y|^p$. $\gamma \in OPT(\mu, \nu)$

Set $\pi_t(x, y) = (1-t)x + ty$, so that

$$(\pi_0)_\# \gamma = \mu, \quad (\pi_1)_\# \gamma = \nu.$$

Define $\mu_t := (\pi_t)_\# \gamma$

$$\gamma_{s,t} = (\pi_s, \pi_t)_\# \gamma \in \Pi(\mu_s, \mu_t)$$

$$W_p(\mu_s, \mu_t) \le \left(\int_{X \times X} |z - z'|^p \, d\gamma_{s,t}(z, z') \right)^{1/p}$$

$$= \left(\int_{X \times X} |\pi_s(x,y) - \pi_t(x,y)|^p \, d\gamma(x,y) \right)^{1/p}$$

$$(1-s)x + sy - [(1-t)x + ty]$$

$$= (s-t)(y-x)$$

$$= |t-s| \left(\int_{X \times X} |x-y|^p \, d\gamma \right)^{1/p} = |t-s| \, W_p(\mu_0, \mu_1).$$

$$W_p(\mu_0, \mu_s) + W_p(\mu_s, \mu_t) + W_p(\mu_t, \mu_1) \qquad 0 \le s < t \le 1.$$

$$\le (s + t-s + 1-t) \, W_p(\mu_0, \mu_1) = W_p(\mu_0, \mu_1)$$

$\ge$ also true by triangle

Hence $W_p(\mu_s, \mu_t) = |t-s| \, W_p(\mu_0, \mu_1)$

$\forall \, 0 \le s, t \le 1$

<u>Def:</u> A curve of measures

$(\mu_t)_{t \in [0,1]} \subset \mathcal{P}_p(\mathbb{R}^d)$ is said to be a constant speed geodesic if

$$W_p(\mu_s, \mu_t) = |t-s| \, W_p(\mu_0, \mu_1)$$

$$\forall \; 0 \leq s, t \leq 1.$$

Any optimal coupling $\gamma \in OPT(\mu, \nu)$ induces a geodesic via $\mu_t := (\pi_t)_\# \gamma$

In particular case when $\gamma = (Id \times T)_\# \mu$ is induced by a map,

$$\mu_t = (\pi_t)_\# (Id \times T)_\# \mu$$

$$= ((1-t) Id + t T)_\# \mu$$

An informal introduction to gradient flows in Hilbert spaces

$\mathcal{H}$: Hilbert space $\quad \phi : \mathcal{H} \to \mathbb{R} \quad C^2$

($\mathcal{H} = \mathbb{R}^d$ as a first example) $\quad x_0 \in \mathcal{H}$

Gradient flow of ϕ starting at x_0 is given by

$$\begin{cases} \dot{x}(t) = -\nabla\phi(x(t)) \\ x(0) = x_0 \end{cases}$$

Then:

$$\frac{d}{dt}\phi(x(t)) = \nabla\phi(x(t)) \cdot \dot{x}(t) = -|\nabla\phi(x(t))|^2 \leq 0$$

- ϕ decreases along the curve $x(t)$;
- $\frac{d}{dt}\phi(x(t)) = 0$ iff $|\nabla\phi|(x(t)) = 0$, i.e.

  $x(t)$ is a critical point of ϕ.

If ϕ has a unique stationary point that coincides with the global minimizer (for instance if ϕ is strictly convex),

then one expects $x(t) \xrightarrow{t\to\infty} x^*$: the minimizer

Rk: • $\phi : \mathcal{H} \to \mathbb{R}$

$$\underbrace{d\phi(x)}_{H^*}[v] = \lim_{\varepsilon\to 0} \frac{\phi(x+\varepsilon v) - \phi(x)}{\varepsilon}$$

H^* : dual space of $\mathcal{H}$.

- If $t \longmapsto x(t) \in \mathcal{H}$ is a A.C. curve, then

$$\dot{x}(t) = \lim_{\varepsilon \to 0} \frac{x(t+\varepsilon) - x(t)}{\varepsilon} \in \mathcal{H}$$

So $\left.\begin{array}{l} \dot{x}(t) \in \mathcal{H} \\ d\phi(x(t)) \in \mathcal{H}^* \end{array}\right\}$ live in different spaces.

<u>Identification</u> $\mathcal{H}$. $\mathcal{H}^*$

$\langle \cdot, \cdot \rangle_{\mathcal{H}}$: scalar product on $\mathcal{H} \times \mathcal{H}$.

One can define the gradient of ϕ at x as the unique element of $\mathcal{H}$ s.t. $(\nabla\phi(x))$

$$\langle \nabla\phi(x), v \rangle_{\mathcal{H}} = d\phi(x)\, v, \quad \forall\, v \in \mathcal{H}.$$

The scalar product allows us to identify the gradient and the differential.

we can then make sense of

$$\dot{x}(t) = -\nabla\phi(x(t)).$$

How to construct a solution to the gradient flow

$\nabla\phi$: Lipschitz, Cauchy–Lipschitz theorem.

$\nabla\phi$: C^0 : Peano theorem on existence

NOT even continuous.

<u>Classical ways to construct solutions</u>:

Implicit Euler Scheme.

$\tau > 0$ small fixed time step

$$\dot{x}(t) \longrightarrow \frac{x(t+\tau) - x(t)}{\tau}$$

now: $$\frac{x(t+\tau) - x(t)}{\tau} = -\nabla\phi(y)$$

$y = x(t)$: explicit Euler scheme

$y = x(t+\tau)$: implicit Euler scheme.

works better

Find $x(t+\tau) \in \mathcal{H}$, s.t.

$$\frac{x(t+\tau) - x(t)}{\tau} = -\nabla\phi(x(t+\tau)).$$

$$x_0^\tau = x_0$$

$k \geq 0$ $x_k^\tau \to x_{k+1}^\tau$ by solving $\dfrac{x_{k+1}^\tau - x_k^\tau}{\tau} = -\nabla\phi(x_{k+1}^\tau)$

or equivalently

$$\nabla_x \left(\frac{\|x - x_k^\tau\|^2}{2\tau} + \phi(x) \right) \Big|_{x = x_{k+1}^\tau}$$

$$= \frac{x_{k+1}^\tau - x_k^\tau}{\tau} + \nabla\phi(x_{k+1}^\tau) = 0,$$

where $\|\cdot\| \leftarrow$ norm induced by the scalar product.

x_{k+1}^τ is a critical point of the function:

$$\psi_k^\tau(x) = \frac{\|x - x_k^\tau\|^2}{2\tau} + \phi(x)$$

Find x_{k+1}^τ via looking for a global minimizer of ψ_k^τ

Assume: $\phi : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ (not necessarily C^1)

$\Downarrow$ $\underline{\text{convex}}$, l.s.c.

one can define sub-differential.

Def: An A.C. curve $x : [0, +\infty) \to \mathcal{H}$ is a gradient flow for the convex and l.s.c. function ϕ with initial point $x_0 \in \mathcal{H}$ if
$$\begin{cases} \dot{x}(t) \in \partial\phi(x(t)), & \text{a.e. } t > 0 \\ x(0) = x_0 \end{cases}$$

$x_0^\tau = x_0$

$k \geq 0$ given x_k^τ, we look for x_{k+1}^τ satisfying

$$\frac{x_{k+1}^\tau - x_k^\tau}{\tau} \in -\partial\phi(x_{k+1}^\tau).$$

$$\Updownarrow$$

$$0 \in \frac{x_{k+1}^\tau - x_k^\tau}{\tau} + \partial\phi(x_{k+1}^\tau) =: \partial\psi_k^\tau(x_{k+1}^\tau),$$

$$\psi_k^\tau(x) := \frac{\|x - x_k^\tau\|^2}{2\tau} + \phi(x).$$

$0 \in \partial\psi_k^\tau(x_{k+1}^\tau) \iff x_{k+1}^\tau$ is a global minimizer of ψ_k^τ.

Given x_k^τ;

$$x_{k+1}^\tau = \arg\min_x \left(\psi_k^\tau(x) := \phi(x) + \frac{\|x - x_k^\tau\|^2}{2\tau} \right)$$

Existence of minimizer ✓.

$\hookrightarrow (x_k^\tau)_{k \geq 0}$.

Setting $x^\tau(0) := x_0$, $x^\tau(t) := x_k^\tau$ for $t \in ((k-1)\tau, k\tau]$,

one obtains a (piecewise constant) curve

$t \mapsto x^\tau(t)$ (which is expected to converge to (GF))

Difficulty: $\tau \to 0$ and prove that $\exists$ a limit curve $x(t)$ that solves
$$\begin{cases} \dot{x}(t) \in -\partial\phi(x(t)), & \text{a.e. } t>0 \\ x(0) = x_0 \end{cases}$$
(GF)

See AGS for details

RK: (Uniqueness and Stability)

Let ϕ be a convex function, and let $x(t)$, $y(t)$ be solutions of (GF) with initial conditions x_0, y_0.

If $\phi \in C^2$, then

$$\frac{d}{dt}\frac{\|x(t)-y(t)\|^2}{2} = \langle x(t)-y(t), \dot{x}(t)-\dot{y}(t)\rangle$$
$$= -\langle x(t)-y(t), \nabla\phi(x(t))-\nabla\phi(y(t))\rangle$$
$$\leq 0 \quad (\text{by convexity})$$

More generally,

if ϕ is convex, but not necessarily C^2,

$$\dot{x}(t) = -p(t), \qquad p(t) \in \partial\phi(x(t)),$$
$$\dot{y}(t) = -q(t), \qquad q(t) \in \partial\phi(y(t)),$$

and therefore

$$\frac{d}{dt}\frac{\|x(t)-y(t)\|^2}{2} = \langle x(t)-y(t), \dot{x}(t)-\dot{y}(t)\rangle$$
$$= -\underbrace{\langle x(t)-y(t), p(t)-q(t)\rangle}_{\text{monotonicity of } \partial\phi} \le 0.$$

In both cases, the gradient flow is unique.

also stability.

Example: Let $\mathcal{H} = L^2(\mathbb{R}^d)$ and

$$\phi(u) = \begin{cases} \frac{1}{2}\int_{\mathbb{R}^d} |\nabla u|^2 dx, & \text{if } u \in W^{1,2}(\mathbb{R}^d) \\ +\infty & \text{otherwise} \end{cases}$$

$\rightsquigarrow$ Dirichlet

Claim: $\partial\phi(u) \neq \phi \iff \Delta u \in L^2(\mathbb{R}^d)$,

and in that case

$$\partial\phi(u) = \{-\Delta u\}$$

(The gradient flow is just $\partial_t u = \Delta u$ ← Heat Eq.)

Pf: $\Rightarrow$ Let $p \in \underbrace{L^2(\mathbb{R}^d)}_{H}$ with $p \in \partial\phi(u)$.

Then by definition, for any $v \in L^2(\mathbb{R}^d)$

$$\phi(v) \geq \phi(u) + \langle p, v-u \rangle_{L^2}.$$

Take $v = u + \varepsilon w$ with $u, w \in W^{1,2}(\mathbb{R}^d)$
$\varepsilon > 0$.

Then the eq. takes the form

$$\int_{\mathbb{R}^d} \frac{|\nabla(u+\varepsilon w)|^2}{2}\,dx - \int_{\mathbb{R}^d} \frac{|\nabla u|^2}{2}\,dx$$

$$= \varepsilon \int_{\mathbb{R}^d} \nabla u \cdot \nabla w\,dx + \frac{1}{2}\varepsilon^2 \int_{\mathbb{R}^d} |\nabla w|^2$$

$$\geq \varepsilon \int_{\mathbb{R}^d} p \cdot w\,dx$$

i.e. $\int \nabla u \cdot \nabla w \geq \int p \cdot w\,dx$ (sending $\varepsilon \to 0+$)
$\forall w \in W^{1,2}(\mathbb{R}^d)$

Replacing w with $-w$ above,

$$\int_{\mathbb{R}^d} \underbrace{-\Delta u}_{\in \mathcal{D}'}\, w = \int_{\mathbb{R}^d} \nabla u \cdot \nabla w\,dx = \int_{\mathbb{R}^d} p \cdot w\,dx$$
$\forall w \in W^{1,2}(\mathbb{R}^d)$

i.e. $\quad -\Delta u = p \in L^2(\mathbb{R}^d)$

$\Leftarrow$ Assume that $\Delta u \in L^2(\mathbb{R}^d)$

By def. of ϕ, for any $w \in W^{1,2}(\mathbb{R}^d)$,

$$\phi(u+w) - \phi(u) = \int_{\mathbb{R}^d} \nabla u \cdot \nabla w \, dx + \frac{1}{2}\int_{\mathbb{R}^d} |\nabla w|^2 \, dx$$

$$\geq \int_{\mathbb{R}^d} \nabla u \cdot \nabla w \, dx = \int_{\mathbb{R}^d} \underbrace{-\Delta u}_{\in L^2(\mathbb{R}^d)} \, w \, dx$$

if $w \notin W^{1,2}(\mathbb{R}^d)$,

$$\phi(u+w) = +\infty \geq \phi(u) + \int -\Delta u \, w \, dx$$

trivially.

Thus $\quad -\Delta u \in \partial\phi(u)$

Heat Eq. as gradient flow

Let $\mathcal{H} = L^2(\mathbb{R}^d)$ and consider the Dirichlet energy

$$\phi(u) := \begin{cases} \frac{1}{2}\int_{\mathbb{R}^d} |\nabla u|^2 \, dx, & \text{if } u \in W^{1,2}(\mathbb{R}^d), \\ +\infty & \text{otherwise} \end{cases}$$

The gradient flow of ϕ w.r.t. L^2-scalar product is the heat eq.

$$\partial_t u(t) \in -\partial \phi(u(t)) \iff \partial_t u(t,x) = \Delta u(t,x).$$

Heat eq. and Optimal Transport : The JKO scheme

Previous discussion:

find solutions to the heat eq.
by solving the gradient flow by Implicit Euler:

$$u_{k+1}^{\tau} \in \arg\min_{u} \left(\phi(u) + \frac{\|u - u_k^{\tau}\|^2_{L^2(\mathbb{R}^d)}}{2\tau} \right)$$

and then $\tau \to 0$

JKO: New and surprising way of constructing solutions of the heat eq. as GFs.

Replace $\int |\nabla u|^2 dx$ by $\int \rho \log \rho$

L^2 - norm   2-Wasserstein distance.

New Implicit Euler Scheme (JKO scheme)

$$\rho_{k+1}^{\tau} \in \underset{\rho \in \mathcal{P}(\mathbb{R}^d)}{\arg\min} \left(\int_{\mathbb{R}^d} \rho \log \rho \, dx + \frac{W_2^2(\rho, \rho_k^{\tau})}{2\tau} \right)$$

Easily extend to the Fokker-Planck eq.

$$\partial_t \rho = \Delta \rho + \operatorname{div}(\rho \nabla V)$$

Let us assume $\rho_0 \in \mathcal{P}(\mathbb{R}^d)$

(but $\rho_0 \in L_+^2(\mathbb{R}^d)$ should also work).

Consider the setting of a bounded convex domain $\Omega \subset \mathbb{R}^d$. $\rho_0 \in \mathcal{P}(\Omega)$ ρ_0: p.d.f.

$$\underbrace{\int_{\Omega} \rho_0 \log \rho_0 \, dx}_{\text{entropy}} < +\infty$$

Fix $\tau > 0$, set $\rho_0^{\tau} := \rho_0$, and given ρ_k^{τ} we define ρ_{k+1}^{τ} as the minimizer of

$$\rho \mapsto \frac{W_2^2(\rho, \rho_k^{\tau})}{2\tau} + \int_{\Omega} \rho \log \rho \, dx$$

Later, show as $\tau \to 0$,

the scheme (ρ_k^{τ}) converges to the solution of the heat eq.

Section 4.4 : Linear Fokker-Planck eq.

(FP) $$\partial_t \rho = \Delta\rho + \mathrm{div}(\rho\nabla V)$$

where $\rho : [0,\infty)\times\mathbb{R}^d \to \mathbb{R}^+$,

$V : \mathbb{R}^d \to \mathbb{R}$ is a C^2 convex function.

density ↑ external confining potential

FP is a gradient flow in the Wasserstein space.

↳ Quantitative convergence rates to equilibrium

i.e. $\rho(t) \to \rho_\infty$ as $t \to +\infty$

(In particular, we prove a Logarithmic Sobolev inequality.)

Consider the functional:

$\mathcal{F} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ as

$$\mathcal{F}[\rho] = \begin{cases} \int_{\mathbb{R}^d} (\rho\log\rho + \rho V)\,dx, & \text{if } \rho \ll Leb, \\ +\infty & \text{otherwise.} \end{cases}$$

Note that writing $\eta = \rho / e^{-V} = e^{V} \rho$

$$\mathcal{F}[\rho] = \int_{\mathbb{R}^d} \eta \log \eta \, e^{-V} dx$$

$$\left(= \int \rho \log \frac{\rho}{e^{-V}} dx \right)$$

Say

$\underbrace{\frac{1}{Z} e^{-V}}$ = p.d.f. $\quad Z = \int_{\mathbb{R}^d} e^{-V} dx$

$$H(\rho \mid \tfrac{1}{Z} e^{-V}) = \int \rho \log \frac{\rho}{\frac{1}{Z} e^{-V}}$$

$$= \int \rho \log \rho + \int \rho V + \underbrace{\boxed{\log Z}}_{\text{constant.}}$$

up to an additive constant,

$\mathcal{F}[\rho]$ is the relative entropy $H(\rho \mid \frac{1}{Z} e^{-V})$.

For simplicity, assume $Z = \int e^{-V} dx = 1$.

Now $\mathcal{F}[\rho] = \int \rho \log \frac{\rho}{e^{-V}} dx \geq 0$

This is a consequence of the fact that

the relative entropy is always ≥ 0.

$$H(\mu|\nu) = \int \mu \log \frac{\mu}{\nu} = \int \underbrace{\frac{\mu}{\nu} \log \frac{\mu}{\nu}}_{h(x) = x\log x \text{ is convex}} d\nu$$

$$\geq \left(\int \frac{\mu}{\nu} d\nu\right) \log \left(\int \frac{\mu}{\nu} d\nu\right)$$

$$= 1.$$

Now we compute the Wasserstein gradient of

$\mathcal{F}$:

$$\text{grad}_{W_2} \mathcal{F}[\rho] = -\text{div}\left(\rho \nabla \frac{\delta \mathcal{F}[\rho]}{\delta \rho}\right)$$

where $\dfrac{\delta \mathcal{F}[\rho]}{\delta \rho} = \log \rho + V$

$$\nabla \frac{\delta \mathcal{F}[\rho]}{\delta \rho} = \nabla \log \rho + \nabla V$$

so $$\text{grad}_{W_2} \mathcal{F}[\rho] = -\text{div}\left(\rho \left(\underbrace{\nabla \log \rho + \nabla V}\right)\right).$$

Now FPE reads:

$$\partial_t \rho = -\mathrm{grad}_{W_2} F[\rho]$$

$$= \mathrm{div}(\rho(\nabla \log \rho + \nabla V)) = \Delta\rho + \mathrm{div}(\rho \cdot \nabla V)$$

$$\frac{d}{dt} F[\rho_t] = \langle \mathrm{grad}_{W_2} F[\rho_t], \partial_t \rho_t \rangle_\rho$$

$$= -\langle \mathrm{grad}_{W_2} F[\rho_t], \mathrm{grad}_{W_2} F[\rho_t] \rangle_{\rho_t}$$

$$= -\int |\nabla \log \rho_t + \nabla V|^2 \rho_t \, dx$$

$\underline{\underline{\rho = \eta e^{-V}}}$

$$-\boxed{\int |\nabla \log \eta|^2 \rho} = \underbrace{\int \rho \left|\nabla \log \frac{\rho}{e^{-V}}\right|^2}_{\text{modified Fisher information}}$$

$\log \rho = \log \eta - V$

$\nabla \log \rho = \nabla \log \eta - \nabla V$

$$= -\int |\nabla \log \eta|^2 \eta e^{-V}$$

$$= -\int \frac{|\nabla \eta|^2}{\eta} e^{-V} dx$$

i.e.

Free energy

$$\frac{d}{dt} F[\rho_t] = \int \eta_t \log \eta_t \, e^{-V} dx$$

$$= -\int \frac{|\nabla \eta|^2}{\eta} e^{-V} dx.$$

Recall that

$$F[\rho] = \int \rho \log \rho + \int \rho V$$

Given $V : \mathbb{R}^d \to \mathbb{R}$, we know that

$F[\rho]$ is displacement convex.

$\boxed{\text{We need } \lambda\text{- uniformly convex to obtain quantitative convergence}}$

Def: (λ-convex)

Consider

$\varphi : I \subset \mathbb{R} \longrightarrow \mathbb{R} \cup \{+\infty\}$ l.s.c.

I: Interval

Given $\lambda \in \mathbb{R}$, the function φ is said to be λ-convex

if

$$(1-s)\varphi(x) + s\varphi(y)$$

$$\geq \varphi((1-s)x + sy) + \frac{\lambda s(1-s)}{2}|x-y|^2, \quad \forall x, y \in I, \ 0 \leq s \leq 1.$$

$\left(\Leftrightarrow \varphi(x) - \frac{\lambda}{2}|x|^2 \triangleq \varphi_\lambda(x) \text{ is convex} \right)$

(Checking: $(1-s)\left(\varphi(x) - \frac{\lambda}{2}|x|^2\right) + s\left(\varphi(y) - \frac{\lambda}{2}|y|^2\right)$

$$\geq \varphi((1-s)x+sy) - \frac{\lambda}{2}|(1-s)x+sy|^2 \Big)$$

A l.s.c. function $\varphi: X \to \mathbb{R}\cup\{+\infty\}$ on a geodesic metric space (X,d) is said to be λ-convex if, given any geodesic $\gamma:[0,1]\to X$, the function $\varphi\circ\gamma:[0,1]\to\mathbb{R}\cup\{+\infty\}$ is λ-convex.

$\lambda>0$: λ-uniformly convex

$\lambda<0$: λ-semi-convex

Lemma: Let $i=1$ or 2,

and $\varphi_i: X\to\mathbb{R}\cup\{+\infty\}$ be λ_i-convex.

Then $(\varphi_1+\varphi_2)$ is $(\lambda_1+\lambda_2)$-convex.

Checking:

$$\left.\begin{array}{l}\varphi_1(x)-\frac{\lambda_1}{2}|x|^2\\ \varphi_2(x)-\frac{\lambda_2}{2}|x|^2\end{array}\right\} \text{convex} \Rightarrow \big(\varphi_1(x)+\varphi_2(x)\big)-\frac{\lambda_1+\lambda_2}{2}|x|^2 \text{ convex.}$$

Given $\varphi \in C^2(\mathbb{R}^d)$, φ is λ-convex, then

$$\psi(z) = \varphi(z) - \frac{\lambda}{2}|z|^2 \quad \text{is convex}$$

$$\Longrightarrow \psi(y) \geq \psi(x) + \langle \nabla\psi(x), y-x \rangle,$$

$$\varphi(y) - \frac{\lambda}{2}|y|^2 \geq \varphi(x) - \frac{\lambda}{2}|x|^2 + \langle \nabla\varphi(x), y-x \rangle - \lambda \langle x, y-x \rangle$$

$$\Rightarrow \varphi(y) \geq \varphi(x) + \langle \nabla\varphi(x), y-x \rangle + \frac{\lambda}{2}|y-x|^2, \quad \forall x, y \in \mathbb{R}^d$$

Exchanging the role of x and y

$$\varphi(x) \geq \varphi(y) + \langle \nabla\varphi(y), x-y \rangle + \frac{\lambda}{2}|y-x|^2$$

Then

$$0 \geq -\langle \nabla\varphi(x) - \nabla\varphi(y), x-y \rangle + \lambda |y-x|^2$$

i.e. $\boxed{\langle \nabla\varphi(x) - \nabla\varphi(y), x-y \rangle \geq \lambda |x-y|^2}$

For a λ-convex function $\varphi \in C^2(\mathbb{R}^d, \mathbb{R})$ with $(\lambda > 0)$

Let x_0 be the unique minimum of x_0.

- $\varphi(x) \geq \varphi(x_0) + \langle \underset{\overset{\shortparallel}{0}}{\nabla\varphi(x_0)}, x-x_0 \rangle + \frac{\lambda}{2}|x-x_0|^2$

i.e. $\varphi(x) \geq \varphi(x_0) + \frac{\lambda}{2}|x-x_0|^2$

$$\Rightarrow \sqrt{\frac{2}{\lambda}(\varphi(x)-\varphi(x_0))} \geq |x-x_0|.$$

- $\varphi(x_0) \geq \varphi(x) + \langle \nabla\varphi(x), x_0-x \rangle + \frac{\lambda}{2}|x-x_0|^2$

or

$$\langle \nabla\varphi(x), x-x_0 \rangle \geq \varphi(x)-\varphi(x_0) + \frac{\lambda}{2}|x-x_0|^2,$$

$$\rotatebox{90}{$\geq$}\ \text{C-S.}$$

$$|\nabla\varphi(x)|\,|x-x_0|$$

and

$$|\nabla\varphi(x)| \geq \frac{\varphi(x)-\varphi(x_0)}{|x-x_0|} + \frac{\lambda}{2}|x-x_0|$$

$$\geq \sqrt{2\lambda(\varphi(x)-\varphi(x_0))}.$$

The results above hold true in Wasserstein space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$.

or the following proposition.

Prop: Given a λ-convex l.s.c. functional $\mathcal{F}: \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ with $\lambda > 0$, and

$$\mathcal{F}[\bar{\rho}] = \min_{\rho \in \mathcal{P}_2(\mathbb{R}^d)} \mathcal{F}[\rho].$$

Then

$$W_2^2(\rho, \bar{\rho}) \le \frac{2}{\lambda}\left(\mathcal{F}[\rho] - \mathcal{F}[\bar{\rho}]\right),$$

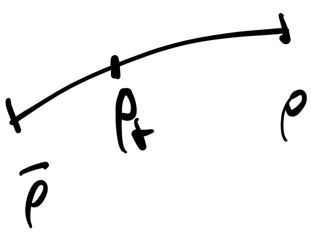
$$\mathcal{F}[\rho] - \mathcal{F}[\bar{\rho}] \le \frac{1}{2\lambda} \langle \mathrm{grad}_{W_2} \mathcal{F}[\rho], \mathrm{grad}_{W_2} \mathcal{F}[\rho] \rangle_\rho$$

$$= \frac{1}{2\lambda} \int \left| \nabla \frac{\delta \mathcal{F}[\rho]}{\delta \rho} \right|^2 d\rho(x)$$

$$\left(\mathrm{grad}_{W_2} \mathcal{F}[\rho] = -\mathrm{div}\left(\bar{\rho} \nabla \underbrace{\frac{\delta \mathcal{F}[\rho]}{\delta \rho}}\right) \right)$$

Pf: Construct a geodesic connecting $\bar{\rho}$ and ρ

$L = W_2(\rho, \bar{\rho})$



$\gamma \in OPT(\bar{\rho}, \rho)$

$$\pi_t(x, y) = \frac{L-t}{L} x + \frac{t}{L} y$$

$t \in [0, L]$

$$\hat{\rho}_t = (\pi_t)_\# \gamma$$

$\hat{\rho}_t$ is a unit-speed W_2-geodesic.

Let $\hat{\varphi} : [0, L] \to \mathbb{R}$ be the composition

$$\hat{\varphi}(t) := \mathcal{F}[\hat{\rho}(t)]$$

Since $\mathcal{F}$ is λ-convex, it follows that $\hat{\varphi}(\tau)$ is λ convex,

$$\sqrt{\frac{2}{\lambda}(\hat{\varphi}(L) - \hat{\varphi}(0))} \geq |L - 0|.$$

clearly.

How about

$$|\nabla \varphi(x)| \geq \sqrt{2\lambda(\varphi(x) - \varphi(x_0))}$$

$$|\nabla \hat{\varphi}(L)| \geq \sqrt{2\lambda(\hat{\varphi}(L) - \hat{\varphi}(0))}$$

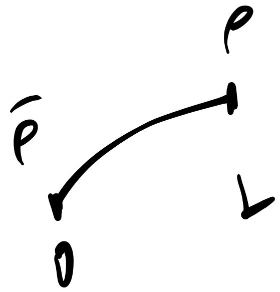
$$= \sqrt{2\lambda(\mathcal{F}(\rho) - \mathcal{F}(\bar{\rho}))}$$

i.e. $\mathcal{F}[\rho] - \mathcal{F}[\bar{\rho}] \leq \frac{1}{2\lambda} |\hat{\varphi}'(L)|^2$

$\hat{\varphi}(t) = \mathcal{F}[\hat{\rho}(t)]$.

$$\frac{d}{dt}\hat{\varphi}(t) = \frac{d}{dt} \mathcal{F}[\hat{\rho}(t)]$$

$$= \int \frac{\delta \mathcal{F}[\hat{\rho}]}{\delta \rho} \cdot \underbrace{\partial_t \hat{\rho}(t)} \, dx$$



$$= -\int \frac{\delta \mathcal{F}[\hat{\rho}]}{\delta \rho} \, \operatorname{div}(\hat{\rho}(t) \cdot v) \, dx$$

$$= \int \nabla \frac{\delta \mathcal{F}[\hat{\rho}]}{\delta \rho} \cdot v \;\; \hat{\rho}(t)$$

$$\leq \left(\int \left| \nabla \frac{\delta \mathcal{F}[\hat{\rho}]}{\delta \rho} \right|^2 \hat{\rho}(t) \, dt \right)^{1/2} \left(\int \underbrace{v^2 \, d\hat{\rho}(t)}_{=1} \right)^{1/2}$$

Okay $\square$

Move back to

$$\mathcal{F}[\rho] = \int_{\mathbb{R}^d} \big(\underbrace{\rho \log \rho}_{\text{convex}} + V \rho \big) \, dx$$

Imposes V : λ-convex

Prop: Assume $V: \mathbb{R}^d \to \mathbb{R}$ is λ-convex,

then $\mathcal{F}[\rho] = \int_{\mathbb{R}^d} \rho \log \rho + \int_{\mathbb{R}^d} \rho V$

is λ-convex.

If $\lambda > 0$, then one has.

* $W_2^2(\rho, \bar{\rho}) \leq \frac{2}{\lambda}\left(\mathcal{F}[\rho] - \mathcal{F}[\bar{\rho}]\right)$

$\bar{\rho}$ ↑ global minimizer

* $\mathcal{F}[\rho] - \mathcal{F}[\bar{\rho}] \leq \frac{1}{2\lambda} \langle \mathrm{grad}_{W_2} \mathcal{F}[\rho], \mathrm{grad}_{W_2} \mathcal{F}[\rho] \rangle_\rho$.

$\Downarrow$

In particular case, this gives us a log-Sobolev inequality

Assume $V: \mathbb{R}^d \to \mathbb{R}$ with $\lambda > 0$.

$$\int_{\mathbb{R}^d} e^{-V(x)} dx = 1.$$

Given $\eta: \mathbb{R}^d \to [0, \infty)$ s.t. $\int_{\mathbb{R}^d} \eta e^{-V} dx = 1$

Then $\rho = \eta e^{-V} \in \mathcal{P}(\mathbb{R}^d)$

$$\mathcal{F}[\rho] = \int_{\mathbb{R}^d} \eta \log \eta \, e^{-V} dx \quad \left(= \int \rho \log \rho + \int \rho V \right)$$

$$\leq \frac{1}{2\lambda} \langle \operatorname{grad}_{W_2} \mathcal{F}[\rho], \operatorname{grad}_{W_2} \mathcal{F}[\rho] \rangle_\rho$$

$$= \frac{1}{2\lambda} \int_{\mathbb{R}^d} \frac{|\nabla \eta|^2}{\eta} e^{-V} dx.$$

i.e.

$$\boxed{\int_{\mathbb{R}^d} \eta \log \eta \, e^{-V} dx \leq \frac{1}{2\lambda} \int_{\mathbb{R}^d} \frac{|\nabla \eta|^2}{\eta} e^{-V} dx, \qquad \forall \eta : \mathbb{R}^d \to [0, \infty), \ \int \eta e^{-V} dx = 1}$$

Log-Sobolev

Convergence to equilibrium

Say λ-convex function $V : \mathbb{R}^d \to \mathbb{R}$

with $\lambda > 0$ and $\int_{\mathbb{R}^d} e^{-V} dx = 1$.

The linear Fokker-Planck

$$\begin{cases} \partial_t \rho = \Delta \rho + \operatorname{div}(\rho \nabla V) \\ \rho(0) = \bar{\rho} \in \mathcal{P}_2(\mathbb{R}^d) \end{cases}$$

is the Wasserstein Gradient Flow associated to F

$$F[\rho] = \int_{\mathbb{R}^d} (\rho \log \rho + \rho V)\, dx.$$

$\boxed{\text{What is the behavior of } \rho_t := \rho(t) \text{ as } t \to \infty\text{ ?}}$

$$\begin{aligned}
\frac{d}{dt} F[\rho_t] &= \langle \mathrm{grad}_{W_2} F[\rho_t], \partial_t \rho_t \rangle_{W_2} \\
&= -\underbrace{\langle \mathrm{grad}_{W_2} F[\rho_t], \mathrm{grad}_{W_2} F[\rho_t] \rangle_{W_2}}_{\overset{LSI}{\le} -2\lambda F[\rho_t].} \\
&\le 0
\end{aligned}$$

$$F[\rho_t] \to F[e^{-V}] = 0. \quad \text{as } t \to \infty.$$

$$\frac{d}{dt} F[\rho_t] \le -2\lambda F[\rho_t],$$

Thus: $$F[\rho_t] \le e^{-2\lambda t} F[\bar{\rho}_0].$$ initial data

Also

$$\begin{aligned}
W_2^2(\rho_t, \underset{\underset{e^{-V}}{\shortparallel}}{\bar{\rho}}) &\le \frac{2}{\lambda}\big(F[\rho_t] - \underset{\underset{0}{\shortparallel}}{F[\bar{\rho}]}\big) \\
&\le \frac{2}{\lambda} F[\bar{\rho}]\, e^{-2\lambda t}, \quad \forall t \ge 0.
\end{aligned}$$

Rk: Since the functional is λ-convex,

a stronger estimate holds:

$$W_2^2(\rho_t, e^{-V}) \le e^{-2\lambda t} W_2^2(\bar{\rho}, e^{-V}).$$

More generally,

say $\rho, \tilde{\rho} : [0, \infty) \to \mathcal{P}_2(\mathbb{R}^d)$ are two gradient flow w.r.t. the functional $\mathcal{F}$, then

$$W_2^2(\rho_t, \tilde{\rho}_t) \le e^{-2\lambda t} W_2^2(\rho_0, \tilde{\rho}_0)$$

(Contractivity)

Basic idea can be given as follows:

Take smooth λ-convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$,

and consider two curves

$$x, y : [0, +\infty) \to \mathbb{R}^d$$

that solve the gradient flow eq.

$$\begin{cases} \dot{x}(t) = -\nabla\varphi(x(t)) \\ \dot{y}(t) = -\nabla\varphi(y(t)) \end{cases}$$

Then

$$\frac{d}{dt}\frac{1}{2}|x(t)-y(t)|^2 = \langle x(t)-y(t), \dot{x}(t)-\dot{y}(t)\rangle$$

$$= -\langle x(t)-y(t), \nabla\varphi(x(t))-\nabla\varphi(y(t))\rangle$$

$$\leq -\lambda|x(t)-y(t)|^2$$

$$\frac{d}{dt}|x(t)-y(t)|^2 + 2\lambda|x(t)-y(t)|^2 \leq 0$$

$$\frac{d}{dt}\left(e^{2\lambda t}|x(t)-y(t)|^2\right) \leq 0$$

$$\Rightarrow |x(t)-y(t)|^2 \leq e^{-2\lambda t}|x(0)-y(0)|^2$$

or $$|x(t)-y(t)| \leq e^{-\lambda t}|x(0)-y(0)|.$$

<u>Remark:</u> Csiszár - Kullback - Pinsker :

$$\frac{1}{2}\|\rho - f\|_{L^1}^2 \leq H(\rho|f)$$
$$= \int \rho \log \frac{\rho}{f}\, dx$$

or $$\|\rho - f\|_{L^1} \leq \sqrt{2\int \rho \log \frac{\rho}{f}}\,.$$

Hence if (ρ_t) solves the linear FPE:

$$\partial_t \rho = \Delta \rho + \operatorname{div}(\rho \nabla V),$$

then
$$\|\rho_t - e^{-V}\|_{L^1}^2 \leq 2F[\rho_t]$$
$$\leq 2e^{-2\lambda t} F[\underset{\substack{\uparrow \\ \text{initial data}}}{\rho_0}].$$

We note that $\sqrt{2}$ in CKP inequality is optimal.
Indeed, if $\rho = \eta e^{-V} \in \mathcal{P}(\mathbb{R}^d)$,
$$\eta : \mathbb{R}^d \to \{1 \pm \varepsilon\} \text{ everywhere},$$
then

$$\mathcal{F}[\rho] = \int_{\mathbb{R}^d} \eta \log \eta \, e^{-V}$$
$$= \int_{\mathbb{R}^d} \left((\eta - 1) + \frac{1}{2} (\eta-1)^2 + O(|\eta-1|^3) \right) e^{-V}$$
$$= \frac{1}{2} \varepsilon^2 + O(\varepsilon^3) = \frac{1+O(\varepsilon)}{2} \| \rho - e^{-V} \|_{L^1}^2 .$$

$$\Big(\varphi(x) = x \log x \qquad \varphi'(x) = \log x + 1 \qquad \varphi''(x) = 1/x$$
$$\varphi(x) = \varphi(1) + \varphi'(1)(x-1) + \frac{1}{2} \varphi''(1)(x-1)^2 + O\big((x-1)^3\big)$$
$$\varphi(\eta) = (\eta-1) + \frac{1}{2}(\eta-1)^2 + O(|\eta-1|^3) \Big)$$

See detailed proof in Figalli & Glaudo

Okay.

Another example:

Wasserstein Gradient Flow arising from Neural Networks:

Mei --- PNAS

Mean-Field analysis of 2-layer NN.

1 Problem A

$D \subseteq \mathbb{R}^n$ $\quad f: D \to \mathbb{R}$. $\quad f$

Given a domain D in $\mathbb{R}^n$ and a function $f : D \to \mathbb{R}$. We train a two-layer neural network that approximates f with the form

$\theta_i := (a_i, b_i, w_i)$

$$f_N(x, \underbrace{(a_i, b_i, \omega_i)_{1 \le i \le N}}_{\text{parameter}}) = \frac{1}{N} \sum_{i=1}^{N} a_i \sigma(\omega_i \cdot x + b_i) = \frac{1}{N} \sum_{i=1}^{N} h(\theta_i, x),$$

where $\theta_i = (a_i, b_i, \omega_i) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^n = \mathbb{R}^d$ are parameters to be optimized and σ is an activation function. The parameters (θ_i) are updated by minimizing the following generalization error

Loss function

$$E(f, f_N) = \frac{1}{2} \int_D \left| f(x) - f_N(x, (\theta_i)_{1 \le i \le N}) \right|^2 \mathrm{d}x.$$

$\theta_i \sim \rho$

$\theta_i \sim \rho$

i) Suppose that $\theta_1, \theta_2, \cdots, \theta_N$ are independent and identically distributed with a common distribution ρ. Give some assumptions and show that as $N \to \infty$,

$$f_N(x, \mu_N) \to f_\rho(x) := \int_{\mathbb{R}^d} h(\theta, x) \rho(\theta) \, \mathrm{d}\theta,$$

$f_N = \underbrace{\frac{1}{N} \sum_i h(\theta_i, x)}$

$\downarrow$ as $N \to \infty$

$\int h(\theta, x) \rho(\theta) \, d\theta$

Loss

and

$$E(f, f_N) \to \underline{E(f, f_\rho)} := \frac{1}{2} \int_D |f(x) - f_\rho(x)|^2 \, \mathrm{d}x.$$

$f_N(x, (\theta_i)_i)$

$= \frac{1}{N} \sum_i h(\theta_i, x)$

$\downarrow$

$f_\rho(x)$

ii) Show that $E(f, f_\rho)$ can be expanded as

$$E(f, f_\rho) = \int_{\mathbb{R}^d \times \mathbb{R}^d} K(\theta, \theta') \rho(\mathrm{d}\theta) \rho(\mathrm{d}\theta') + \int_{\mathbb{R}^d} V(\theta) \rho(\theta) + C_f.$$

(continuum problem)

Give the exact formulas for the functions K, V and the constant C_f.

iii) Consider the minimization problem with the entropy regularization

$$\min_{\rho \in \mathcal{P}(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)} \Big(\mathcal{F}_\sigma(\rho) := E(f, f_\rho) + \sigma \underline{\int_{\mathbb{R}^d} \rho(x) \log \rho(x) \, \mathrm{d}x} \Big),$$

$E(f, f_\rho)$

$= H(\rho_\theta)$

$\theta \in \mathbb{R}^d$

where $\mathcal{P}(\mathbb{R}^d)$ denotes the space of probabilities on $\mathbb{R}^d$ and $\sigma > 0$ is a small parameter.

Show that the gradient flow in the probability space, that is

$$\partial_t \rho_t = \mathrm{div}_\theta \Big(\rho_t \nabla_\theta \frac{\delta \mathcal{F}_\sigma}{\delta \rho_t} \Big),$$

has the form

$$\boxed{\partial_t \rho_t = \mathrm{div}_\theta \Big(\rho_t \Big(\nabla_\theta V + \int_{\mathbb{R}^d} \nabla_\theta K(\theta, \theta') \rho_t(\mathrm{d}\theta') \Big) \Big) + \sigma \Delta_\theta \rho_t.} \tag{1.1}$$

$E(f, f_\rho)$

$= \frac{1}{2} \int_D |f(x) - f_\rho(x)|^2 \, dx,$ where $f_\rho = \int h(\theta, x) \rho(\theta) \, d\theta$

$$= \underbrace{\frac{1}{2} \int_D |f(x)|^2}_{= C_f} - \int_D f(x) f_\rho(x) \, dx + \underbrace{\frac{1}{2} \int_D |f_\rho(x)|^2 \, dx}_{\text{interaction}}$$

$$-\int_D f(x) \int_{\mathbb{R}^d} h(\theta, x)\rho(\theta)\, d\theta$$

$$= \int_{\mathbb{R}^d} d\theta\, \rho(\theta) \boxed{\int_D f(x)\, h(\theta, x)\, dx} = V(\theta)$$

McKean-Vlasov PDE

iv) By taking $V = 0$, and $\nabla_\theta K(\theta, \theta') = F(\theta - \theta')$ and $\sigma = 1$ above, Eq. (1.1) becomes

$$\partial_t \rho_t = \operatorname{div}\Big(\rho_t F * \rho_t\Big) + \Delta \rho_t.$$

(nonlinear Fokker-Planck aggregation-diffusion eq.)

Assume that $F \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ is anti-symmetric and Lipschitz, i.e. $F(\theta) = -F(-\theta)$ and $\|\nabla F\|_{L^\infty} \leq L$. Show that starting from two integrable initial data ρ_1^0 and $\rho_2^0 \in \mathcal{P}_2(\mathbb{R}^d)$, the corresponding two solutions $\rho_1 = \rho_1(t, \theta)$ and $\rho_2 = \rho_2(t, \theta)$ with $(t, \theta) \in \mathbb{R}_+ \times \mathbb{R}^d$ satisfy that

$\partial_t \rho_t = \operatorname{div}(\rho_t F * \rho_t)$

$$\mathcal{W}_2(\rho_1(t), \rho_2(t)) \leq \mathcal{W}_2(\rho_1^0, \rho_2^0) \exp(2Lt).$$

Lt

$W_p(\rho_1(t), \rho_2(t))$ $\mu_1(t)$

$\leq W_p(\rho_1(0), \rho_2(0))\, e^{2Lt}$. $\mu_1(0)$

$(t > 0)$

$\Downarrow$

N-particle approximation estimate (Mean-Field Limit)

Note: We write $\mathcal{P}_2(\mathbb{R}^d)$ as the space of probability measures on $\mathbb{R}^d$ with finite 2−th moments. Then the Wasserstein-2 distance in $\mathcal{P}_2(\mathbb{R}^d)$ is defined as the following

$$\mathcal{W}_2(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\ \nu)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \, \mathrm{d}\pi(x, y) \right)^{1/2} = \inf_{X \sim \mu, Y \sim \nu} \Big(\mathbb{E}[|X - Y|^2] \Big)^{1/2},$$

where $\Pi(\mu, \nu)$ is the set of joint probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν respectively and (X, Y) are all possible couplings of random variables with μ and ν as their marginal laws respectively.

Solutions of Problem A i) Using the empirical measure $\mu_N = \frac{1}{N}\sum_i \delta_{\theta_i}$, the leaning function f_N reads

$$f_N(x, \mu_N) := f_N(x, (\theta_i)_{1 \leq i \leq N}) = \int_{\mathbb{R}^d} h(\theta, x) \mu_N(\,\mathrm{d}\theta).$$

If $h \in C_b$ and θ_i are i.i.d. with the common law ρ, then by Law of Large Numbers,

$$f_N(x, \mu_N) \to f_\rho(x) := \int_{\mathbb{R}^d} h(\theta, x)\rho(\theta)\,\mathrm{d}\theta, \quad \text{as } N \to \infty.$$

Hence the generalization error converges to

- Boltzmann (Villani)

$$E(f, f_\rho) := \frac{1}{2}\int_D |f(x) - f_\rho(x)|^2 \,\mathrm{d}x.$$

- Vlasov-Poisson (Loeper 2006)

ii)By completing the square and then using Fubini

$$E(f, f_\rho) = \frac{1}{2}\int_D |f(x)|^2 \,\mathrm{d}x + \int_{\mathbb{R}^d} \Big(\int_D -f(x)h(\theta, x)\,\mathrm{d}x \Big) \rho(\,\mathrm{d}\theta)$$
$$+ \int_{\mathbb{R}^d \times \mathbb{R}^d} \Big(\frac{1}{2}\int_D h(\theta, x)h(\theta', x)\,\mathrm{d}x \Big) \rho(\,\mathrm{d}\theta)\rho(\,\mathrm{d}\theta')$$

W_p - distance as a metric to do some stability estimate in PDE.

$\Updownarrow$

Mean-Field Limit / N-particle approximation.

Propagation of Chaos

$$\partial_t \rho = \nabla\cdot(\rho F * \rho) \qquad W_1(\rho(t), \mu_N(t)) \le e^{2\|\nabla F\|_{L^\infty} t}\, \underbrace{W_1(\rho(0), \mu_N(0))}_{\sim \frac{1}{N^\theta}}$$

$$\dot{X}_i = \frac{1}{N}\sum_{j=1}^N F(X_i - X_j), \quad i = 1, 2, \dots, N$$

$$\mu_N(t) = \frac{1}{N}\sum_{i=1}^N \delta_{X_i(t)}$$

$X_i(0) \sim$ i.i.d. $\rho(0)$

We thus obtain the formula

$$H(\rho) = E(f, f_\rho) = \underbrace{\int_{\mathbb{R}^d\times\mathbb{R}^d} K(\theta,\theta')\rho(\mathrm{d}\theta)\rho(\mathrm{d}\theta')}_{\text{interaction energy}} + \underbrace{\int_{\mathbb{R}^d} V(\theta)\rho(\theta)}_{\text{potential energy}} + \cancel{C_f},$$

with the definitions

$$K(\theta,\theta') = \frac{1}{2}\int_D h(\theta,x)h(\theta',x)\,\mathrm{d}x, \quad V(\theta) = \int_D -f(x)h(\theta,x)\,\mathrm{d}x.$$

iii) Consider the minimization problem with the entropy regularization

$$\min_{\rho\in\mathcal{P}(\mathbb{R}^d)\cap L^1(\mathbb{R}^d)} \Big(E(f,f_\rho) + \sigma \underbrace{\int_{\mathbb{R}^d}\rho(x)\log\rho(x)\,\mathrm{d}x}_{\text{entropy}}\Big),$$

where $\mathcal{P}(\mathbb{R}^d)$ denotes the space of probabilities on $\mathbb{R}^d$ and $\sigma > 0$ is a small parameter. Now the functional reads

Free entropy

$$\mathcal{F}_\sigma(\rho) = \int K(\theta,\theta')\rho(\mathrm{d}\theta)\rho(\mathrm{d}\theta') + \int V(\theta)\mu(\mathrm{d}\theta) + \sigma\int \rho(\theta)\log\rho(\theta)(\mathrm{d}\theta).$$

Hence the 1st variation of F_σ reads W_2 - gradient.

$$\frac{\delta\mathcal{F}_\sigma}{\delta\rho}(\rho) = 2\int_{\mathbb{R}^d} K(\theta,\theta')\rho(\mathrm{d}\theta') + V(\theta) + \sigma\log\rho(\theta).$$

Hence the gradient flow PDE which given by

$$v = -\nabla\frac{\delta F}{\delta\rho}$$

$$\partial_t\rho_t = \mathrm{div}_\theta\Big(\rho_t\nabla_\theta\frac{\delta\mathcal{F}_\sigma}{\delta\rho_t}\Big),$$

L^2 - 1st variation

now becomes

$$\partial_t\rho_t = \mathrm{div}_\theta\Big(\rho_t\Big(\nabla_\theta V + \int_{\mathbb{R}^d}\nabla_\theta K(\theta,\theta')\rho_t(\mathrm{d}\theta')\Big)\Big) + \sigma\Delta_\theta\rho_t. \quad \bullet\ F = -\nabla W$$

iv) By taking $V = 0$, and $\nabla_\theta K(\theta,\theta') = F(\theta-\theta')$ and $\sigma = 0$, one arrives a simple PDE

interaction force

McKean–Vlasov PDE. $v = F * \rho_t$

$$\partial_t\rho_t = \mathrm{div}_\theta(\rho_t F * \rho_t) + \Delta\rho_t.$$

aggregation-diffusion

We now proceed to show that

$F \in W^{1,\infty}$, $F(0) = 0$

global Lipschitz.

$\{|F(x)| \le C(1+|x|)\}$

$$W_2(\rho_1(t), \rho_2(t)) \le W_2(\rho_1^0, \rho_2^0)\exp(Lt).$$

We construct two stochastic processes X_t and Y_t such that

Itô

$$\mathrm{d}X_t = -F * \rho_t^1(X_t) + \sqrt{2}\,\mathrm{d}W_t, \quad X_0 \sim \rho_1(0),$$

$X_t \sim \rho_t^1$ ✓

$\boxed{\text{self-consistent}}$

or Distribution Dependent

$$L = \|F\|_{Lip} = \sup_{x\neq y}\frac{|F(x)-F(y)|}{|x-y|}$$

$F = \nabla W$

$$\mathcal{W}_2^2(\rho_1(t), \rho_2(t)) \le \boxed{\mathbb{E}|X_t - Y_t|^2}$$

$X_t \sim \rho_1(t)$

$Y_t \sim \rho_2(t)$

and

$$\mathrm{d}Y_t = -F * \rho_t^2(Y_t) + \sqrt{2}\,\mathrm{d}W_t, \quad Y_0 \sim \rho_2(0),$$

where X_t and Y_t are driven by the same standard Brownian motion and of course $X_t \sim \rho_1(t)$ and $Y_t \sim \rho_2(t)$. Let us choose a particular initial coupling such that

optimal coupling

$$\mathbb{E}|X_0 - Y_0|^2 = \left(\mathcal{W}_2(\rho_1^0, \rho_2^0)\right)^2.$$

We write that the law of coupling (X_t, Y_t) as $\pi_t \in \Pi(\rho_1(t), \rho_2(t))$, hence by definition

$$\mathcal{W}_2^2(\rho_1(t), \rho_2(t)) \le \mathbb{E}|X_t - Y_t|^2.$$

study the evolution of $\mathbb{E}|X_t - Y_t|^2$

By Ito's formula,

$$\mathrm{d}\left(\frac{1}{2}|X_t - Y_t|^2\right) = (X_t - Y_t) \cdot (\mathrm{d}X_t - \mathrm{d}Y_t).$$

$$\mathrm{d}X_t - \mathrm{d}Y_t = -F * \rho_t^1(X_t) - F * \rho_t^2(Y_t)$$

Note that

$$\mathrm{d}X_t - \mathrm{d}Y_t = \int_{\mathbb{R}^d \times \mathbb{R}^d} \Big(F(X_t - x') - F(Y_t - y')\Big)\pi_t(\mathrm{d}x'\,\mathrm{d}y').$$

Combining the above two formulas and taking expectations on both sides, one arrives that

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \pi_t(\mathrm{d}x\,\mathrm{d}y) = \frac{1}{2}\mathbb{E}|X_t - Y_t|^2$$

$$= -\int_{\mathbb{R}^d \times \mathbb{R}^d}\int_{\mathbb{R}^d \times \mathbb{R}^d} (x - y) \cdot \Big(F(x - x') - F(y - y')\Big)\pi_t(\mathrm{d}x'\,\mathrm{d}y')\pi_t(\mathrm{d}x\,\mathrm{d}y).$$

By the fact that F is anti-symmetric, it equals to

$F(x) = -F(-x)$

F is odd

$$-\frac{1}{2}\int_{\mathbb{R}^d \times \mathbb{R}^d}\int_{\mathbb{R}^d \times \mathbb{R}^d} (x - y - x' + y') \cdot \Big(F(x - x') - F(y - y')\Big)\pi_t(\mathrm{d}x'\,\mathrm{d}y')\pi_t(\mathrm{d}x\,\mathrm{d}y),$$

$$((x-x') - (y-y')) \cdot (F(x-x') - F(y-y'))$$

which can be bounded by

$$\frac{L}{2}\int\int |x - y - (x' - y')|^2 \pi_t(\mathrm{d}x\,\mathrm{d}y)\pi_t(\mathrm{d}x'\,\mathrm{d}y') \le L|x-y-(x'-y')|^2$$

$$\le 2L\int |x - y|^2 \pi_t(\mathrm{d}x\,\mathrm{d}y)$$

$$|a+b|^2 \le 2a^2 + 2b^2$$

by the Cauchy-Schwarz inequality. Now we can conclude the proof by Gronwall's lemma.

Furthermore, if $F = \nabla W$, W is λ-convex, $(\lambda > 0)$

$$\underbrace{\frac{d}{dt}\frac{1}{2}\int |x-y|^2 \pi_t(dx\,dy)}_{Q(t)} = -\frac{1}{2}\int\int \big((x-x') - (y-y')\big)\big(\nabla W(x-x') - \nabla W(y-y')\big)\pi_t^{\otimes 2}$$

$$\le -\frac{1}{2}\iint \lambda\,|(x-x') - (y-y')|^2 \pi_t^{\otimes 2}$$

$$= -\frac{\lambda}{2}\iint \left|(x-y)-(x'-y')\right|^2 \pi_t(dx\,dy)\,\pi_t(dx'\,dy')$$

$$= -\lambda \int |x-y|^2 \pi_t(dx, dy)$$

$$\Rightarrow \quad Q'(t) \leq -2\lambda\, Q(t)$$

$$\Rightarrow \quad Q(t) \leq e^{-2\lambda t}\, Q(0)$$

$$\mathbb{E}|X_t - Y_t|^2 \leq e^{-2\lambda t}\, \underbrace{\mathbb{E}|X_0 - Y_0|^2}_{W_2(\rho_1^0, \rho_2^0)^2}$$

$$\Leftarrow \quad \boxed{W_2(\rho_1^t, \rho_2^t) \leq e^{-\lambda t}\, W_2(\rho_1^0, \rho_2^0)}$$

$$\left(2\iint |x-y|^2 \pi_t - 2\int (x-y)\cdot(x'-y')\,\pi_t(dx\,dy)\,\pi_t(dx', dy')\right)$$

$$= \left(2\iint |x-y|^2 \pi_t - 2\underbrace{\left(\int (x-y)\,\pi_t(dx\,dy)\right)}_{\parallel}{}^2\right)$$

0

We only need

$$\begin{cases} \int x \, d\rho_1^t(x) = \int y \, d\rho_2^t(y) = c \\ \Uparrow \\ F(x) = -F(-x) \end{cases}$$

Conclusion:

For $F = \nabla W$, W is λ-convex, $(\lambda > 0)$

$F(x) = -F(-x)$ (or W is even)

then for the aggregation - diffusion PDE:

$$\partial_t \rho = \nabla \cdot (\rho \nabla W * \rho) + \sigma \Delta \rho$$

has the following stability estimate:

$$W_2(\rho_1^t, \rho_2^t) \le \exp(-\lambda t) \, W_2(\rho_1^0, \rho_2^0)$$

where $\int_{\mathbb{R}^d} x \rho_1^0(x) \, dx = \int_{\mathbb{R}^d} x \rho_2^0(x) \, dx = \text{const.}$

Reference: Optimal Transport, Old and New

Chapter $\underline{\underline{6}}$ page 93 – 106

Recall the notions

$$\underset{\Uparrow}{C(\mu,\nu)} = \inf_{\pi\in\Pi(\mu,\nu)} \int \underset{\substack{\uparrow \\ \text{cost function}}}{c(x,y)}\, d\pi(x,y),$$

The value.

Def: (Wasserstein Distance)

Let $(\mathcal{X}, d)$ be a Polish metric space, and (complete separable)

let $p\in[1,\infty)$. For any $\mu, \nu \in P(\mathcal{X})$, the
($p=\infty$ don't consider this case)

Wasserstein distance of order p between μ and

ν is defined as

$$W_p(\mu,\nu) = \left(\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}} \underbrace{d(x,y)^p}_{= c(x,y)}\, d\pi(x,y) \right)^{1/p}$$

$$= \inf\left\{ \left[\underset{\substack{X\sim\mu \\ Y\sim\nu}}{E}\, d(X,Y)^p \right]^{1/p} \right\}$$

$\mathrm{Law}(X) = \mu$

$\mathrm{Law}(Y) = \nu$

($p=1$ $\underbrace{W_1}$ Kantorovich – Rubinstein distance)

Trivially, $W_p(\delta_x, \delta_y) = d(x,y)$ $\forall p\geq 1$.

$x\in X \mapsto \delta_x$ isometry

Proof that W_p satisfies the axioms of a distance (metric): $(\mathcal{P}_p(\mathcal{X}), W_p)$

- Clearly $W_p(\nu,\mu) = W_p(\mu,\nu)$ (Symmetry)
- Next, let $\mu_1, \mu_2, \mu_3 \in \mathcal{P}(X)$, $W_p(\mu_1,\mu_3) \leq W_p(\mu_1,\mu_2) + W_p(\mu_2,\mu_3)$

let (X_1, X_2) be an optimal coupling of (μ_1,μ_2),
(Z_2, Z_3) an optimal coupling of (μ_2,μ_3) $(c = d^p)$

By the Gluing lemma, (Villani Chapter 1)

X_1 X_2
$(\mu_1,$ $\mu_2)$
Z_2 . Z_3
μ_2 μ_3

$\exists$ random variables (X_1', X_2', X_3') with

$$\begin{cases} \text{Law}(X_1', X_2') = \text{law}(X_1, X_2) \\ \text{law}(X_2', X_3') = \text{law}(Z_2, Z_3). \end{cases}$$

In particular, (X_1', X_3') is a coupling of (μ_1,μ_3),

so

$$W_p(\mu_1,\mu_3) \leq \left(\mathbb{E}\, d(X_1', X_3')^p\right)^{1/p}$$

$(\|g_1+g_2\|_{L^p} \leq \|g_1\|_{L^p} + \|g_2\|_{L^p})$

$$\leq \left(\mathbb{E}\left(d(X_1', X_2') + d(X_2', X_3')\right)^p\right)^{1/p}$$

Minkowski in $L^p(\mathbb{P})$

$$\leq \left(\mathbb{E}\, d(X_1', X_2')^p\right)^{1/p} + \left(\mathbb{E}\, d(X_2', X_3')^p\right)^{1/p}$$

$$\leq W_p(\mu_1,\mu_2) + W_p(\mu_2,\mu_3)$$

- Finally, assume that $W_p(\mu,\nu) = 0$, $\Rightarrow \mu = \nu$

then $\exists$ a transference plan which is

$W_p(\mu,\nu) = \int_{X\times X} d^p(x,y)\, d\pi^*(x,y) = 0$, $\pi^* \in OPT(\mu,\nu)$

concentrated on the diagonal

$\{(\underbrace{x, x}_{\boxed{y=x}}) \mid x \in \mathcal{X}\} \subseteq \mathcal{X} \times \mathcal{X}$. So $\nu = Id_{\#}\mu = \mu$.

To ensure that $\boxed{W_p(\mu, \nu)} < +\infty$,

it is natural to restrict W_p to a proper subset of $\boxed{\mathcal{P}(X)}$, usually $\boxed{\mathcal{P}_p(X)}$:

$$\mathcal{P}_p(\mathcal{X}) := \left\{ \mu \in \mathcal{P}(\mathcal{X}) \,\Big|\, \int_{\mathcal{X}} d(x_0, x)^p \mu(dx) < +\infty \right\}$$

$x_0 = 0$ $\int_{\mathbb{R}^d} |x|^p d\mu(x) < +\infty$

$x_0 \in X$ is arbitrary

This space does not depend on x_0.

Then W_p defines a (finite) distance on $\mathcal{P}_p(\mathcal{X})$.

$(\Leftarrow$ $W_p^p(\mu, \nu) \leq \int d(x, y)^p \, d\pi(x, y)$. $(a+b)^p \leq 2^{p-1}(a^p + b^p)$ $a, b \geq 0$

$$\leq 2^{p-1} \int \left[d(x, x_0)^p + d(x_0, y)^p \right] d\pi(x, y)$$

$< \infty$. $\quad " \; 2^{p-1} \Big[\int d(x, x_0)^p d\mu(x)$

$+ \int d(x_0, y)^p \, d\nu(y) \Big]$

Remark:

Very useful duality formula for W_1: W_p

For $\mu, \nu \in \mathcal{P}_1(X)$, $W_1, W_2, \boxed{W_\infty}$

$\boxed{\text{dual}}$ $\int \varphi \, d\mu + \int \varphi^c \, d\nu$

is ... d. ormal

For ψ Lip

$$W_1(\mu,\nu) = \sup_{\|\psi\|_{Lip}\le 1} \left\{ \int_X \psi \, d(\mu-\nu) \right\} \quad \left(\operatorname{ess} \sup_{(x,y)\in \operatorname{supp}(\pi)} |x-y| \right)$$

$$\left| \int \psi \, d\mu - \int \psi \, d\nu \right| \le \|\nabla \psi\|_{L^\infty} W_1(\mu,\nu)$$

An applications: if $\int f \, d\mu = 1$, $f \ge 0$

i.e. f is a prob. density w.r.t. μ.

$$\Big(\underbrace{\int f \, d\mu}_{=1}\Big)\Big(\int g \, d\mu\Big) - \int (fg) \, d\mu \quad \text{(correlation function)}$$

$$= \underbrace{\int g \, d(\mu - f\mu)}$$

$$\le \|g\|_{Lip} W_1(f\mu, \mu)$$

Remark: Hölder inequality $\Rightarrow$ $L^\infty(d\mu) \cdots \subseteq L^p(d\mu) \subseteq \cdots \subseteq L^1(d\mu)$ (μ is a probability measure)

$1 \le p \le q < \infty \Rightarrow W_p(\mu,\nu) \le W_q(\mu,\nu)$

$$\int 1\cdot d(x,y)^p \, d\pi(x,y) \le \left(\int 1^* \, d\pi\right)^{\frac{1}{*}} \left(\int \left(d^p\right)^{\frac{q}{p}}(x,y) \, d\pi\right)^{\frac{p}{q}}$$

$$\left[\int d^p(x,y) \, d\pi(x,y)\right]^{1/p} \le \left[\int d^q(x,y) \, d\pi(x,y)\right]^{1/q}.$$

$$W_1 \le W_p \le W_q \le W_\infty \qquad 1 \le p \le q \le \infty$$

W_1, W_2, W_∞ $\rightsquigarrow$ for instance in Hwang–Jabi

W_1: duality formula; W_2: geometric feature

$$\left|\int fg \, d\mu\right| \le \left(\int f^p \, d\mu\right)^{1/p} \left(\int g^q \, d\mu\right)^{1/q}$$

$1/p + 1/q \le 1$, if μ

is a probability

Convergence in Wasserstein sense

Characterization $W_p(\mu_k, \mu) \to 0$

Narrow convergence: (in Villani: weak convergence)

$\Downarrow$

$\mu_k \to \mu$ (μ_k converges weakly to μ)

(i.e.) $\int \varphi \, d\mu_k \to \int \varphi \, d\mu \quad \forall \varphi \in C_b(X)$

finite p-th moment

Def: (Weak convergence in $\mathcal{P}_p$).

Let (X,d) be a Polish space, $p \in [1, \infty)$.

Let $(\mu_k)_{k \in \mathbb{N}} \subset \mathcal{P}_p(X)$, $\mu \in \mathcal{P}_p(X)$.

Then (μ_k) is said to converge weakly in $\mathcal{P}_p(X)$

if any one of the following equivalent properties is satisfied for some (and then any) $x_0 \in X$,

i) $\mu_k \to \mu$ and $\int d(x_0, x)^p \, d\mu_k(x) \to \int d(x_0,x)^p \, d\mu(x)$,

(convergence of p-th moment)

ii) $\mu_k \to \mu$ and $\limsup_{k\to\infty} \int d(x_0,x)^p \, d\mu_k(x) \le \int d(x_0,x)^p \, d\mu(x)$;

$\liminf_{k\to\infty} \int d(x_0,x)^p \, d\mu_k(x) \overset{\text{l.s.c}}{\ge} \int d(x_0,x)^p \, d\mu(x)$

iii) $\mu_k \to \mu$ and $\lim_{R\to\infty} \limsup_{k\to\infty} \int_{d(x_0,x)\ge R} d(x_0,x)^p \, d\mu_k(x) = 0$;

$\mathcal{P}_p(X)$

automatically

iv) $\forall \varphi$ continuous, $\varphi \in C_b(\mathcal{X})$

$$|\varphi(x)| \le C(1 + d(x_0, x)^p), \quad C \in \mathbb{R}_+,$$

one has

$$\int \varphi(x)\, d\mu_k(x) \to \int \varphi(x)\, d\mu(x).$$

Thm (W_p metrizes $\mathcal{P}_p$)

Let $(\mathcal{X}, d)$ be a Polish space, and $p \in [1, \infty)$; then the Wasserstein distance W_p metrizes the weak convergence in $\mathcal{P}_p(\mathcal{X})$.

That is, if $(\mu_k) \subset \mathcal{P}_p(\mathcal{X})$, $\mu \in \mathcal{P}(\mathcal{X})$, then

$$\boxed{\begin{array}{l} \mu_k \to \mu \text{ weakly in } \mathcal{P}_p \\ \iff W_p(\mu_k, \mu) \to 0 \end{array}}$$

RK: Convergence in W_p $\Rightarrow$ Convergence of moments of order p. ✓

If $\mathcal{X}$ is locally compact length space ($\mathbb{R}^d$, $\mathbb{T}^d$, ...), then one has a stronger statement that the map $\mu \mapsto \left(\int d(x_0, x)^p \mu(dx)\right)^{1/p}$ is 1-Lip w.r.t. W_p

$\mathcal{P}_p(\mathcal{X})$

$$\left| \left(\int_{\mathcal{X}} d(x_0,x)^p \, d\mu(x)\right)^{1/p} - \left(\int_{\mathcal{X}} d(x_0,x)^p \, d\nu(x)\right)^{1/p} \right| \le \underbrace{W_p(\mu,\nu)}$$

How to prove it, say $\mathcal{X} = \mathbb{R}^d$?

Corollary: $(\mathcal{P}_p(\mathcal{X}), W_p)$ topology : weak convergence

If $(\mathcal{X}, d)$ is a Polish space, and $p \in [1, \infty)$.

Then W_p is continuous on $\mathcal{P}_p(\mathcal{X})$.

More explicitly, if μ_k (resp ν_k) converges to μ (resp. ν) weakly in $\mathcal{P}_p(\mathcal{X})$ as $k \to \infty$, then

$$\boxed{W_p(\mu_k, \nu_k) \to W_p(\mu, \nu)}.$$

(For usual convergence, one only obtains that

$$W_p(\mu,\nu) \le \boxed{\liminf_{k\to\infty}} W_p(\mu_k, \nu_k) \quad \boxed{\text{l.s.c}}$$

As we did for $\underbrace{C(\mu,\nu)}$ = $\inf\limits_{\pi \in \Pi(\mu,\nu)} \int c(x,y)\, d\pi(x,y)$

c. l.s.c. $c \ge 0$.

Corollary (Metrizability of weak topology).

Let $(\mathcal{X}, d)$ be a Polish space. If $\tilde{d}$ is a bounded distance inducing the same topology as d (such as $\tilde{d} = d/_{1+d}$), then the convergence in Wasserstein sense for the distance $\tilde{d}$ is $\tilde{d} \leq c$ equivalent to the usual weak convergence of probability measures in $\mathcal{P}(X)$.

$\int d(x_0, x)^p \, d\mu(x) \leq c^p$

$(\mathcal{P}_p(X), W_p) = (\mathcal{P}(X), W_p)$

Other ways to metrize weak convergence:

$\mu, \nu \in \mathcal{P}(X)$ or r.v.s X, Y with $\mathrm{law}(X) = \mu$, $\mathrm{law}(Y) = \nu$.

- Lévy–Prokhorov distance

$$d_p(\mu, \nu) = \inf\{\varepsilon > 0 : \exists X, Y, \inf \mathbb{P}[d(X,Y) > \varepsilon] \leq \varepsilon\}$$

- bounded Lipschitz distance: (compare to W_1)

$$d_{bL}(\mu, \nu) = \sup\left\{ \int \varphi \, d\mu - \int \varphi \, d\nu \;\middle|\; \|\varphi\|_\infty + \|\varphi\|_{Lip} \leq 1 \right\}$$

- the weak-* distance (on a locally compact metric space):

$$d_{w*}(\mu, \nu) = \sum_{k \in \mathbb{N}} 2^{-k} \left| \int \varphi_k \, d\mu - \int \varphi_k \, d\nu \right|,$$

where $(\varphi_k)_{k\in\mathbb{N}}$ is a dense sequence in $C_0(\mathcal{X})$;

- Toscani distance (on $\mathcal{P}_2(\mathbb{R}^n)$)

$$d_T(\mu,\nu) = \sup_{\xi\in\mathbb{R}^n\setminus\{0\}} \left(\frac{1}{|\xi|^2} \left| \int e^{-i x\cdot\xi}\, d(\mu(x) - \nu(x)) \right| \right)$$

(Need $\int x\, d\mu(x) = \int y\, d\nu(y)$

$$e^{-ix\cdot\xi} = 1 - ix\cdot\xi + \frac{1}{2}(-ix\cdot\xi)^2 + \cdots$$

Just comparing all moments ---)

$\sup_{\varphi} \left| \int \varphi\, d\rho_t - \int \varphi\, d g_t \right| \le \|\nabla\varphi\|_{L^\infty} W_1(\rho_t, g_t)$

So why bother with Wasserstein distance?

- Strong, take care large distance in $\mathcal{X}$.
- Convenient in problems where O.T. is involved.

  (stability in PDEs) $\quad W_1(\rho_1(t), \rho_2(t)) \le e^{-Lt} W_1(\rho_1(0), \rho_2(0))$

  $\begin{cases} \bullet \text{ Continuity Eq.} \\ \bullet \text{ Boltzmann Eq / Landau Eq.} \end{cases}$

- W_1 $\boxed{\text{duality form}}$
- $W_p^p(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int d^p(x,y)\, d\pi(x,y) \le \int d^p(x,y)\, d\tilde{\pi}(x,y)$

  any plan.

Easy to bound from above

f - C-Lip $\quad f_\# \mu \quad \|\nabla f\|_{L^\infty} \le C$

$$W_p^p(\mu, f_\#\mu) \le \underbrace{\int |x - f(x)|^p\, d\mu(x)}$$ one upper bound

$$W_p^p(f_\# \mu, f_\# \nu) \overset{=}{\le} \int |f(x)-f(y)|^p \, d\underbrace{\pi(x,y)}_{(f,f)_\# \pi}$$

choose $\pi \in OPT(\mu,\nu)$

OPTIMAL PLAN

$$\le C^p \underbrace{\int |x-y|^p \, d\pi(x,y)}_{\| \atop W_p^p(\mu,\nu)}$$

$\nu \overset{\pi}{\underset{\mu}{\#}} \xrightarrow{(f,f)} f_\#\nu \overset{\tilde\pi}{\underset{f_\#\mu}{\#}}$

i.e. $\mu \mapsto f_\# \mu$

is also C-Lipschitz

$\left(W_p(f_\#\mu, f_\#\nu) \le \|\nabla f\|_{L^\infty} W_p(\mu,\nu) \right)$

- Wasserstein distances (W_2) incorporate a lot of geometry of the space.

$\|f-g\|_{L^p} = \|f - \tau_t \circ g\|_{L^p}$

trivially: $x \mapsto \delta_x$ (isometric embedding of X into $\mathcal{P}_p(X)$)

$\boxed{W_p}$ $d\mu = f\,dx$, $d\nu = g\,dy$ on $\mathbb{R}^d$

$$\underbrace{\left(\int_{\mathbb{R}^d} |f(x)-g(x)|^p dx\right)^{1/p}}_{L^p} + \underbrace{\left(\int_{\mathbb{R}^n} |\nabla f(x) - \nabla g(x)|^p dx\right)^{1/p}}_{W^{1,p}}$$

Now we prove the theorem.

We shall use the following Lemma

Lemma (Cauchy sequences in W_p are tight)

Let X be a Polish space, let $p \ge 1$ and let $(\mu_k)_{k\in\mathbb{N}}$ be a Cauchy sequence in $(\mathcal{P}_p(X), W_p)$. Then (μ_k) is tight.

Pf of this lemma:

$(\mu_k)_{k\in\mathbb{N}}$ Cauchy Sequence in $\mathcal{P}_p(X)$,

i.e. $W_p(\mu_k,\mu_l)\to 0$ as $k,l\to\infty$

Then in particular $\forall k$,

$$\int d(x_0,x)^p \, d\mu_k(x) = W_p(\delta_{x_0},\mu_k)^p$$
$$\le \left(W_p(\delta_{x_0},\mu_1)+W_p(\mu_1,\mu_k)\right)^p$$
$$\le C < \infty.$$

Since $W_p \geq W_1$,

(μ_k) is also Cauchy in W_1 sense.

Take $\varepsilon>0$, let $N\in\mathbb{N}$ s.t.

$W_1(\mu_N,\mu_k)\le\varepsilon^2$, when $k\ge N$.

Then $\forall k\in\mathbb{N}$, $\exists j\in\{1,2,\cdots,N\}$, s.t.

$W_1(\mu_j,\mu_k) < \varepsilon^2$. ↑ ε-net.

($k\ge N$, choose $j=N$; otherwise choose $j=k$.)

Since the finite set $\{\mu_1,\cdots,\mu_N\}$ is tight,

$\exists$ cpt set K, st $\mu_j[X\setminus K] < \varepsilon$, $\forall j\in\{1,2,\cdots,N\}$.

By compactness, $\underbrace{K}_{\text{cpt}}$ can be covered by a finite # of small balls, say

$$K\subset B(x_1,\varepsilon)\cup\cdots\cup B(x_m,\varepsilon).$$

Now write

$$\mathcal{U} := B(x_1,\varepsilon) \cup \cdots \cup B(x_m,\varepsilon) ;$$

$$\mathcal{U}_\varepsilon := \{ x \in \mathcal{X} : d(x,\mathcal{U}) < \varepsilon \}$$

$$\subseteq B(x_1, 2\varepsilon) \cup \cdots \cup B(x_m, 2\varepsilon)$$

$$\phi(x) = \left(1 - \frac{\overset{>\varepsilon}{d(x,\mathcal{U})}}{\varepsilon} \right)_+$$

Note that $\quad 1_{\mathcal{U}} \le \phi \le 1_{\mathcal{U}_\varepsilon}$

and ϕ is $1/\varepsilon$ - Lipschitz. w.r.t d.

For $j \le N$, and any k,

$$\begin{aligned}
\mu_k[\mathcal{U}_\varepsilon] &\ge \int \phi \, d\mu_k \qquad (\text{since } 1_{\mathcal{U}_\varepsilon} \ge \phi) \\
&= \int \phi \, d\mu_j + \left(\int \phi \, d\mu_k - \int \phi \, d\mu_j \right) \\
&\ge \int \phi \, d\mu_j - \frac{1}{\varepsilon} W_1(\mu_k, \mu_j) \\
&\ge \mu_j(\mathcal{U}) - \varepsilon
\end{aligned}$$

$$\Rightarrow \mu_k[\mathcal{U}_\varepsilon] \ge 1 - \varepsilon - \varepsilon = 1 - 2\varepsilon .$$

We have showed that: $\forall \varepsilon > 0$, $\exists$ finite family $(x_i)_{1 \le i \le m}$ s.t. all measures μ_k give mass at least $1 - 2\varepsilon$ to the set $Z := \bigcup B(x_i, 2\varepsilon)$.

The point is that ℓ might not be cpt.

Remedy: Repeat the reasoning with ε replaced by

$2^{-(\ell+1)}\varepsilon$, $\ell = 1, 2, 3, \ldots$

So $\exists\ (x_i)_{1 \le i \le m(\ell)}$ s.t.

$$\mu_k\Big[\mathcal{X} \setminus \bigcup_{1 \le i \le m(\ell)} B(x_i, 2^{-\ell}\varepsilon)\Big] \le 2^{-\ell}\varepsilon$$

Thus $\quad \mu_k[\mathcal{X} \setminus S] \le \varepsilon.$

$$S := \bigcap_{p=1}^{\infty} \bigcup_{i=1}^{m(p)} \overline{B(x_i, \varepsilon 2^{-p})}$$

By construction, S can be covered by finitely many balls of radius δ, (δ can be arbitrarily small).

Thus S is totally bounded (i.e. it can be covered by finitely many balls of arbitrarily small radius).

S is also closed. Since $\mathcal{X}$ is a complete metric space, then S is compact. $\boxtimes$

Now we can prove the theorem.

Let (μ_k) be s.t. $\mu_k \to \mu$ in distance W_p;

firstly we show that $\mu_k \to \mu$ in $P_p(\mathcal{X})$.

Trivially (μ_k) is of course a Cauchy sequence, by lemma above, $(\mu_k)_k$ is tight. (Prokhorov)

so $\exists$ a subsequence $(\mu_{k'})$ s.t. $\mu_{k'} \to \tilde{\mu}$ weakly (narrow)

Hence

l.s.c.

$$W_p(\tilde{\mu}, \mu) \le \liminf_{k' \to \infty} \underbrace{W_p(\mu_{k'}, \mu)}_{\to 0} = 0.$$

($\mu_k \to \mu$ weakly)

So $\tilde{\mu} = \mu$ and the whole sequence (μ_k) has to converge to μ (since the only possible limit is μ).

We already show the weak convergence in the usual sense (narrow convergence.), but not yet the convergence in $\mathcal{P}_p(X)$.

For any $\varepsilon > 0$, $\exists$ a constant $C_\varepsilon > 0$, s.t.

$\forall a, b \ge 0$, s.t.

$$(a+b)^p \le (1+\varepsilon) a^p + C_\varepsilon b^p. \qquad \text{(Check)}$$

Combining with the triangle inequality,

$$\underbrace{d(x_0,x)^p}_{\le (d(x_0,y) + d(x,y))^p} \le (1+\varepsilon) d(x_0, y)^p + C_\varepsilon d(x,y)^p$$

两边关于 π_k 积分.

Take $\pi_k \in OPT(\mu_k, \mu)$, then

$$\int d(x_0,x)^p \, d\mu_k(x) \le (1+\varepsilon)\int d(x_0,y)^p \, d\mu(y)$$

$$+ C_\varepsilon \underbrace{\int d(x,y)^p \, d\pi_k(x,y)}_{\substack{\| \\ W_p^p(\mu_k,\mu) \xrightarrow{k\to\infty} 0}}$$

thus

$$\limsup_{k\to\infty} \int d(x_0,x)^p \, d\mu_k(x) \le (1+\varepsilon)\int d(x_0,x)^p \, d\mu(x)$$

Letting $\varepsilon \to 0$, we obtain that

$$\limsup_{k\to\infty} \int d(x_0,x)^p \, d\mu_k(x) \le \int \underbrace{d(x_0,x)^p \, d\mu(x)}_{\le \liminf \int d(x_0,x)^p \, d\mu_k(x))}.$$

(given $\mu_k \to \mu$ narrowly,

i.e. Property ii) in def holds true,

so $\mu_k \to \mu$ in $P_p(X)$

Conversely, assume that $\mu_k \to \mu$ weakly in $P_p(\mathcal{X})$, and for each k, $\pi_k \in OPT(\mu_k,\mu)$, s.t.

$$\int d(x,y)^p \, d\pi(x,y) \longrightarrow 0$$

By Prokhorov's theorem, (μ_k) is tight

also $\{\mu\}$ is tight. Hence $\{\pi_k\}$ is itself tight in

$P(X\times X)$. So up to extraction of a subsequence, still denoted by (π_k), one may assume that

$$\pi_k \to \pi \quad \text{weakly in } P(X\times X)$$

By stability of optimal transport (Theorem 8.20), since each π_k is optimal, π is an optimal coupling of μ and μ, it is the trivial coupling $\pi = (Id, Id)_\# \mu$.

Further, since this is independent of the extracted subsequence, $\pi_k \to \pi$ for the whole sequence.

Now take $x_0 \in \mathcal{X}$, $R > 0$.

If $d(x,y) > R$, then $\max\{d(x,x_0), d(x_0,y)\} > R/2$, i.e.

$$1_{d(x,y)\geq R}$$

$$\leq 1_{[d(x,x_0)\geq R/2 \text{ and } d(x,x_0)\geq \frac{d(x,y)}{2}]}$$

$$+ 1_{[d(x_0,y)\geq R/2 \text{ and } d(x_0,y)\geq \frac{d(x,y)}{2}]}$$

So

$$[d(x,y)^p - R^p]_+$$

$$\leq d(x,y)^p 1_{[\,\cdot\,]} + d(x,y)^p 1_{[\,\cdot\,]}$$

$$\leq 2^p d(x,x_0)^p \mathbb{1}_{d(x,x_0)\geq R/2}$$

$$+ 2^p d(x_0,y)^p \mathbb{1}_{d(x_0,y)\geq R/2}$$

It follows that.

$$W_p(\mu_k,\mu)^p = \int d(x,y)^p \, d\pi_k(x,y)$$

$$= \int [d(x,y)\wedge R]^p d\pi_k(x,y) + \int [d(x,y)^p - R^p]_+ d\pi_k(x,y)$$

$$\leq \underbrace{\int (d(x,y)\wedge R)^p d\pi_k(x,y)}_{\swarrow \text{ as } k\to\infty \atop 0} + 2^p \int_{d(x,x_0)\geq R/2} d(x,x_0)^p \underbrace{d\pi_k(x,y)}_{d\mu_k(x)}$$

$$+ 2^p \int_{d(x_0,y)> R/2} d(x_0,y)^p \underbrace{d\pi_k(x,y)}_{d\mu(y)}$$

Hence

$$\limsup_{k\to\infty} W_p(\mu_k,\mu)^p$$

$$\leq \lim_{R\to\infty} 2^p \limsup_{k\to\infty} \left[\int_{d(x,x_0)\geq R/2} d(x,x_0)^p \, d\mu_k(x) + \cdots\right] = 0$$

$d\mu = f\,dx$

$d\nu = g\,dx$

☆ Control by total variation (or $\underline{\underline{\|f-g\|_{L^1}}}$)

Def:

$$\|\mu-\nu\|_{TV} = 2 \inf_{\substack{X\sim\mu \\ Y\sim\nu}} \mathbb{P}[X\neq Y], \quad \Downarrow$$

Kantorovich duality for $\underbrace{c(x,y) = 1_{x\neq y}} = \begin{cases} 0 & \text{if } x=y \\ 1 & \text{if } x\neq y \end{cases}$

Thm (Wasserstein distance is controlled by weighted total variation). (if no weights, this is wrong!) $\|f-g\|_{L^1}\le 2$ $\quad \|f-g\|_{L^1} \le \|f\|_{L^1}+\|g\|_{L^1} \le 2$

Let $\mu, \nu \in P(X)$, (X,d) Polish. Let $p\in[1,\infty)$, $x_0\in X$. Then

$$\boxed{W_p(\mu,\nu) \le 2^{1/p'} \left(\int d(x_0,x)^p \, d|\mu-\nu|(x)\right)^{1/p},}$$

where $1/p + 1/p' = 1$.

$\Downarrow$ $p=1$, $p'=\infty$

Particular case: $p=1$, if $\mathrm{diam}(X) \le D$, this bounds implies $\underbrace{W_1(\mu,\nu)} \le \underbrace{D\|\mu-\nu\|_{TV}}_{\text{diameter}}$.

RK: $\left(\int d(x_0,x)^p \, d|\mu-\nu|(x)\right)^{1/p}$

W_1 for $c(x,y) = [d(x_0,x) + d(x_0,y)]\, 1_{x\neq y}$.

$D = \sup_{\substack{x\in \mathrm{supp}(\mu) \\ y\in\mathrm{supp}(\nu)}} d(x,y)$

Pf: Taking π be the transport plan obtained

$\pi \in \Pi(\mu,\nu)$

by keeping fixed all the mass shared by μ and ν, and distributing the rest uniformly :

this is

taking the smaller one

$$\pi = (Id, Id)_{\#}(\mu \wedge \nu) + \frac{1}{a} (\mu-\nu)_+ \otimes (\mu-\nu)_-$$

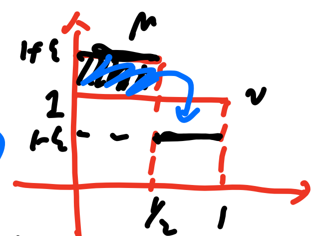
where

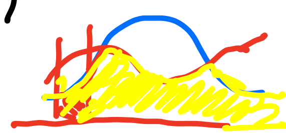
$$\mu \wedge \nu = \mu - (\mu-\nu)_+ ,$$

and $\qquad a = (\mu-\nu)_-[X] = (\mu-\nu)_+[X]$.

Or a more readable way

$$\pi(dx\,dy) = (\mu \wedge \nu)\, dx\, \delta_{y=x} + \frac{1}{a}(\mu-\nu)_+(dx)\,(\mu-\nu)_-(dy)$$





Hence

$$W_p(\mu,\nu)^p \le \int d(x,y)^p \, d\pi(x,y)$$

$$= \frac{1}{a} \int d(x,y)^p \, d(\mu-\nu)_+(x)\, d(\mu-\nu)_-(y)$$

$(a+b)^p \le 2^{p-1}(a^p+b^p)$

$$\le \frac{2^{p-1}}{a} \int \left[d(x,x_0)^p + d(x_0,y)^p \right] d(\mu-\nu)_+(x)\, d(\mu-\nu)_-(y)$$

$$\leq 2^{p-1}\left[\int d(x,x_0)^p \, d\left[(\mu-\nu)_+(x) + (\mu-\nu)_-(x)\right]\right]$$

$$= 2^{p-1}\int d(x,x_0)^p \, d|\mu-\nu|(x). \quad \checkmark$$

Topological properties of the Wasserstein space.

$(\mathcal{P}_p(X), W_p)$ inherits several properties of the base space $\mathcal{X}$.

$(\mathcal{P}_p(X), W_p)$ Polish $\Leftarrow$ if $(\mathcal{X}, d)$ is Polish

Thm: The Wasserstein space over a Polish space is itself a Polish space.

(Complete separable metric space)

Moreover, any probability measure can be approximated by a sequence of prob. measures with finite support.

$$\sum_{j=0}^{N} a_j \cdot \delta_{x_j} \leftarrow \left(\sum_{j=0}^{N} a_j = 1,\ a_j > 0\right)$$

Remark: If $\mathcal{X}$ is compact, then $\mathcal{P}_p(X)$ is also compact; but if $\mathcal{X}$ is only locally compact, then $\mathcal{P}_p(X)$ is not locally compact.

Pf of the above theorem

$(\mathcal{P}_p(X), W_p)$ is a metric space ✓

Need to check: a) separability

b) completeness.

a) Let $\mathcal{D}$ be a dense sequence in X, and let $\mathcal{P}$ be the space of probability measures that can be written as $\sum_j a_j \delta_{x_j}$, $a_j \in \mathbb{Q} \leftarrow$ rational

x_j finite many in $\mathcal{D}$

It turns out that $\mathcal{P}$ is dense in $\mathcal{P}_p(X)$.

To prove this, let $\varepsilon > 0$ given, and let x_0 be an arbitrary element of $\mathcal{D}$.

If $\mu \in \mathcal{P}_p(X)$, then $\exists\ K \subset X$, compact, s.t.

$$\int_{X \setminus K} d(x_0, x)^p \, d\mu(x) \leq \varepsilon^p.$$

Cover K by a finite family of balls $B(x_k, \varepsilon/2)$, $1 \leq k \leq N$, define

$$B_k' = B(x_k, \varepsilon) \setminus \bigcup_{j<k} B(x_j, \varepsilon)$$

Then all B_k' are disjoint and still cover K.

Define f on X by

$$f(B_k' \cap K) = \{x_k\}$$

$$f(\mathcal{X} \setminus K) = \{x_0\}$$

Then, $\forall\, x \in K$, $d(x, f(x)) \leq \varepsilon$.

So

$$\int d(x, f(x))^p \, d\mu(x)$$

$$\leq \varepsilon^p \int_K d\mu(x) + \int_{\mathcal{X} \setminus K} d(x, x_0)^p \, d\mu(x)$$

$$\leq \varepsilon^p + \varepsilon^p = 2\varepsilon^p$$

Since (Id, f) is a coupling of μ and $f_\# \mu$,

$$W_p(\mu, f_\# \mu) \leq 2\varepsilon^p$$

Of course $f_\# \mu$ can be written as $\sum a_j \delta_{x_j}$, $0 \leq j \leq N$.

That is μ might be approximated, with arbitrary precision, by a finite combination of Dirac masses.

To conclude, it is sufficient to note that

$a_j \rightsquigarrow b_j \in \mathbb{Q}$ and

$$W_p\Big(\sum_{j \leq N} a_j \delta_{x_j}, \sum_{j \leq N} b_j \delta_{x_j}\Big)$$

$$\leq 2^{1/p'} \Big[\max_{k,l} d(x_k, x_l)\Big] \sum_{j \leq N} |a_j - b_j|^{1/p}.$$

b) <u>Completness</u>

Let $(\mu_k)_{k\in\mathbb{N}} \subset \mathcal{P}_p(X)$ be a Cauchy sequence.

Hence, it admits a subsequence $(\mu_{k'})$ which converges weakly (i.e. narrow convergence) to some measure μ. Then

$$\int d(x_0,x)^p \, d\mu(x) \leq \liminf_{k'\to\infty} \int d(x_0,x)^p \, d\mu_{k'}(x) < +\infty,$$

so $\mu \in \mathcal{P}_p(X)$.

Moreover, by l.s.c of W_p,

$$W_p(\mu,\mu_{l'}) \leq \liminf_{k'\to\infty} W_p(\mu_{k'},\mu_{l'})$$

So

$$\limsup_{l'\to\infty} W_p(\mu,\mu_{l'}) \leq \limsup_{k',l'\to\infty} W_p(\mu_{k'},\mu_{l'}) = 0,$$

i.e. $\mu_{l'} \to \mu$ in W_p.

Since (μ_k) is Cauchy, then the whole sequence is converging.

Some applications :

- Stability Estimates for the linear eq. in Wasserstein
- Stability Estimates for the nonlinear continuity eq.
  - ↳ Dobrushin's estimate
    Classical result of Mean Field Limit

Mean-Field Limit & Coupling

* Start from Linear continuity eq to Non-linear PDEs

<u>PDE</u>: $\partial_t \rho = \nabla\cdot(\rho \underbrace{\nabla W * \rho}) \; (+\, \sigma \cancel{\Delta_x}\rho)$ $\xrightarrow{} 0$

$$v(t,x) = -\nabla W * \rho_t(x)$$
$$= -\int_{\mathbb{R}^d} \nabla W(x-y)\,\rho_t(dy) \quad \text{(Average or expectation)}$$
$$\xleftarrow{\text{as } N\to\infty} -\frac{1}{N}\sum_{j=1}^{N} \nabla W(x - X_j(t)) \quad \text{(Empirical mean)}$$

(in particular, if $X_j(t) \overset{\text{i.i.d.}}{\sim} \rho_t$)

<u>ODEs</u>:
$$\begin{cases} \dfrac{dX_i^t}{dt} = -\dfrac{1}{N}\displaystyle\sum_{j=1}^{N} \nabla W(X_i^t - X_j^t) \\ X_i\big|_{t=0} = X_i^0 \end{cases} \qquad i=1,2,\dots N$$

mean-field scaling

Goal: If $-\nabla W = F$ is globally Lipschitz, then as $N\to\infty$, ODEs of N-particle system approximates PDE in the following sense:

<u>Mean-Field Limit</u>:
$$W_1(\mu_N(t), \rho_t) \le e^{2\|\nabla F\|_{\infty} t}\, \underbrace{W_1(\mu_N(0), \rho_0)}_{\to 0 \text{ as } N\to\infty},$$
$W_1(\mu_N(t), \rho_t) \to 0$ as $N\to\infty$

where
$$\mu_N(0) = \underbrace{\frac{1}{N}\sum_{i=1}^{N} \delta_{X_i^0}}$$ (empirical measure for initial data of ODEs),

and

$$\mu_N(t) = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i^t}$$

and ρ_t is the unique weak solution to the PDE with initial data ρ_0.

Remark: If initially $X_i^0 \sim \rho_0$, i.i.d., then of course

$$\mathbb{E}\, W_1(\mu_N(0), \rho_0) \lesssim \frac{1}{N^\theta} \to 0$$

See (Fournier & Guérin)

$\theta \sim$ depends on dimension d, some moments of ρ_0

See Golse: Mean-Field Limit for large Systems of Interacting Particles

Standard Method to prove Mean-Field Limit. (Classical): Dobrushin's Estimate

Linear Continuity Eq: (or simply conservation of mass)

$$\partial_t \rho = \nabla \cdot (\rho \nabla V)$$

where the velocity field $v(t,x) = -\nabla V(x)$

$$\left(\text{Corresponding ODE:} \quad \begin{cases} \dot{x}_t = -\nabla V(x_t) \\ x_t|_{t=0} = x_0 \end{cases} \right)$$

$v = -\nabla V$

$V: \mathbb{R}^d \to \mathbb{R}$, $\quad V \in C_b^2$, s.t. $D^2 V(x) \geq \lambda Id$, with $\lambda > 0$.

Also V has a unique global minimum at 0, also with minimal value 0, i.e. $V(0) = 0$.

Continuity Eq: $\begin{cases} \partial_t \rho = \nabla \cdot (\rho \nabla V) \\ \rho|_{t=0} = \rho_0 \end{cases}$

Seek Weak Solutions $\rho \in C([0,T], \mathcal{P}_p(\mathbb{R}^d))$
$(\rho_t \in \mathcal{P}_p(\mathbb{R}^d)$ $\quad)$ $\quad W_p$ $(1 \leq p < \infty)$

Def: We call $\rho \in C([0,T], \mathcal{P}_p(\mathbb{R}^d))$ is a solution to the linear Continuity Eq above with initial data ρ_0 if $\forall \varphi \in C_0^\infty([0,\infty) \times \mathbb{R}^d)$, we have

$$\underline{\int_0^T \int_{\mathbb{R}^d} \partial_t \varphi \, \rho} + \int_{\mathbb{R}^d} \varphi(0,\cdot) \rho_0(x) \, dx$$

$$= \int_0^T \int_{\mathbb{R}^d} \nabla_x \varphi \cdot \nabla V \rho + \int_{\mathbb{R}^d} \varphi(T,x) \, d\rho(T,x).$$

(Explain:

$$\int_0^T \int_{\mathbb{R}^d} \underbrace{\psi\, \partial_t \rho}_{\substack{\| \\ \partial_t(\psi\rho) \\ -(\partial_t\psi)\rho}} = \int_0^T \int_{\mathbb{R}^d} \psi \cdot \nabla_x \cdot (\rho \nabla V)\, dx\, dt$$

$$= -\int_0^T \int_{\mathbb{R}^d} \nabla_x \psi \cdot \nabla V\, \rho(t, dx)\, dt,$$

$$-\int_0^T \int (\partial_t \psi)\rho + \int_{\mathbb{R}^d} \psi(T,x)\rho(T,dx) - \int_{\mathbb{R}^d} \psi(0,x)\rho_0(dx)$$

$$= -\int_0^T \int_{\mathbb{R}^d} \nabla_x \psi \cdot \nabla V\, \rho(t,dx)\, dt)$$

Or the weak formulation of PDE can be rewritten as

$$\int_0^T \int_{\mathbb{R}^d} (\partial_t \psi - \underbrace{\nabla_x V \cdot \nabla_x \psi}) \rho(t, dx)\, dt$$

$$= \int_{\mathbb{R}^d} \psi(T,x)\, d\rho(T,dx) - \int_{\mathbb{R}^d} \psi(0,x)\, \rho_0(dx).$$

$\nabla V \in C_b^2 \Leftarrow V \in C_b^3$

Basic Facts : $\begin{cases} \dot{x}_\tau = -\nabla V(x_\tau) & s < \tau < t \\ x(s) = x \in \mathbb{R}^d \end{cases}$

Cauchy–Lipschitz

- globally Lipschitz
- $|\nabla V(x)| \le |\nabla V(x) - \nabla V(0)| + |\nabla V(0)| \le C(1+|x|)$ linear growth

Flow map: $\Phi_{s,t}(x) : x \longmapsto \Phi_{s,t}(x)$

a family of diffeomorphism $\mathbb{R}^d \circlearrowleft$

$$\Phi_{s,s}(x) = x$$

$$\Phi_{0,t}(x) = \Phi_t(x)$$

$$\Phi_t(x) = x + \int_0^t -\nabla V(\Phi_s(x))\, ds$$

$$|\Phi_t(x) - x| \le c \int_0^t |\Phi_t(x)|\, ds$$

$$\Rightarrow \quad |\Phi_t(x)| \lesssim_T 1 + |x|.$$

- Using Duality argument to show the unique weak solution to $\begin{cases} \partial_t \rho = \nabla\cdot(\rho \nabla V) \\ \rho|_{t=0} = \rho_0 \end{cases}$

is given by $\rho_t = \Phi_{t\#}\rho_0$

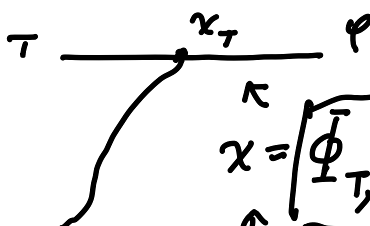
Consider the Cauchy Problem (dual of the linear PDE above)

$$\begin{cases} \partial_t \psi - \overbrace{\nabla V}^{v = \text{velocity field}} \cdot \nabla \psi = 0, & \text{for } t < T,\ x \in \mathbb{R}^d. \\ \psi(T, x) = \varphi(x) \in C_0^\infty(\mathbb{R}^d) \end{cases}$$

$\downarrow$ Very Easy to solve this transport eq.

Solution:

$$\boxed{\psi(t,x) = \varphi(\Phi_{t,T}(x))}$$



$$x = \boxed{\Phi_{T,t}(x_T)}$$

$\Updownarrow$

- Prove the uniqueness of weak solution to linear PDE. $\quad X_T = \Phi_{t,T}(x)$

Recall the weak formulation:

$$\int_0^T \int_{\mathbb{R}^d} \cancel{\Big(\underbrace{\frac{\partial \psi}{\partial t} - \nabla V \cdot \nabla \psi}_{\equiv 0}\Big)}^{=0} \rho(t, dx)\,dt$$

$$= \int_{\mathbb{R}^d} \underbrace{\psi(T, x)}_{\varphi(x)} \rho(T, dx) - \int_{\mathbb{R}^d} \psi(0, x) \underbrace{\rho_0(dx)}_{=0}$$

$$\begin{cases} \partial_t \rho = \nabla \cdot (\rho \nabla V) \\ \rho_0^i, \ i = 1, 2, \ \rho_0 = \underbrace{\rho_0^1 - \rho_0^2}_{\equiv 0} \end{cases}$$

Then take $\psi(t, x) = \varphi(\Phi_{t,T}(x))$

$$\frac{d}{dt}\psi(t, X_t) = \partial_t \psi + \nabla\psi \cdot \dot{X}_t$$
$$= \partial_t \psi - \nabla V(X_t) \cdot \nabla\psi(t, X_t) = 0$$

$$\Rightarrow \forall \varphi \in C_0^\infty(\mathbb{R}^d), \quad \int_{\mathbb{R}^d} \varphi(x) \rho(t, dx) = 0$$

$$\Rightarrow \boxed{\rho(t, \cdot) \equiv 0}$$

This proves the uniqueness. $\square$

(Duality Method)

The unique solution is indeed given by

$$\underbrace{\rho_t = \Phi_{t\#} \rho_0}$$ (Just check it).

Why $\rho_t = \Phi_{t\#}\rho_0 \in C([0,T], \mathcal{P}_p(\mathbb{R}^d))$

give $\rho_0 \in \mathcal{P}_p(\mathbb{R}^d)$?

$$(\Leftarrow \quad W_p(\Phi_{t\#}\rho_0, (\Phi_s)_\#\rho_0) \qquad (t > s)$$

$$= W_p(\underbrace{\Phi_{t-s\#}}(\underbrace{\Phi_{s\#}\rho_0}_{=\mu}), \underbrace{(\Phi_s)_\#\rho_0}_{=\mu}) \qquad \underset{C|t-s|-\text{Lipschitz}}{(\Phi_{t-s})_\#}$$

$$\le C\,|t-s| \quad)$$

globally Lipschitz

Later $\to \nabla W * \rho <$ linear growth

Stability in W_1: $\boxed{v(x)} = \boxed{-\nabla V} \in W^{1,\infty}$

can be general $v = v(t,x)$

globally Lipschitz

$$(\quad \|\nabla v\|_{L^\infty} \le L \quad)$$

FACT: The flow map Φ_t is also Lipschitz with constant e^{Lt}. (Will use Gronwall inequality repeatedly)

Computations: $\dot{x}_t = -\nabla V(x_t)$

$$\begin{cases} \Phi_t(x) = x - \int_0^t \nabla V(\Phi_s(x))\,ds \\ \Phi_t(y) = y - \int_0^t \nabla V(\Phi_s(y))\,ds \end{cases}$$

Then

$$(\ \|\nabla^2 V\|_{L^\infty} \le L)$$

$$|\Phi_t(x) - \Phi_t(y)| \le |x-y| + L\int_0^t |\Phi_s(x) - \Phi_s(y)|\, ds$$

$\Downarrow$ Gronwall inequality

$$|\Phi_t(x) - \Phi_t(y)| \le \boxed{|x-y|\, e^{tL}}$$

Take a function φ s.t. $\|\nabla\varphi\|_{L^\infty} \le 1$,

then

$$\int_{\mathbb{R}^d} \varphi(x)\, d(\underbrace{\Phi_{t\#}\mu_1}_{\rho_t^1} - \underbrace{\Phi_{t\#}\mu_2}_{\rho_t^2})$$

NOT necessarily invariant data

$$= \int_{\mathbb{R}^d} \varphi(\Phi_t(x))\, d(\mu_1 - \mu_2)(x)$$

$$\overset{\pi_0 \in OPT(\mu_1,\mu_2)}{=} \int_{\mathbb{R}^d} \big(\varphi(\Phi_t(x)) - \varphi(\Phi_t(y))\big)\, d\pi_0(x,y)$$

$\|\nabla\varphi\|_{L^\infty} \le 1$

$\le |\Phi_t(x) - \Phi_t(y)|$

$$\le e^{tL}\int_{\mathbb{R}^d} |x-y|\, d\pi_0(x,y)$$

$$= e^{tL}\, W_1(\mu_1,\mu_2)$$

Or we obtain:

$$W_1(\underbrace{\Phi_{t\#}\mu_1}_{\rho_1^t}, \underbrace{\Phi_{t\#}\mu_2}_{\rho_2^t}) \le e^{Lt}\, W_1(\mu_1,\mu_2)$$

$\|\nabla V\|_{L^\infty} \le L$ or have $\|\nabla^2 V\|_{L^\infty} \le L$.

This is the stability estimate (gives well-posedness) of solutions $\rho(t) = \underbrace{\Phi_{t\#}}\mu \in \underbrace{C([0,T], \mathcal{P}_1(\mathbb{R}^d))}$ in W_1 - distance.

Lecture 18

Consider the solution $x = x(t)$ to the finite dimensional gradient flow

$$\begin{cases} \frac{dx(t)}{dt} = -\underbrace{\nabla V(x(t))} & \text{in } t > 0 \\ x(0) = x_0 \in \mathbb{R}^d \end{cases}$$

Correspondingly $\begin{cases} \rho_0 = \delta_{x_0} \\ \rho(t) = \delta_{x(t)} \end{cases}$ solves the linear PDE:

$$\begin{cases} \partial_t \rho = \operatorname{div}(\rho \nabla V) \\ \rho|_{t=0} = \delta_{x_0} \end{cases}$$

Assumptions: $\begin{cases} \nabla^2 V \geq \lambda I d \\ V(x) \geq V(0) = 0 \quad \forall x \in \mathbb{R}^d. \\ V \in C_b^2 \end{cases}$

$W_2(\rho(t), \delta_0) \leq e^{-\lambda t} W_2(\rho_0, \delta_0)$

This implies the exponential convergence rate of weak solutions ρ_t of the linear PDE towards the equilibrium ρ_∞.

As a basic illustration, given any two

solutions $x_1(t)$ and $x_2(t)$ of $\frac{dx(t)}{dt} = -\nabla V(x(t))$,

one has

$$W_2(\delta_{x_1(t)}, \delta_{x_2(t)}) \leq e^{-\lambda t} W_2(\delta_{x_1(0)}, \delta_{x_2(0)}),$$

or here

$$|x_1(t) - x_2(t)| \leq e^{-\lambda t} |x_1(0) - x_2(0)|.$$

($\Leftarrow$ Computations: Recall $\dot{x}_i(t) = -\nabla V(x_i(t))$

$$\frac{d}{dt}|x_1(t) - x_2(t)|^2 = 2\,(x_1(t) - x_2(t)) \cdot \underbrace{(\dot{x}_1(t) - \dot{x}_2(t))}$$

here using $\nabla^2 V \geq \lambda Id$ i.e. V is λ-convex

$$= -2\,\underbrace{(x_1(t) - x_2(t)) \cdot (\nabla V(x_1(t)) - \nabla V(x_2(t)))}_{\geq \lambda |x_1(t)-x_2(t)|^2}$$

$$\leq -2\underbrace{\lambda\,|x_1(t) - x_2(t)|^2}$$

λ-convex

okay.)

In general, we have

Thm: Given $V \in C_b^2$, s.t. $D^2 V(x) \geq \lambda Id$ in $\mathbb{R}^d$ with $\lambda > 0$, and $V(x) \geq V(0) = 0$, $\forall x \in \mathbb{R}^d$.

Given any two weak solutions $\rho_1(t)$, $\rho_2(t)$ of the linear eq. $\partial_t \rho = \operatorname{div}(\rho \nabla V)$ in $C([0,T], \mathcal{P}_2(\mathbb{R}^d))$, we have

$$\boxed{W_2(\rho_1(t), \rho_2(t)) \le e^{-\lambda t} W_2(\rho_1(0), \rho_2(0))}$$

$\in \mathcal{P}_2(\mathbb{R}^d)$

and as a consequence (taking $\rho_\infty = \delta_0$)

$$W_2(\rho_1(t), \rho_\infty) = W_2(\rho_1(t), \delta_0)$$

$$\le e^{-\lambda t} W_2(\rho_1(0), \delta_0).$$

$\frac{d}{dt} \underbrace{W_2^2(\rho_1(t), \rho_2(t))}$

$\frac{d}{dt} \underbrace{\mathbb{E}|X(t)-Y(t)|^2}$

Pf: (An alternative one,

"different" to the coupling method previously)

Let π_0 be the optimal coupling between $\rho_1(0)$ and $\rho_2(0)$ for W_2-distance. $\pi_0 \in OPT(\rho_1(0), \rho_2(0))$

Of course, $\rho_i(t) = \Phi_{t\#}\rho_i(0)$, $i = 1, 2.$

Define $\pi_t = (\Phi_t, \Phi_t)_\# \pi_0 \in \Pi(\rho_1(t), \rho_2(t))$,

Then $W_2^2(\rho_1(t), \rho_2(t)) \le \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 \, d\pi_t(x, y)$

push-forward formula

$\Phi_{t,s}(x)$ Φ_t

0 t

$$\le \int_{\mathbb{R}^d\times\mathbb{R}^d} |\Phi_t(x) - \Phi_t(y)|^2 \, d\pi_0(x, y)$$

$-2\cdot(x-y)\cdot(\nabla V(x) - \nabla V(y))$

Claim:

$$\boxed{\frac{d}{dt}\Big|_{t=0^+} \frac{1}{2} W_2^2(\rho_1(t), \rho_2(t)) \le -\int_{\mathbb{R}^d\times\mathbb{R}^d} (x-y)\cdot(\nabla V(x) - \nabla V(y)) \, d\pi_0}$$

$\le -\lambda |x-y|^2$

$\Downarrow$

$$\text{LHS} = \lim_{t\to 0^+} \frac{1}{2t}\left(W_2^2(\rho_1(t), \rho_2(t)) - W_2^2(\rho_1(0), \rho_2(0))\right)$$

$$= \lim_{t\to 0^+} \int \frac{1}{2t}\, \underbrace{\frac{|\Phi_t(x)-\Phi_t(y)|^2 - |x-y|^2}{t}}\, d\pi_0(x,y)$$

Try to use Lebesgue Dominated Convergence theorem

$\downarrow t \to 0^+$

(Fill the gaps here: Now we only have the limit $-(x-y)\cdot(\nabla V(x) - \nabla V(y)) \le 0$.)

$$-(x-y)\cdot(\nabla V(x) - \nabla V(y))$$

$$\left(\text{also} \le -\lambda |x-y|^2 \le 0\right)$$

Applying Fatou's lemma (NOT RIGHT)

~~(A sequence of functions $\{f_n\}$, $f_n \ge 0$, E measurable, then~~

$$\cancel{\int_E \liminf_{n\to\infty} f_n \, dx \le \liminf_{n\to\infty} \int_E f_n \, dx.}$$ ~~)~~

The claim then leads to

$$\left.\frac{d}{dt}\right|_{t=0^+} W_2^2(\rho_1(t), \rho_2(t)) \le -2\lambda\, W_2^2(\rho_1(0), \rho_2(0)).$$

This inequality can also be derived at any time $t \ge 0$, we obtain

$$\frac{d^+}{dt} W_2^2(\rho_1(t), \rho_2(t)) \le -2\lambda\, W_2^2(\rho_1(t), \rho_2(t)),$$

$$\forall\, t \ge 0.$$

Integrating in time will give the conclusion.

Go to nonlinear setting:

Dobrushin's Approach:

Existence, stability, and derivation of the aggregation eq.

Assumptions: $\underline{W \in C_b^2}$, $\underline{\overset{\nabla}{W}(x) = W(-x)}$, even, $\nabla W(0) = 0$

$v(t,x) = -\nabla W * \rho_t(x)$ Fix point argument.

consider PDE: $\partial_t \rho = \operatorname{div}(\rho(\nabla W * \rho))$

(W: interaction potential)

* Well-posedness in $C([0,T], \mathcal{P}_1(\mathbb{R}^d))$ with initial $\rho_0 \in \mathcal{P}_1(\mathbb{R}^d)$, $\rho_t \in \mathcal{P}_1(\mathbb{R}^d)$

Idea: Fix point argument $\Rightarrow \begin{cases} \text{existence} \\ \text{uniqueness} \end{cases}$

Given $\rho \in C([0,T], \mathcal{P}_1(\mathbb{R}^d))$ ← (assume the law is given a priori)

define the corresponding velocity field

$$v(\rho)(t,x) = -\nabla W * \rho_t(x) = \int_{\mathbb{R}^d} -\nabla W(x-y)\, \rho_t(dy)$$

By assumptions on W, $\in C_b^2$ one has $\nabla W(0) = 0$,

$$|\nabla W(x)| \leq C(1+|x|) \quad \leftarrow \text{actually } |\nabla W(x)| \leq C|x|$$

and $\|\nabla^2 W\|_{L^\infty} \leq L$. $\|\nabla^2 W\|_{L^\infty} \leq L$

Hence for $\underbrace{v(\rho)}(t,x) = -\nabla W * \rho_t(x)$

$$= -\int_{\mathbb{R}^d} \nabla W(x-y)\,\rho(t,dy),$$

• $v(\rho)(t,x)$ C^0 $W^{1,\infty}$ linear growth

• $$|v(\rho)(t,x)| \le C\int (1+|x-y|)\,d\rho(t)(y)$$

Linear growth in x.

$$\le C[1+|x| + M(\rho)], \qquad t\in[0,T].$$

and $$M(\rho) = \max_{0\le t\le T}\int_{\mathbb{R}^d} |x|\,d\rho(t)(x). \checkmark$$

• $|v(\rho)(t,x) - v(\rho)(t,y)|$

$v = v(\rho)(t,x)$ Lipschitz in x.

$$\le \int_{\mathbb{R}^d} |\nabla W(x-z) - \nabla W(y-z)|\,d\rho(t,z)$$

$$\le L|x-y|, \qquad \forall t\in[0,T]$$

• $v = v(\rho)(t,x)$ is <u>also continuous in time</u>

$|v(\rho)(t,x) - v(\rho)(s,x)|$

$\rho \in C([0,T], \mathcal{P}_1(\mathbb{R}^d))$

$$\le \left|-\int \underbrace{\nabla W(x-y)}_{\text{as a test function}}\, d(\rho_t - \rho_s)(y)\right|$$

$$\le \|\nabla^2 W\|_{L^\infty}\, W_1(\rho_t,\rho_s)$$

Hence, by Cauchy-Lipschitz theory, one has well-defined flow map associated to the ODE

$$\begin{cases} \dfrac{dx(t)}{dt} = \underbrace{v(\rho)(t,x)}_{\text{velocity field generated by } \rho} = -\nabla W * \rho_t(x(t)), & t \geq 0 \\ x(0) = x \in \mathbb{R}^d \end{cases}$$

We denote the flow map by $\boxed{\Phi_t(\rho)}$,

of course $\rho = \rho_t \in C([0,T], \mathcal{P}_1(\mathbb{R}^d))$.

One has the properties:

① $\forall t \in [0,T]$, $\exists C(T) \sim M(\rho)$, s.t.

$$|\Phi_t(\rho)(x)| \leq C(T)(1+|x|), \quad \forall x \in \mathbb{R}^d.$$

(Linear growth)

② The map $x \mapsto \Phi_t(\rho)(x)$ is Lipschitz:

$$|\Phi_t(\rho)(x) - \Phi_t(\rho)(y)| \leq e^{Lt} |x-y|;$$

$\forall t \geq 0$, $\forall x, y \in \mathbb{R}^d$.

where $L = \|\nabla^2 W\|_{L^\infty}$

iii) Continuity in t: $\forall T > 0$, $\exists C(T) > 0$, $C(T) \sim (M(\rho), T)$, s.t.

$$|\Phi_t(\rho)(x) - \Phi_s(\rho)(x)| \quad \Leftarrow \begin{cases} \Phi_t(\rho)(x) = x + \int_0^t v(\rho)(r, \Phi_r(\rho)(x))\, dr \\ \Phi_s(\rho)(x) = x + \int_0^s \cdots \, dr \end{cases}$$

$\leq C(T)(1+|x|)|t-s|$, for all $0 \leq t, s \leq T$.

$v(\rho)(t,x) = -\nabla W * \rho(t,x)$ and $x \in \mathbb{R}^d$.

Pf: i) Note simply that

good given

$$\Phi_t(\rho)(x) = x + \int_0^t \boxed{v(\rho)}(t, \Phi_s(\rho)(x))\, ds$$

given

Hence

$$|\Phi_t(\rho)(x)| \le |x| + \int_0^t C(T)\,(1 + |\Phi_s(\rho)(x)|)\, ds.$$

Applying Gronwall will then conclude.

ii)

$$|\Phi_t(\rho)(x) - \Phi_t(\rho)(y)|$$

$$\le |x-y| + \int_0^t |v(\rho)(t, \Phi_s(\rho)(x)) - v(\rho)(t, \Phi_s(\rho)(y))|\, ds.$$

$$\le |x-y| + L\int_0^t |\Phi_s(\rho)(x) - \Phi_s(\rho)(y)|\, ds,$$

$$\begin{cases} \partial_t \rho = \nabla\cdot(\rho\, \nabla W * \rho) \\ \rho_i^0 \end{cases}$$

along. Φ_t is the same then the difference comes from the natural data

Stability estimate between $\underbrace{\text{two different flow maps}}_{\Uparrow}$

$\cdot\rho_1^0,\ \rho_2$

$\rho_1^0,\ \rho_2^0$ $\Phi_t^1\ \Phi_t^2$ induced by two given curves $\Phi_t^i = \Phi_t(\rho_i)$

$\rho_1^t\ \ \rho_2^t$ e^{2Lt}

$\rho_i \in \mathcal{C}([0,T], \mathcal{P}_1(\mathbb{R}^d))$

Metric: $D_{1,T}(\rho_1, \rho_2) := \max_{0 \le t \le T} \underline{W_1(\rho_1(t), \rho_2(t))}$ ↙ path space this makes the path space complete for all $T > 0$.

We will do fixed point argument in $(C([0,T], P_1(\mathbb{R}^d)), D_{1,T})$.

Lemma. Given $\rho_i \in C([0,T], P_1(\mathbb{R}^d))$,

and $v^i = v^i(\rho_i)$, $\Phi_t^i = \Phi_t(\rho_i)$, $i = 1, 2$,

$= -\nabla W * \rho_i^t(x)$

then

$$|\Phi_t^1(x) - \Phi_t^2(x)| \le L \int_0^t e^{L(t-s)} W_1(\rho_1(s), \rho_2(s))\, ds,$$

for all $0 \le t \le T$, $\forall x \in \mathbb{R}^d$, and consequently

integrating over x w.r.t. ρ_0

also $W_1(\rho_1(s), \rho_2(s)) \le D_{1,T}(\rho_1, \rho_2)$

$$\star\ \underbrace{W_1(\Phi^1_{t\#}\rho_0, \Phi^2_{t\#}\rho_0) \le (e^{Lt} - 1)\, D_{1,T}(\rho_1, \rho_2)}$$

for all $\begin{cases} 0 \le t \le T \\ \rho_0 \in P_1(\mathbb{R}^d) \end{cases}$.

$\rho_i \rightsquigarrow ((\Phi_t^i)_\# \rho_0)_{t \in [0,T]}$

Take $t \ll 1$, s.t. $|e^{Lt} - 1| < 1$

Pf: For all $t \in [0,T]$, $i = 1, 2$, one has

$$\Phi_t^i(x) = x + \int_0^t v^i(s, \Phi_s^i(x))\, ds \qquad \rho_i$$

Then

$$|\Phi_t^1(x) - \Phi_t^2(x)|$$

$$\le \int_0^t |v^1(s, \Phi_s^1(x)) - v^2(s, \Phi_s^2(x))|\, ds$$

$$\le \int_0^t |v^1(s, \Phi_s^1(x)) - v^1(s, \Phi_s^2(x))|\, ds$$

Lipschitz property of v^1

$$+ \int_0^t \underbrace{\left| v^1(s, \Phi_s^2(x)) - v^2(s, \Phi_s^2(x)) \right|}_{\le \underline{L\, W_1(\rho_1(s), \rho_2(s))}} ds$$

$$\le \underline{L} \int_0^t \left| \Phi_s^1(x) - \Phi_s^2(x) \right| ds$$

$v^1(s,x) = -\int \underline{\nabla W}(x-y) \rho_1(s,dy)$

$v^2(s,x) = -\int \underline{\nabla W}(x-y) \rho_2(s,dy)$

$$+ \underline{L} \int_0^t W_1(\rho_1(s), \rho_2(s))\, ds.$$

By Gronwall, $\left(\underset{\shortparallel}{f(t)} \le L \int_0^t \underset{\shortparallel}{f(s)}\, ds + L \int_0^t \alpha(s)\, ds \right)$

$$\left| \Phi_t^1(x) - \Phi_t^2(x) \right| \le L \int_0^t e^{L(t-s)} W_1(\rho_1(s), \rho_2(s))\, \underset{\shortparallel}{ds}. \quad (*)$$

(Leave as an exercise)

Taking the transference plan $x \sim \rho_0(x)$

$$(\Phi_t^1, \Phi_t^2)_\# \rho_0 \in \Pi(\underline{\Phi_t^1{}_\# \rho_0}, \underline{\Phi_t^2{}_\# \rho_0}),$$

then

$$W_1(\underline{\Phi_t^1{}_\#} \rho_0, \Phi_t^2{}_\# \rho_0)$$

$$\le \underline{\int_{\mathbb{R}^d} \left| \Phi_t^1(x) - \Phi_t^2(x) \right| d\rho_0(x)}$$

$$\le L \int_0^t e^{L(t-s)} \underline{W_1(\rho_1(s), \rho_2(s))}\, ds$$

$$\le \underline{\underline{L \int_0^t e^{L(t-s)} ds}} \; \underbrace{D_{1,T}(\rho_1, \rho_2)}_{\shortparallel\ \sup_{t \in [0,T]} W_1(\rho_1^t, \rho_2^t)}$$

$$= \underline{(e^{Lt}-1)}\, D_{1,T}(\rho_1,\rho_2).$$

okay.

Banach fixed point argument

(Similar strategy to Picard's iteration as in Cauchy-Lipschitz theorem)

Thm: Given $W \in C_b^2(\mathbb{R}^d)$, then $\exists !$ global in time weak solution ρ in $C([0,\infty), \mathcal{P}_1(\mathbb{R}^d))$ to the non-linear PDE

$$\begin{cases} \partial_t \rho = \nabla\cdot(\rho \boxed{\nabla W * \rho}) \\ \rho|_{t=0} = \rho_0 \end{cases}.$$

CT Lipschitz

Iteration: $\rho^0(t,x) \equiv \rho_0(x) \longrightarrow \rho^1(t,\cdot) = F(\rho_0)_t$

$\rho^2(t,\cdot) = F(\rho^1)_t$ ...

Pf: (Fill the details later)

$T \ll \frac{1}{L}$

Let $T>0$, to be chosen later.

Define $X = (C([0,T], \mathcal{P}_1(\mathbb{R}^d)), D_{1,T})$

Define the map: $F: X \to X$ as

Fix initial data $\rho(0) = \mu \in \mathcal{P}_1(\mathbb{R}^d)$, $\rho = (\rho_t)_{t\in[0,T]}$

$$F(\rho) = \left(\left(\Phi_t(\rho)\right)_\# \mu\right)_{t\in[0,T]}$$

Note: $$\partial_t \underline{\underline{(\mathcal{F}(\rho))_t}} = \nabla\cdot\left(\underline{\underline{(\mathcal{F}(\rho))_t}} \cdot \underbrace{\nabla W * \rho_t}_{\text{given}}\right)$$

By the above lemma,

$$W_1\Big(\underbrace{\Phi_t(\rho_1)_\# \mu}_{\mathcal{F}(\rho_1)},\ \underbrace{\Phi_t(\rho_2)_\# \mu}_{\mathcal{F}(\rho_2)}\Big) \le (e^{Lt}-1)\, D_{1,T}(\rho_1,\rho_2)$$

$\sup_{t\in[0,T]}$

or $$D_{1,T}(\mathcal{F}(\rho_1), \mathcal{F}(\rho_2)) \le (e^{Lt}-1)\, D_{1,T}(\rho_1,\rho_2) \le \frac{1}{2} D_{1,T}(\rho_1,\rho_2)$$

given $0<T \ll 1$.

Then by Banach fixed point theorem,

we have $\left\{\begin{array}{l}\text{existence} \\ \text{uniqueness}\end{array}\right.$ & of weak solutions to

$$\begin{cases} \partial_t \rho = \nabla\cdot(\rho\, \nabla W * \rho) \\ \rho|_{t=0} = \mu. \end{cases}$$

$\blacksquare$

Thm (Dobrushin Stability Estimate)

Given $W \in C_b^2(\mathbb{R}^d)$. Consider two solutions ρ_i, $i=1,2$, in $C([0,\infty), \mathcal{P}_1(\mathbb{R}^d))$ to the (McKean)-Vlasov PDE, then

$\partial_t \rho = \nabla\cdot(\rho \nabla W * \rho) + \sigma\Delta\rho$

$$W_1(\rho_1(t), \rho_2(t)) \le e^{2Lt} W_1(\rho_1(0), \rho_2(0)),$$

for all $t \ge 0$.

e^{2Lt} (Question: Can we do better than $2L$?)

Pf: We write $\Phi_t^i = \Phi_t(\rho_i)$ the flow maps

Then

self-consistency (the representation is always true)

$$W_1(\rho_1(t), \rho_2(t)) = W_1(\Phi^1_{t\#}\rho_1(0), \Phi^2_{t\#}\rho_2(0))$$

$$\le W_1(\Phi^1_{t\#}\rho_1(0), \Phi^2_{t\#}\rho_1(0)) + W_1(\Phi^2_{t\#}\rho_1(0), \Phi^2_{t\#}\rho_2(0))$$

$$\le \underbrace{\int_{\mathbb{R}^d} |\Phi_t^1(x) - \Phi_t^2(x)| \, d\rho_1(0)} + \underbrace{W_1(\Phi^2_{t\#}\rho_1(0), \Phi^2_{t\#}\rho_2(0))}$$

$$\le L\int_0^t e^{L(t-s)} W_1(\rho_1(s), \rho_2(s))\, ds + \boxed{}$$

$\boxed{} \le e^{Lt} W_1(\rho_1(0), \rho_2(0))$

As argued before,

$$W_1(\Phi^2_{t\#}\rho_1(0), \Phi^2_{t\#}\rho_2(0))$$

$$\le \int_{\mathbb{R}^d\times\mathbb{R}^d} |\Phi_t^2(x) - \Phi_t^2(y)| \, d\pi_0(x,y)$$

$$\leq e^{Lt} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y| \, d\pi_0(x,y)$$

$$\leq \underline{e^{Lt} W_1(\rho_1(0), \rho_2(0))}$$

Hence

$$\underline{W_1(\rho_1(t), \rho_2(t))} \leq L \int_0^t e^{L(t-s)} \underline{\underline{W_1(\rho_1(s), \rho_2(s))}} \, ds$$

$$+ \; e^{Lt} \underline{\underline{W_1(\rho_1(0), \rho_2(0))}}$$

$\Downarrow$ Gronwall

By Gronwall, one has.

$$W_1(\rho_1(t), \rho_2(t)) \leq e^{2Lt} W_1(\rho_1(0), \rho_2(0)).$$

($\Leftarrow$ write $x(t) \triangleq e^{-Lt} W_1(\rho_1(t), \rho_2(t))$, then

$$x(t) \leq L \int_0^t x(s) \, ds + x(0)$$

$$\left(x(t) - L \int_0^t x(s) \, ds \right) e^{-Lt} \leq x(0) e^{-Lt}$$

$$\left(e^{-Lt} \int_0^t x(s) \, ds \right)' \leq x(0) e^{-Lt}$$

$$e^{-Lt} \int_0^t x(s) \, ds \leq x(0) \int_0^t e^{-Ls} \, ds$$

$$= -x(0) \cdot \frac{1}{L} e^{-Ls} \Big|_0^t$$

$$= \frac{1}{L} x(0)\left(1 - e^{-Lt}\right)$$

i.e.

$$L \int_0^t x(s)\, ds \le x(0)\left(e^{Lt} - 1\right)$$

$$\Rightarrow \quad x(t) \le x(0)\, e^{Lt}. \Big)$$

Mean-Field Limit:

N particle system given by ODEs: $\quad K = -\nabla W$

$$(\text{ODE}_s)\begin{cases} \dfrac{dX_t^i}{dt} = -\dfrac{1}{N} \displaystyle\sum_{j \neq i} \nabla W(X_t^i - X_t^j), & i = 1, 2, \ldots, N. \\ X_t^i \big|_{t=0} = X_0^i \sim \mu_N(0) \end{cases}$$

Given $W \in C_b^2$ (not necessarily this strong), ODE (or PDE) has a unique globally defined empirical measure solution: 經驗測度

$$\underset{\pi}{\mu_N(t)} = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}. \qquad \partial_t \rho = \nabla \cdot (\rho \nabla W * \rho)$$

Now the velocity field:

$$v^N(t,x) = \boxed{-\nabla W * \mu_t^N(x)}$$

$$= -\frac{1}{N}\sum_{j=1}^{N} \nabla W(x - X_t^j) \quad \checkmark$$

$$(\nabla W(0) = 0).$$

then $$\frac{dX_t^i}{dt} = v^N(t, X_t^i), \quad i = 1, 2, \cdots, N.$$

Flow map: $\Phi_t^N = \Phi_t(\mu^N)$,

then $X_t^i = \Phi_t^N(X_0^i)$, $i = 1, 2, \cdots, N$.

Then of course,

$$\mu^N \in C([0,\infty), P_1(\mathbb{R}^d))$$

is the unique solution to

(McKean)-Vlasov type PDE: $$\begin{cases} \partial_t \rho = \nabla \cdot (\rho \nabla W * \rho) \\ \rho|_{t=0} = \mu^N(0), \end{cases}$$

where we recall that

$$\mu^N(0) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_0^i},$$

(initial data chosen as "particle form")

and

$$\boxed{\mu^N(t) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_t^i}},$$

where X_t^i, $i = 1, \cdots, N$ is the unique solution to the N-particle ODEs with initial data X_0^i, $i = 1, 2, \cdots, N$.

RK: Note empirical measures are weak solutions to the PDE with "particle" initial data.

ODEs ≈ PDE

Dobrushin's Estimate gives a direct derivation of the Vlasov PDE from the ODEs for interacting particle systems. (or Mean-Field Limit)

$\boxed{\text{We write it as a theorem.}}$

W is even

Given $W \in C_b^2$: Take a sequence of empirical measure

$$\mu^N(0) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_0^i}$$

with initial data X_0^i, $i = 1, 2, \dots N$, s.t.

$$W_1(\mu^N(0), \rho(0)) \to 0, \quad \text{as} \quad N \to \infty$$

for $\rho_1(0) \in \mathcal{P}_1(\mathbb{R}^d)$

Define $\mu^N(t) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_t^i}$, where X_t^i solves the ODEs with initial data X_0^i, $i = 1, 2, \dots N$.

Then

$$W_1(\mu^N(t), \rho(t)) \le e^{2Lt} W_1(\mu^N(0), \rho(0)) \to 0 \quad \text{as} \quad N \to \infty.$$

for all $t \in [0, T]$,

Note: Both $\mu^N(t)$, $\rho(t)$ solve $\partial_t \rho = \nabla \cdot (\rho \nabla W * \rho)$ weakly, so Dobrushin's estimate applies.

where $\rho(t)$ is the unique weak solution in $C([0,\infty), \mathcal{P}_1(\mathbb{R}^d))$ to PDE with initial data ρ_0.

Note:

For deterministic system:

$$\begin{cases} \partial_t \rho = \nabla \cdot (\rho \nabla W * \rho) & \leftarrow \text{MFE.} \\ \rho|_{t=0} = \rho_0 \end{cases}$$

$$\Updownarrow$$

$$\begin{cases} \dfrac{dX_t^i}{dt} = -\dfrac{1}{N} \sum_{j=1}^{N} \nabla W(X_t^i - X_t^j) & \leftarrow \text{ODEs} \\ & i = 1, 2, \dots N \\ X_t^i|_{t=0} = X_0^i \end{cases}$$

For any t,

$\to 0$ as $N \to \infty$.

as $N \to \infty$ $\to 0$

$$W_1(\rho_1(t), \mu_N(t)) \le e^{2L|t|} W_1(\rho_1(0), \mu_N(0))$$

$t \in [0, T]$ where $L = \|D^2 W\|_{L^\infty} < \infty$.

Numerics: consistent (discrete $\Rightarrow$ continuous)

Note: Sym $\nabla W(0) = 0$

$$\begin{cases} dX_t^i = -\dfrac{1}{N} \sum_{j=1}^{N} \nabla W(X_t^i - X_t^j)\, dt \\ \quad + \sqrt{2\sigma}\, dW_t^i \end{cases}$$

$\downarrow$ | $X_0^i \sim \bar{\rho}_0$ i.i.d.

or $\mathrm{Law}(X_0^1 \cdots X_0^N) \sim \rho_N^0 \in P_{sym}((\mathbb{R}^d)^N)$

$$\mu_t^N := \frac{1}{N} \sum_{j=1}^{N} \delta_{X_t^j}$$ ← This is a random probability measure

$$\partial_t \mu_t^N = \nabla \cdot (\mu_t^N \nabla W * \mu_t^N) + \sigma \Delta_x \mu_t^N$$

$$\boxed{+ \underbrace{\text{Martingale Term}}_{\downarrow \text{ as } N \to \infty \atop 0}}$$

will vanish by Law of Large Number

Limit PDE: $\partial_t \rho = \nabla \cdot (\rho \nabla W * \rho) + \sigma \Delta_x \rho$

Later: $\boxed{\text{Gradient Flows}}$

Exercise: Extend Dobrushin's estimate to W_2.

$$\partial_t \rho = \nabla \cdot (\rho \nabla W * \rho)$$

$$d X(t) = - \nabla W * \rho_1^t (X(t))\, dt$$

$$\mathrm{Law}(X(t)) = \rho_1(t)$$

$$d Y(t) = - \nabla W * \rho_2^t (Y(t))\, dt$$

$$\mathrm{Law}(Y(t)) = \rho_2(t)$$

$$\frac{d}{dt}|X(t) - Y(t)|^2 = \cdots$$

Under assumptions $\int x \, d\rho_1(0,x) = \int y \, d\rho_2(0,y)$

expect exponential decay

given $\nabla^2 W \geq \lambda \, \mathrm{Id}$.

RK: Classical method for MFL.

- Dobrushin's estimate (Golse's notes)
- Coupling Method (Sznitmann: Topics in Propagation of Chaos)

Follow the paper:

{Euclidean, metric, and Wasserstein} gradient flows:
An overview

by F. Santambrogio

Main part: Wasserstein Gradient Flows.

$(\mathcal{P}_2(\mathbb{R}^d), W_2)$

Main Reference: [AGS]

Ambrosio, Gigli and Savaré:

$\boxed{\text{Gradient Flows in Metric Spaces and in the Spaces of Probability Spaces.}}$

Gradient Flows (Steepest Descent Curves):

$F: X \to \mathbb{R} \cup \{+\infty\}$ $\rightsquigarrow$ Find the global minimum of $F(x)$

(easily $X = \mathbb{R}^n$) (if $F \in C^1$, $\nabla F(x) = 0$)

Consider

$$\begin{cases} \dot{x}(t) = -\nabla F(x(t)) \\ x(0) = x_0 \end{cases}$$

then $x = x(t)$ is a curve starting x_0, trying to minimize the function F as fast as possible.

X : Hilbert Space $\quad X \simeq X'$

Eg: Heat Eq. $\partial_t u = \Delta u$

can be viewed as a gradient flow in L^2-Hilbert space of the Dirichlet energy

$$F(u) = \begin{cases} \frac{1}{2}\int |\nabla u|^2, & \text{if } u \in H^1 \\ +\infty & \text{otherwise} \end{cases}$$

$$\frac{\delta F(u)}{\delta u} = -\Delta u$$

$$\left\langle \frac{\delta F(u)}{\delta u}, \underline{\underline{h}} \right\rangle = \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} F(u+\varepsilon h)$$

$$= \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \frac{1}{2}\int |\nabla u + \varepsilon h|^2$$

$$= \int \nabla u \cdot \nabla h \, dx = -\int \underline{\underline{h}}\, \Delta u \, dx$$

$$\Rightarrow \frac{\delta F(u)}{\delta u} = -\Delta u$$

$$\partial_t u \underset{\Pi}{=} - \underbrace{\text{“}\frac{\delta F(u)}{\delta u}\text{”}}_{\substack{\text{gradient} \\ \text{or 1st order variation}}} = \Delta u$$

Renovated interests around 2000s $\begin{cases} \text{JKO scheme} \\ \text{Otto Calculus} \end{cases}$

Ambrosio et al ~

Continuity Eq. $\begin{cases} \partial_t \rho + \mathrm{div}(\rho v) = 0 \\ v = -\nabla \frac{\delta F(\rho)}{\delta \rho} \end{cases}$

$F: P_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$

(Later)

Today we focus on Euclidean & metric setting.

First, consider gradient flows on Euclidean space $\mathbb{R}^n$, $F: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, C^∞, $x_0 \in \mathbb{R}^n$

A Gradient flow is defined as a curve $x(t)$, with initial position $x(0) = x_0$, as the solution to the Cauchy Problem

$$(*) \begin{cases} x'(t) \; (= \dot{x}(t)) = -\nabla F(x(t)), \text{ for all } t > 0. \\ x(0) = x_0 \end{cases}$$

If $F \in C^{1,1}$ (or $\nabla F \in W^{1,\infty} \Leftarrow$ Lipschitz), then by Cauchy–Lipschitz (ODE well-posedness) theorem, then $(*)$ has unique solution $x = x(t)$.

But existence and uniqueness can hold without this strong regularity.

Assume F is convex (F is real-valued. F is a.e. C^2), then F can be non-differentiable.

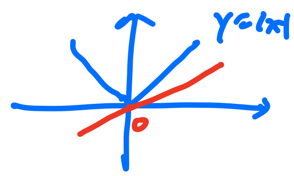
$$\nabla F(x) \longrightarrow \partial F(x)$$

$\uparrow$ subdifferential

We now consider differential inclusion

$x = X(t)$
A.C. curve s.t. $\begin{cases} x'(t) \in -\partial F(x) & \text{for a.e. } t > 0. \\ X(0) = X_0 \end{cases}$

where the sub-differential



$\partial F(x)$

$\triangleq \{ p \in \mathbb{R}^n \mid F(y) \geq F(x) + p \cdot (y-x) \ \forall y \in \mathbb{R}^n \}$

$\begin{cases} \partial F(0) = [-1, 1] \\ F = |x| \end{cases}$

In particular if F is differential at x, then $\partial F(x) = \{ \nabla F(x) \}$

- $\partial F(x)$ is a convex set.
- If F is real valued, then $\partial F(x) \neq \phi$ (or for $y \in \{F < +\infty\}^0$, $\partial F(y) \neq \phi$)

Prop: Suppose
- F is convex;
- Let X_i $(i=1,2)$ be two

solutions of (DI) $\begin{cases} x'(t) \in -\partial F(X(t)) \\ x(0) = x_i^0 \end{cases}$

then $t \mapsto |X_1(t) - X_2(t)|$ is non-increasing.

Consequently, we have uniqueness of the Cauchy problem of the differential inclusion eq. (DI).

Pf: Let $g(t) = \frac{1}{2} |X_1(t) - X_2(t)|^2$

$$\frac{d}{dt} g(t) = (X_1(t) - X_2(t))\big(\underbrace{X_1'(t)}_{= -p_1} - \underbrace{X_2'(t)}_{-p_2}\big) \le 0$$

F is convex $\Rightarrow$ $p_i \in \partial F(X_i(t))$

$$\underbrace{(X_1 - X_2)(p_1 - p_2) \geq 0}$$

$\blacksquare$

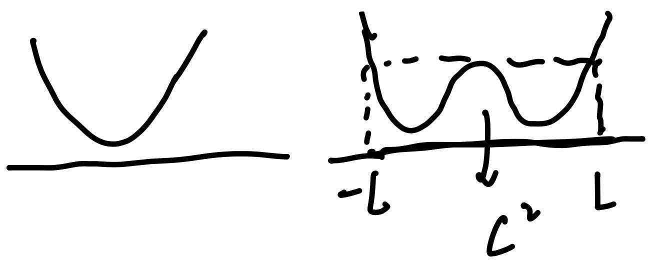
When F is semi-convex, (λ-convex)

if $x \mapsto F(x) - \frac{\lambda}{2}|x|^2$ is convex.

$$\begin{cases} \lambda > 0: & \text{stronger than convex} \\ & \lambda\text{-uniformly convex} \\ \lambda < 0: & \text{weaker than convex.} \end{cases}$$

$\Updownarrow$

$$\boxed{\nabla^2 F \geq \lambda\, Id}$$



Note: On a compact set, any C^2 functions are

λ-convex for some $\lambda < 0$.

For λ-convex F, its sub-differential

$$\partial F(x) = \{ p \in \mathbb{R}^n \mid F(y) \geq F(x) + p\cdot(y-x) + \frac{\lambda}{2}|y-x|^2, \ \forall y \in \mathbb{R}^n \}.$$

$$\left(\quad \tilde{F}(x) = \underbrace{F(x) - \frac{\lambda}{2}|x|^2} \text{ is convex.} \right.$$

$$\tilde{F}(y) \geq \tilde{F}(x) + \underset{\substack{\parallel \\ (p-\lambda x)(y-x)}}{\tilde{p}\cdot(y-x)} \qquad \left. \tilde{p} = \nabla F(x) - \lambda x. \right)$$

As above, if F is λ-convex, then

$$\frac{d}{dt}\frac{1}{2}|x_1(t) - x_2(t)|^2 \qquad \underbrace{-x_1'(t)}_{p_1(t)} \in \partial F(x_1(t))$$

$$= \underbrace{-(x_1(t) - x_2(t))\cdot(p_1(t) - p_2(t))}_{\geq \lambda |x_1(t) - x_2(t)|^2}$$

$$\leq -\lambda |x_1(t) - x_2(t)|^2$$

$$\Rightarrow |x_1(t) - x_2(t)| \leq e^{-\lambda t}|x_1(0) - x_2(0)|$$

$x'(t) \in -\partial F(x(t))$

But if F is λ-convex ($\lambda > 0$)

(F is strictly convex),

$$F(x) \geq F(x_0) + p\cdot(x-x_0) + \frac{\lambda}{2}|x-x_0|^2$$

[F. l.s.c] $\qquad (p \in \partial F(x_0))$

$$\geq F(x_0) + \frac{\lambda}{4}|x-x_0|^2 - C > -\infty.$$

So F admits a $\underbrace{\text{unique}}_{(F \text{ is strictly convex})}$ minimizer.

$$\bar{x} = \underset{x\in\mathbb{R}^n}{\arg\min}\, F(x)$$

then $x(t) = \bar{x}$ is a solution to

$$\begin{cases} x'(t) \in -\partial F(x(t)) \\ x|_{t=0} = \bar{x} \end{cases}$$

Then by the estimate above

$$|x(t) - \bar{x}| \leq e^{-\lambda t}|x(0) - \bar{x}|$$

RK: Actually $x'(t) \in -\partial F(x(t))$,

becomes $x'(t) = -\partial F^0(x(t))$

Here $\partial F^0(x(t))$ is the element in $\partial F(x(t))$ with the minimal norm.

($\Leftarrow$ $p \in \partial F(x(t_0))$

assume $t \mapsto x(t)$ and $t \mapsto F(x(t))$ are differentiable at $t = t_0$.

Since F is λ-convex

$$F(x(t)) \geq F(x(t_0)) + p\cdot(x(t)-x(t_0)) + \frac{\lambda}{2}|x(t)-x(t_0)|^2, \quad \forall t$$

$$\Downarrow$$

("$\geq$" becomes "$=$" when $t = t_0$)

$$h(t) = F(x(t)) - F(x(t_0)) - p\cdot(x(t)-x(t_0)) - \frac{\lambda}{2}|x(t)-x(t_0)|^2$$

$$\left.\frac{d}{dt}\right|_{t=t_0} h(t) = 0$$

$$\Rightarrow \left.\frac{d}{dt}F(x(t))\right|_{t=t_0} = p\cdot x'(t_0) \quad \forall\, p \in \partial F(x(t_0))$$

This means

$$\partial F(x(t_0)) \subset \underbrace{\{p \mid p\cdot x'(t_0) = \text{const}\}}_{\text{hyperplane}}$$

$$x'(t_0) \in -\partial F(X(t_0))$$

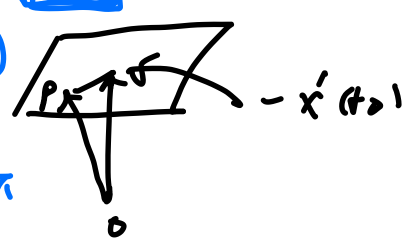
So $\quad -x'(t_0)\, x'(t_0) = p\cdot x'(t_0)$

or $\quad (p+\underline{x'(t_0)})\cdot \underline{x'(t_0)} = 0$

$$\boxed{x'(t_0) = -\partial^0 F(X(t_0))}$$

$$(\underset{\pi}{\underset{\perp}{p}} - (\underbrace{-x'(t_0)}_{\perp}))$$

Hyper π



Discretization in time

Fix a small time step $\tau > 0$

Look for a sequence of points (x_k^τ)

defined via the iterated scheme

(Minimizing Movement Scheme)

$$x_{k+1}^\tau \in \underset{x}{\operatorname{argmin}} \left(F(x) + \frac{|x - x_k^\tau|^2}{2\tau} \right)$$

Here F is l.s.c. and $F(x) \geq C_1 - C_2|x|^2$

(if F is λ-convex (even $\lambda < 0$),

the assumption holds true.)

$\{x_k^\tau\}$: the approximate value of the "limit" $x(t)$

at times $t = 0, \tau, 2\tau, \cdots k\tau, \cdots$

$$x_{k+1}^\tau \in \operatorname{arg\,min} \left(\underbrace{F(x) + \frac{|x - x_k^\tau|^2}{2\tau}} \right)$$

Assume $F \in C^2$,

$\Downarrow$ necessary

$$\nabla F(x_{k+1}^\tau) + \frac{x_{k+1}^\tau - x_k^\tau}{\tau} = 0$$

or

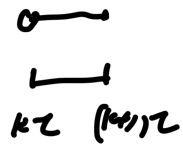
$$\frac{x_{k+1}^\tau - x_k^\tau}{\tau} = - \nabla F(x_{k+1}^\tau) \quad \Phi$$

Need to solve a linear system

This is Implicit Euler Scheme for

$$\dot{x} = -\nabla F(x) \quad ✓$$

(Compare to Explicit Euler Scheme:



$$\frac{x_{k+1}^\tau - x_k^\tau}{\tau} = -\nabla F(x_k^\tau)$$

$$x_{k+1}^\tau = x_k^\tau - \tau \nabla F(x_k^\tau) \;)$$

We know:

$$\dot{x}(t) = -\nabla F(x(t)) \qquad F \in C^2$$

$$\frac{d}{dt} F(x(t)) = \nabla F(x(t))\, \dot{x}(t)$$

$$= -|\nabla F(x(t))|^2 \le 0$$

This property holds for the Implicit Euler Scheme as well.

Recall

$$x_{k+1}^\tau \in \arg\min \Big(\underbrace{F(x) + \frac{|x - x_k^\tau|^2}{2\tau}}\Big)$$

$$F(x_0) = F(x_0^\tau)$$

So

$$F(x_{k+1}^\tau) + \frac{|x_{k+1}^\tau - x_k^\tau|^2}{2\tau} \le F(x_k^\tau)$$

$k = 0, 1, 2, \cdots, L$

If $F(x_0) < +\infty$ and $\inf F > -\infty$,

then summing k from 0 to L

we obtain

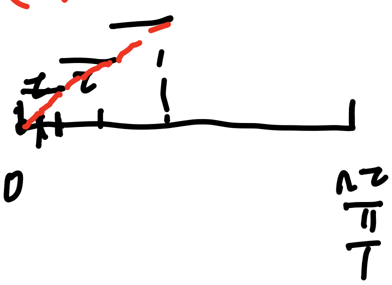
$$F(x^\tau_{L+1}) + \sum_{k=0}^{L} \frac{|x^\tau_{k+1} - x^\tau_k|^2}{2\tau} \le F(x_0)$$

or

$$\boxed{\sum_{k=0}^{L} \frac{|x^\tau_{k+1} - x^\tau_k|^2}{2\tau} \le F(x_0) - F(x^\tau_{L+1}) \le \underset{<+\infty}{C}}$$

$$\Downarrow$$

$$(x_k^\tau) \xrightarrow{\text{interpolation}} \begin{cases} x^\tau(t) \ \checkmark \\ \tilde{x}^\tau(t) \end{cases} \quad \boxed{\text{equi-continuous}}$$



$\Downarrow$ Ascoli's Lemma

Pass to limit.

Two interpolations of (x_k^τ)

$$\begin{cases} x^\tau(t) = x^\tau_{k+1} & \rightarrow \text{(Piece-wise constant)} \\ \quad \text{if } t \in (k\tau, (k+1)\tau], \\ \tilde{x}^\tau(t) = x^\tau_k + (t-k\tau)\, v^\tau_{k+1} & \leftarrow \text{(piecewise linear interpolation)} \end{cases}$$

$$v^\tau(t) = v^\tau_{k+1} = \frac{x^\tau_{k+1} - x^\tau_k}{\tau} = -\nabla F(x^\tau_{k+1})$$

for $t \in (k\tau, (k+1)\tau]$

Easy to check:

$$\begin{cases} (\tilde{x}^\tau)'(t) = v^\tau(t) \\ \tilde{x}^\tau : \text{continuous, piecewise-affine} \\ \quad \text{(hence A.C.)} \end{cases}$$

x^τ : NOT continuous

$$v^\tau(t) = \frac{x^\tau_{k+1} - x^\tau_k}{\tau} \in -\partial F(\underbrace{x^\tau_{k+1}}_{= \tilde{x}^\tau(t)})$$

$t \in [k\tau, (k+1)\tau)$

i.e. $v^\tau(t) \in -\partial F(\bar{x}^\tau(t))$ for any t.

Recall $L = \lfloor T/\tau \rfloor$, $\boxed{F(x^\tau_{k+1}) + \frac{|x^\tau_{k+1} - x^\tau_k|^2}{2\tau} \le F(x^\tau_k)}$

$$\underbrace{\sum_{k=0}^{L} \frac{|x^\tau_{k+1} - x^\tau_k|^2}{2\tau}}_{} \le F(x_0) - F(x^\tau_{L+1}) \le C < \infty$$

$$= \frac{1}{2}\tau \sum_{k=0}^{L} \left(\frac{|x^\tau_{k+1} - x^\tau_k|}{\tau}\right)^2 = \frac{1}{2}\tau \sum_{k=0}^{L} |v^\tau_k|^2$$

$$= \frac{1}{2}\sum_{k=0}^{L}\int_{k\tau}^{(k+1)\tau} |(\tilde{x}^\tau)'(t)|^2 \, dt < C < \infty$$

$$= \boxed{\frac{1}{2}\int_0^T |(\tilde{x}^\tau)'(t)|^2 \, dt < C < \infty}$$

Then by Cauchy-Schwarz

We have $|\tilde{x}^\tau(t) - \tilde{x}^\tau(s)| \lesssim |t-s|^{1/2}$

$$|\tilde{x}^\tau(t) - \tilde{x}^\tau(s)| \le \int_s^t |(\tilde{x}^\tau)'(r)| \, dr$$

$$\le \left(\int_s^t |(\tilde{x}^\tau)'(r)|^2 \, dr\right)^{1/2} |t-s|^{1/2}$$

$$\le \sqrt{2}\, C \; |t-s|^{1/2}$$

Applying Ascoli to $\tilde{x}^{\tau}(t)$

Fix initial x_0 $[0,T]$

i.e. $\left\{\begin{array}{l} \bullet \; |\tilde{x}^{\tau}(t) - \tilde{x}^{\tau}(s)| \le C\,|t-s|^{1/2} \\ \quad \text{(linear)} \qquad \text{(Equi-continuous)} \\ \bullet \; |\tilde{x}^{\tau}(t) - x^{\tau}(t)| \le |x_{k+1}^{\tau} - x_{k}^{\tau}| \le C\sqrt{\tau} \\ \quad t \in [k\tau, (k+1)\tau] \end{array}\right.$

Prop: Let $\tilde{x}^{\tau}$, x^{τ} and v^{τ} be constructed as above using the minimizing movement scheme. Suppose $F(x_0) < +\infty$ and $\inf F > -\infty$. Then up to a subsequence $\tau_j \to 0$, (still denoted as τ), both $\tilde{x}^{\tau}$, x^{τ} converge uniformly to a same curve $x \in H^1$, $\dot{x} = v$ and v^{τ} weakly converges in L^2 to a vector field v, s.t. $x' = v$. and

i) If F is λ-convex, we have
$$v(t) \in -\partial F(x(t)), \quad \text{a.e. } t.$$

ii) If F is C^1, $v(t) = -\nabla F(x(t))$.

$\square$

Start of Lecture 20

Pf of the above proposition:

i) * $\tilde{x}^{\tau}(0) = x_0$ is fixed. $\to$ Uniform bound

* $\frac{1}{2}\int_0^T |(\tilde{x}^{\tau})'(t)|^2 dt \le C < \infty$

$\hookrightarrow$ $|\tilde{x}^{\tau}(t) - \tilde{x}^{\tau}(s)| \le C |s-t|^{1/2}$

i.e. $(\tilde{x}^{\tau}(t))_{t\in[0,T]}$ is $1/2$-Hölder

Then Applying Ascoli's lemma to $(\tilde{x}^{\tau})$ to get a uniform converging subsequence, i.e. $\tilde{x}^{\tau} \to x$ in $L^\infty([0,T])$

Also by

$$|\tilde{x}^{\tau}(t) - x^{\tau}(t)| \le C\,\tau^{1/2},$$

on the same subsequence, $x^{\tau} \to x$ in $L^\infty([0,T])$

where $x = x(t) : [0,T] \to \mathbb{R}^n$.

Then $v^{\tau} = (\tilde{x}^{\tau})'$, a.e. $t \in [0,T]$,

and

$$\frac{1}{2}\int_0^T |(\tilde{x}^{\tau})'(t)|^2 dt \le C < \infty$$

$$= \frac{1}{2}\int_0^T |v^{\tau}(t)|^2 dt \le C < \infty$$

Hence up to an extra subsequence,

$$v^{\tau} \to v \quad \text{weakly in } L^2([0,T])$$

($\langle v^{\tau}, \phi\rangle \to \langle v, \phi\rangle$, for any $\phi \in L^2([0,T])$.

By $\begin{cases} * \ \tilde{x}^\tau, \ x^\tau \to x & \text{in } L^\infty([0,T]) \\ * \ v^\tau \to v & \text{weakly in } L^2([0,T]) \\ * \ (\tilde{x}^\tau)' = v^\tau \end{cases}$

by the distributional limit, $\boxed{v = x'}$

Going back to prove i): i.e. if F is λ-convex, then $v(t) \in -\partial F(x(t))$ a.e. t.

$x(t)$ is a solution of $\dot{x}(t) \in -\partial F(x(t))$

Fix any $y \in \mathbb{R}^n$, since F is λ-convex,

$$(*) \quad F(y) \geq F(x^\tau(t)) - v^\tau(t) \cdot (y - x^\tau(t)) + \frac{\lambda}{2}|y - x^\tau(t)|^2$$

(note $v^\tau(t) \in -\partial F(x^\tau(t))$)

Then multiply by a positive measurable function

$a : [0,T] \to \mathbb{R}_+$ $(a \geq 0)$ and integrate

$$\int_0^T a(t)\Big(F(y) - F(x^\tau(t)) + v^\tau(t) \cdot (y - x^\tau(t)) - \frac{\lambda}{2}|y - x^\tau(t)|^2 \Big)\, dt \geq 0$$

strong $\to L^2$

We can now pass to limit as $\tau \to 0$

$\forall$ any ϕ $\langle f_n, \phi \rangle \to \langle f, \phi \rangle$ $n \to \infty$

$\|g_n - g\|_{L^2} \to 0$

$\Big\} \Rightarrow \langle f_n, g_n \rangle \to \langle f, g \rangle$

For the term $\displaystyle -\!\!\int_0^1 F(x^\tau(t))\, a(t)\, dt$,

we need the l.s.c. of the function F.

($\displaystyle F(x) \le \liminf_{x_n \to x} F(x_n)$)

$$-\int_0^T a(t) F(x(t))\, dt \geq -\int_0^T \liminf_{\tau \to 0} F(x^\tau(t))\, a(t)\, dt$$

$$\geq - \liminf_{\tau \to 0} \int_0^T a(t) F(x^\tau(t))\, dt$$

($\displaystyle \int_E \liminf_{n \to \infty} f_n(x) \le \liminf_{n\to\infty} \int_E f_n$ $\leftarrow$ Fatou) ≥ 0

Hence in the limit $\tau \to 0$, one has

$$\int_0^T a(t) \Big(F(y) - F(x(t)) + v(t)\cdot(y - x(t)) - \frac{\lambda}{2}|y - x(t)|^2 \Big)\, dt \geq 0$$

$$\Downarrow$$

For a.e. t., for fixed y,

$$(*)\quad F(y) \geq F(x(t)) - v(t)\cdot(y - x(t)) + \frac{\lambda}{2}|y - x(t)|^2$$

Using y in dense countable set in $\{F < +\infty\}$,

(where F is continuous)

we get a.e. t, $v(t) \in -\partial F(x(t))$.

ii) $\quad -\nabla F(x^\tau_{(t)}) = v^\tau(t) = (\tilde{x}^\tau)'(t)$

$\qquad\quad \wr$? $\qquad\qquad\qquad \wr$ weakly in L^2

$$-\nabla F(x(t)) \qquad \dot{v}(t)$$

Rk: No rate here. See AGS for convergence rates, usually in order of τ.

($\Leftarrow$ x_k^τ ✓

$$x_{k+1}^\tau \in \arg\min_x \left(F(x) + \frac{|x - x_k^\tau|^2}{2\tau} \right)$$

Modification: x_k^τ ✓

$$x_{k+1}^\tau \in \arg\min_x \left(2F\left(\frac{x + x_k^\tau}{2}\right) + \frac{|x - x_k^\tau|^2}{2\tau} \right)$$

$$\nabla F\left(\frac{x_{k+1}^\tau + x_k^\tau}{2}\right) + \frac{x_{k+1}^\tau - x_k^\tau}{\tau} = 0$$

i.e. $$\frac{x_{k+1}^\tau - x_k^\tau}{\tau} = -\nabla F\left(\frac{x_{k+1}^\tau + x_k^\tau}{2}\right)$$

convergence is of order τ^2.

$\boxed{\text{Metric Setting}}$

(x_k^τ) Iterated Minimizing Scheme

(to define Generalized Minimizing Movement

De Giorgi: the limit of the discrete scheme) $\Leftarrow$ GMM Scheme)

(X, d) Metric Space $\quad d$: metric

$(\mathcal{P}_2(\mathbb{R}^d), W_2)$

$F: X \to \mathbb{R} \cup \{+\infty\}$ l.s.c. (w.r.t. d-topology)

We can also define (given $x_0^\tau = x_0$) ← metric on X.

$$x_k^\tau \in \underbrace{\operatorname*{argmin}_x \left(F(x) + \frac{d(x, x_k^\tau)^2}{2\tau} \right)}$$

$\{x_k^\tau\}_{k \in \mathbb{N}}$. Next Step is to study possible limit when $\tau \to 0$.

Use the piecewise constant interpolation

$$x^\tau(t) := x_k^\tau \quad \text{for every } t \in ((k-1)\tau, k\tau].$$

and then study the limit of x^τ, as $\tau \to 0$

Def: A curve $x: [0, T] \to X$ is called GMM if $\exists\ \tau_j \to 0$, s.t. $x^{\tau_j}(t) \to x(t)$ uniformly in $[0, T]$.

Again

$$F(x_{k+1}^\tau) + \frac{d(x_{k+1}^\tau, x_k^\tau)}{2\tau} \leq F(x_k^\tau)$$

Summing up for $k = 0, 1, \cdots, l$, $\quad l = \lfloor T/\tau \rfloor$.

$$\sum_{k=0}^{L} d(x_{k+1}^{\tau}, x_k^{\tau})^2 \le 2\tau\left(F(x_0) - F(\underbrace{x_{L+1}^{\tau}}_{>-\infty})\right)$$

$$\le \boxed{2\tau C}$$

$$(F > -\infty, \quad F(x_0) < +\infty)$$

Again by Cauchy–Schwarz,

For $t < s$, $\quad t \in [k\tau, (k+1)\tau)$

$s \in [L\tau, (L+1)\tau)$,

$$|L-k| \le |t-s|/\tau + 1$$

$d(x^{\tau}(t), x^{\tau}(s))$ (recall we use constant interpolation)

$$\le \sum_{j=k}^{L} d(x_{j+1}^{\tau}, x_j^{\tau})$$

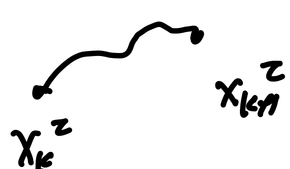
$$\le \underbrace{\left(\sum_{j=k}^{L} d^2(x_{j+1}^{\tau}, x_j^{\tau})\right)^{1/2}}_{\le C\sqrt{\tau}} \underbrace{\left(\sum_{j=k}^{L} 1\right)^{1/2}}_{\left[\frac{|t-s|}{\tau}+1\right]^{1/2}}$$

$$\lesssim |t-s|^{1/2} + \tau^{1/2}$$

→ negligible when $\tau \to 0$.

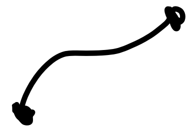
(X, d) need more structure

X: geodesic space



$$x = x(t) \qquad x'(t) = \lim_{h\to 0} \frac{x(t+h) - x(t)}{h}$$

A curve
$$w = w(t) : [0,1] \to X$$



we cannot define "$w'(t)$"

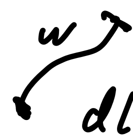
but we can define the "speed" $|w'|(t)$

Metric Derivative (the modulus of the velocity)

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

provided the limit exists.

$w: [0,T] \to X$



$d(w(t), w(s)) \le L|t-s|$

Rudmacher Theorem:

if w is Lipschitz, then the metric derivative $|w'|(t)$ exists for a.e. t. Also we have,

for $t_0 < t_1$,
$$d(w(t_0), w(t_1)) \le \boxed{\int_{t_0}^{t_1} |w'|(s)\, ds.}$$

Def: (Absolutely Continuous Curve in X)

We call a curve $w: [0,1] \to X$ is A.C.,

if $\exists\, g \in L^1([0,1])$, s.t.
$$d(w(t_0), w(t_1)) \le \int_{t_0}^{t_1} g(t)\, dt,$$
for all $t_0 < t_1$

$AC(X) = \{ w : [0,1] \to X \mid w \text{ is AC.} \}$

Length:

For $w : [0,1] \to X$, define

$$\text{Length}(w) := \sup \left\{ \sum_{k=0}^{n-1} d(w(t_k), w(t_{k+1})) \;\middle|\; \begin{array}{l} n \geq 1 \\ 0 = t_0 < t_1 < \cdots < t_n = 1 \end{array} \right\}$$

For $w \in AC(X)$,

then $\text{Length}(w) \leq \int_0^1 s(t)\,dt < \infty$

Prop: Given an A.C. curve, $w : [0,1] \to X$, we have

$$\text{Length}(w) = \int_0^1 |w'|(t)\,dt.$$

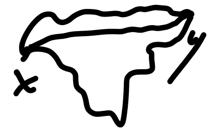
Def: A geodesic is a length minimizing curve.

i.e. a curve $w : [0,1] \to X$ is said to be a geodesic between $x_0 \in X$ and $x_1 \in X$, if

$$w(0) = x_0, \quad w(1) = x_1$$

and $\text{Length}(w) = \min \left\{ \text{Length}(\tilde{w}) \;\middle|\; \begin{array}{l} \tilde{w}(0) = x_0 \\ \tilde{w}(1) = x_1 \end{array} \right\}.$

(X, d) : <u>length space</u>



$\forall x, y \in X$,

$$d(x,y) = \underset{\substack{\Updownarrow \\ \min}}{\inf} \{ \text{Length}(w) : w \in AC(X),\ w(0) = x,\ w(1) = y \}.$$

(X, d) $\boxed{\text{geodesic space}}$

Energy Dissipation Equality.

and **Evolution Variational Inequality.** EDE EVI

Do calculations in Euclidean and smooth setting.

first

$$\dot{x}(t) = -\nabla F(x(t))$$

$0 \le s \le t \le T$ $\qquad F \in C^2$

$$F(x(s)) - F(x(t)) \qquad (\text{purely F.T.C.})$$

$$= -\int_s^t \nabla F(x(r))\, x'(r)\, dr \qquad \boxed{x'(r) = -\nabla F(x(r))}$$

$$\le \frac{1}{2}\int_s^t \underbrace{|x'(r)|^2}_{|x'|} dr + \frac{1}{2}\int_s^t \underbrace{|\nabla F(x(r))|^2}_{\text{slope } |\nabla^- F|} dr$$

"=" ← when take inf "=" ?

"=" iff $x'(r) = -\nabla F(x(r))$ $\forall r \in [s,t]$.

(Indeed: Right $-$ Left $= \frac{1}{2}\int_s^t |\nabla F(x(r)) + x'(r)|^2 dr$)

The "=" condition is called

Energy Dissipation Equality (EDE):

$$\begin{cases} F(x(s)) - F(x(t)) \underset{(\geq)}{=} \frac{1}{2}\int_s^t |x'(r)|^2 dr \\ (s \leq t) \qquad\qquad + \frac{1}{2}\int_s^t |\nabla F(x(r))|^2 dr, \\ \qquad\qquad \Updownarrow \\ x'(t) = -\nabla F(x(t)) \quad \text{a.e.} \end{cases}$$

$F: X \to \mathbb{R} \cup \{+\infty\}$

$$\begin{cases} |\nabla^- F|(x) := \limsup_{y \to x} \frac{(F(x) - F(y))_+}{d(x,y)} \\ |x'|(t) := \lim_{h \to 0} \frac{d(x(t+h), x(t))}{|h|}, \end{cases}$$

$\boxed{\text{Lecture 21}}$

Another characterization of Gradient Flows

If $F:\mathbb{R}^d \to \mathbb{R}$ is convex, then

$$F(y) \geq F(x) + p\cdot(y-x), \quad \forall y \in \mathbb{R}^d$$

(Characterization of $p \in \partial F(x)$,

if $F \in C^1$, $p = \nabla F(x)$.)

If F is λ-convex, the inequality that characterizes the gradient is

$$F(y) \geq F(x) + p\cdot(y-x) + \frac{\lambda}{2}|y-x|^2, \quad \forall y \in \mathbb{R}^d.$$

Hence we can pick a curve

$$x = x(t)$$

and a point y

Compute

$$\frac{d}{dt}\frac{1}{2}|x(t)-y|^2 = (y - x(t))(-x'(t))$$

$$-x'(t) = \nabla F(x(t)) \in \partial F(x(t))$$

Then

$$F(y) \geq F(x(t)) - x'(t)(y - x(t)) + \frac{\lambda}{2}|y - x(t)|^2$$

$$\frac{d}{dt}\frac{1}{2}|x(t)-y|^2 \leq F(y) - F(x(t)) - \frac{\lambda}{2}|y - x(t)|^2$$

memorize

π

EVI : Evolution Variational Inequality
(or more precisely EVI_λ)

All terms have metric setting counterpart

$$|x(t)-y|^2 \Rightarrow d(x(t),y)^2$$

Easy to show uniqueness and stability:
Take two curves $x(t)$, $y(t)$, (EVI_λ)

$$\frac{d}{dt}\frac{1}{2}d(x(t),y(s))$$

$$\leq F(y(s)) - F(x(t)) - \frac{\lambda}{2}d(x(t),y(s))^2 \quad ①$$

$+$

$$\frac{d}{ds}\frac{1}{2}d(x(t),y(s))^2 \quad ②$$

$$\leq F(x(t)) - F(y(s)) - \frac{\lambda}{2}d(x(t),y(s))^2$$

$\Downarrow$

$$\Rightarrow \quad \frac{d}{dt}\frac{1}{2}d(x(t),y(t))^2 \leq -\lambda d(x(t),y(t))^2$$

✓ EDE

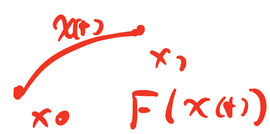
Uniqueness / Stability i.e. $\frac{d}{dt}E(t) \leq -2\lambda E(t)$, ✓ $\boxed{EVI_\lambda}$

$$E(t) = d(x(t),y(t))^2$$

§7. General theory in $\boxed{\text{metric space}}$.

Notions in metric setting. $v(t) = \lim_{h \to 0} \frac{x(t+h) - x(t)}{h}$

* Speed of a curve (Metric derivative)
* Slope of a function
  (~ modulus of its gradient) $f: \mathbb{R}^d \to \mathbb{R}$, $\vec{x}$
* geodesic convexity



Metric derivative: ← not a vector space

A curve $w : [0,T] \to (X, d)$

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

provided the limit exists.

Slope and modulus of the gradient:

Upper gradient of F is a function $g: X \to \mathbb{R}$, s.t. $\forall$ Lipschitz curve w, we have

$$|F(w(0)) - F(w(t))|$$

$$\leq \int_0^1 g(w(t))\, |w'|(t)\, dt .$$

$F : X \to \mathbb{R} \cup \{+\infty\}$

(RK: If $F \in W^{1,\infty}$,

one can choose $g = |\nabla F|(x)$, i.e.

$$|\nabla F|(x) := \overline{\lim_{y \to x}} \frac{|F(x) - F(y)|}{d(x,y)} .)$$

* Descending Slope.

(adapted to the minimization of a function).

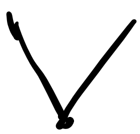
For a function F, which is l.s.c.,

$$|\nabla^- F|(x) := \limsup_{y \to x} \frac{(F(x) - F(y))_+}{d(x,y)}$$

(at local minimum point x_0,

$$|\nabla^- F|(x_0) = 0 \;)$$

$F(x) = |x|$



at $x = 0$ $\quad |\nabla^- F|(0) = 0$

$$|\nabla F|(0) = \overline{\lim_{y \to 0}} \frac{|0 - F(y)|}{|y|} = 1 .$$

i.e. in general $|\nabla^- F|(x) \neq |\nabla F|(x)$.

* Geodesic convexity $\Leftarrow$ convex along a geodesic

Assume further (X, d) is a geodesic space.

$\forall\, (x(0), x(1))$, $\exists$ a geodesic x with constant speed, connecting $x(0)$ and $x(1)$, s.t.

$$F(x(t)) \leq (1-t) F(x(0)) + t F(x(1)).$$

λ-convex: $\left(- \lambda \frac{t(1-t)}{2} d^2(x(0), x(1)) \right)$.

<u>Existence of Gradient Flows</u> (EDE)

Fix any $\tau > 0$, construct sequential minimization along a discrete scheme along a discrete scheme.

$$x^\tau_{k+1} \in \arg\min_{x \in X} \left(\underbrace{F(x) + \frac{d^2(x, x^\tau_k)}{2\tau}} \right)$$

Given

- (X, d)
- $F : X \to \mathbb{R} \cup \{+\infty\}$
- $\underline{\forall c \in \mathbb{R},\ \{F \leq c\} \text{ compact in } X}$.

Iterated Scheme gives

$$(*) \quad F(x_{k+1}^{\tau}) + \frac{d^2(x_{k+1}^{\tau}, x_k^{\tau})}{2\tau} \le F(x_k^{\tau})$$

(No optimality condition used.)

This estimate is not sufficient to characterize the limit curve.

We shall exploit how much x_{k+1}^{τ} is better than x_k^{τ}.

De Giorgi: "Variational interpolation" between x_k^{τ} and x_{k+1}^{τ}.

Fix x_k^{τ}, introduce $\underline{\underline{\theta}} \in (0,1]$.

Consider $\displaystyle \min_x \left(F(x) + \frac{d^2(x, x_k^{\tau})}{2\theta\tau} \right)$

$x(\theta)$: minimizer

$\varphi(\theta)$: the minimal value.

As $\theta \to 0^+$, $x(\theta) \to x_k^{\tau}$.

$\varphi(\theta) \to F(x_k^{\tau})$

for $\theta = 1$, we recover the original problem with minimizer x_{k+1}^{τ}.

Of course $\varphi(\theta) \downarrow$. (non-increasing)

hence a.e. differentiable.

$$\varphi'(\theta) = \frac{d}{d\theta}\left(\underbrace{\left(\theta \mapsto F(x) + \frac{d^2(x, x_k^{\tau})}{2\theta\tau}\right)}\Big|_{x = x(\theta)} \right)$$

Write:

$$g(x, \theta) = \underbrace{F(x) + \frac{d^2(x, x_k^{\tau})}{2\theta\tau}}$$

$$\varphi(\theta) = g(x(\theta), \theta)$$

$$\frac{d\varphi(\theta)}{d\theta} = \frac{d}{d\theta} g(x(\theta), \theta) = \underbrace{\nabla g(x(\theta), \theta)}_{= 0} \cdot x'(\theta) + \frac{\partial g}{\partial \theta}\Big|_{x = x(\theta)}$$

$$= \frac{\partial g}{\partial \theta}\Big|_{x = x(\theta)}.$$

i.e.

$$\boxed{\varphi'(\theta) = -\frac{d^2(x(\theta), x_k^{\tau})}{2\theta^2 \tau}}$$

τ

Recall ↓ Descending slope

$$|\nabla F^-|(x) = \limsup_{y \to x} \frac{[F(x) - F(y)]_+}{d(x,y)} \qquad \pi$$

Claim: $\boxed{|\nabla^- F|(x(\theta)) \leq \dfrac{d(x(\theta), x_k^\tau)}{\theta\tau}}$

Pf: Consider minimization of a function

$$x \longmapsto F(x) + c\, d^2(x, \bar{x})$$

for fixed $c > 0$ and $\bar{x}$.

$$\begin{cases} c = \dfrac{1}{2\theta\tau} \\ \bar{x} = x_k^\tau \end{cases}$$

Consider a competitor y. If x is optimal, then

$$F(x) + c\,d^2(x, \bar{x}) \leq F(y) + c\, d^2(y, \bar{x}),$$

which implies

$$\begin{aligned} F(x) - F(y) &\leq c\left(d^2(y, \bar{x}) - d^2(x, \bar{x})\right) \\ &= c\left(d(y, \bar{x}) + d(x, \bar{x})\right) \\ &\qquad \cdot \left(d(y, \bar{x}) - d(x, \bar{x})\right) \\ &\leq c\left(d(y, \bar{x}) + d(x, \bar{x})\right) d(y, x) \end{aligned}$$

Hence

$$\frac{(F(x) - F(y))_+}{d(y, x)} \leq c\left(d(y, \bar{x}) + d(x, \bar{x})\right)$$

$$\limsup_{y \to x} \frac{(F(x) - F(y))_+}{d(x,y)} \leq 2c\, d(x, \bar{x})$$

This gives:

$$|\nabla^- F|(x(\theta)) \leq \frac{d(x(\theta), x_k^\tau)}{\theta \tau}.$$

For the function φ,

(A) $\varphi(0) - \varphi(1) \geq -\int_0^1 \underbrace{\varphi'(\theta)}_{\leq 0}\, d\theta$ ⇐

("$\geq$" due to possible singular part of the derivative for monotone functions)

$$-\varphi'(\theta) = \frac{d(x(\theta), x_k^\tau)^2}{2\theta^2 \tau} = \frac{\tau}{2} |\nabla^- F(x(\theta))|^2$$

Now.

(A) above implies:

$$F(x_k^\tau) - \left(F(x_{k+1}^\tau) + \frac{d(x_{k+1}^\tau, x_k^\tau)^2}{2\tau} \right)$$

$$\geq \frac{\tau}{2} \int_0^1 |\nabla^- F(x(\theta))|^2\, d\theta.$$

$$\Rightarrow F(x_{k+1}^{\tau}) + \frac{d(x_{k+1}^{\tau}, x_k^{\tau})^2}{2\tau} \le F(x_k^{\tau}) - \frac{\tau}{2}\int_0^1 |\nabla^- F(x(\theta))|^2 \, d\theta$$

Summing up for $k = 0, 1, 2, \cdots, \lfloor t/\tau \rfloor$,

$$\underbrace{F(x_L^{\tau})}_{F(x(t))} + \underbrace{\frac{1}{2}\sum_{k=0}^{L-1}\int_{k\tau}^{(k+1)\tau} \overbrace{\left(\frac{d(x_{k+1}^{\tau}, x_k^{\tau})}{\tau}\right)^2}^{|x'|(t)} dr}$$

$$\le F(x_0) - \underbrace{\frac{1}{2}\sum_{k=0}^{L-1}\int_{k\tau}^{(k+1)\tau} dr \left(\int_0^1 |\nabla^- F(x_k^{\tau}(\theta))|^2 \, d\theta\right)}$$

As $\tau \to 0$,

we can prove for every Generalized Minimizing Movement $x = x(t)$,

$$F(x(t)) + \frac{1}{2}\int_0^t |x'(r)|^2 \, dr + \frac{1}{2}\int_0^t |\nabla^- F|^2 (x(r)) \, dr$$

$\boxed{\begin{matrix} s \\ 0 \end{matrix}}$

$$\le F(x(0))$$

under some assumptions to check.

$$\begin{cases} * \ \text{l.s.c. of } F \\ * \ \text{l.s.c. of } |\nabla^- F| . \end{cases}$$

Recall EDE: For $s \le t$

$$F(x(s)) - F(x(t))$$
$$= \int_s^t -\nabla F(x(r)) \cdot x'(r)\, dr$$
$$\le \frac{1}{2}\int_s^t |x'(r)|^2 dr + \frac{1}{2}\int_s^t |\nabla F(x(r))|^2 dr$$

$$\boxed{\overset{?}{=}}$$

EDE (equivalent to "$\geq$")

We also write it as for $s \le t$

$$F(x(t)) + \frac{1}{2}\int_s^t |x'(r)|^2 dr + \frac{1}{2}\int_s^t |\nabla F(x(r))|^2 dr$$

(This is EDE) $\le \underline{F(x(s))}$

Energy Dissipation Equality

We have obtained

$$F(x(t)) + \frac{1}{2}\int_0^t |x'(r)|^2 dr + \frac{1}{2}\int_0^t |\nabla^- F|^2 (x(r))\, dr$$

$(\forall t$, but $s = 0)$ "$=$" $\le F(x(0))$

This is indeed equality

• $F(x(0)) - F(x(t)) \le \int_0^t |\nabla^- F(x(r))| \, |x'|(r) \, dr$

(Since $\overline{\lim_{s \to t}} \dfrac{(F(x(t)) - F(x(s)))_+}{|s-t|}$

$\le \overline{\lim_{s \to t}} \dfrac{(F(x(t)) - F(x(s)))_+}{d(x(s), x(t))} \cdot \dfrac{d(x(s), x(t))}{|s-t|}$

$\Uparrow \qquad \Uparrow$ Exists. a.e. t.

$\le |\nabla^- F|(x(t)) \, |x'|(t)$.)

$F(x(0)) - F(x(t))$

$\le \frac{1}{2}\int_0^t |x'|^2(r) \, dr + \frac{1}{2}\int_0^t |\nabla^- F|^2(x(r)) \, dr$,

$\Rightarrow$ Combining with the inequality we obtain via De Giorgi variational interpolation.

$F(x(t)) + \frac{1}{2}\int_0^t |x'|(r)^2 \, dr + \frac{1}{2}\int_0^t |\nabla^- F|^2(x(r)) \, dr$

$= F(x(0))$

$F(x(s)) + \frac{1}{2}\int_0^s \cdot \quad + \frac{1}{2}\int_0^s \cdot \quad \shortparallel$

$\Rightarrow$ $\boxed{EDE}$ $\circledast$

$$F(x(s)) = F(x(t)) + \frac{1}{2}\int_s^t |x'|^2(r)\, dr + \frac{1}{2}\int_s^t |D^- F|^2(x(r))\, dr.$$

RK: (*) The EDE condition is not in general sufficient to guarantee uniqueness of gradient flow

Eg: $X = \mathbb{R}^2$ with metric L^∞

$$d((x_1, x_2), (y_1, y_2)) = \max\{|x_1 - y_1|, |x_2 - y_2|\}$$

Take $F(x_1, x_2) = x_1$

Consider $x(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}$, with $x_1'(t) = -1$, $|x_2'(t)| \leq 1$ ✓

What EDE means now?

For $s \leq t$,

$$\underbrace{F(x(t))}_{\checkmark} + \frac{1}{2}\int_s^t \underbrace{|x'|^2(r)}_{\mathbb{R}^2}\, dr + \frac{1}{2}\int_s^t \underbrace{|D^- F|}(x(r))^2\, dr$$

$\leq F(x(0))$

In this example, the 2nd dimension does not matter.

(*) While generally, it is hard to prove existence of gradient flows in the sense of $EVI_{(\lambda)}$, but uniqueness / stability is almost trivial.

<u>Evolution PDE</u>

- $\partial_t \rho + \mathrm{div}(\rho\, \underset{\sim}{v[\rho]}) = \sigma \Delta \rho$

  ρ: a probability density

  $$\partial_t \rho + \mathrm{div}(\rho(-\nabla V - \nabla W * \rho)) = \sigma \Delta \rho$$
  $$= \sigma \Delta \rho^m$$

  Aggregation – Convection – Diffusion

- Kinetic Equations

- Vlasov - Poisson Eq.

$$\begin{cases} \partial_t f + v \cdot \nabla_x f + E \cdot \nabla_v f = 0 \\ E = -\nabla_x \phi \\ -\Delta_x \phi = \int f \, dv \end{cases}$$

where $f = f(t, x, v) \in L^1_{x,v}$.

$f \in \mathcal{P}(\mathbb{R}^3 \times \mathbb{R}^3)$, $\int_{\mathbb{R}^3 \times \mathbb{R}^3} f(t,x,v) \, dx \, dv = 1$.

- Boltzmann Eq.

$$\partial_t f + v \cdot \nabla_x f + \underset{\text{electrostatic force}}{E} \cdot \nabla_v f = \underbrace{Q(f,f)}_{\text{collision kernel}}$$

$f = f(t, x, v)$

space-homogeneous $f = f(t, v)$

$$\begin{cases} \partial_t f = Q(f,f) \\ \qquad = \int_{S^2} \int_{\mathbb{R}^3} Q(v, v', \omega) [f(v'^*) f(v^{*\prime}) - f(v) f(v')] \, dv' \, d\omega \end{cases}$$

Villani's book.

- Landau Eq.

$$\partial_t f + v \cdot \nabla_x f + E \cdot \nabla_v f = \mathcal{L}(f,f)$$

$f = f(t, x, v)$

space-homogeneous Landau

$$\partial_t f = \mathcal{C}(f,f)$$ ← where $\mathcal{C}(\cdot,\cdot)$ is the Landau collision kernel.

Note where $v, v^* \in \mathbb{R}^3$ ($\mathbb{R}^d$)

$$\mathcal{C}(f,f)(v)$$

$$= \nabla_v \cdot \int_{\mathbb{R}^3} A(v-v^*)\left[f(v_*)\nabla_v f(v) - f(v)\nabla_{v_*} f(v_*)\right] dv_*$$

$$= \nabla_v \cdot \int_{\mathbb{R}^3} A(v-v^*)\left[\nabla_v \log f(v) - \nabla_{v_*} \log f(v_*)\right] f(v) f(v_*)\, dv_*.$$

where the Landau kernel

$$A(z) = |z|^{\gamma+2}\left(Id - \frac{z\otimes z}{|z|^2}\right)$$

(3×3-matrix)

$\gamma \in [-d-1, 1]$.

In 3D, $\gamma = -3$ is the Coulomb case.

$\mathbb{R}^3$, $A(z) = \frac{1}{|z|}\left(Id - \frac{z\otimes z}{|z|^2}\right)$.

In the easiest form

$$\begin{cases} \partial_t f = \mathcal{C}(f,f) \\ f = f(t,v), \quad v \in \mathbb{R}^3. \quad \gamma = -3. \end{cases}$$

i.e.

Space-homogeneous Landau in 3D with Coulomb interactions.

RK: ① Landau eq. as a continuity eq.

$$C(f,f) = \nabla_v \cdot \int_{\mathbb{R}^3} A(v-v_*)[\nabla_v \log f(v) - \nabla_{v_*} \log f(v_*)] \underline{f(v) f(v_*)} \, dv_*$$

$$\partial_t f = C(f,f)$$

$$\partial_t f + \operatorname{div}_v \left(\boxed{\int_{\mathbb{R}^3} A(v-v_*)[\nabla_v \log f(v) - \nabla_{v_*} \log f(v_*)] f(v_*) \, dv_*} \, f(v) \right) = 0$$

$$\partial_t f + \operatorname{div}_v \left(f(v) \underbrace{V[f](t,v)}_{\text{velocity field}} \right) = 0$$

Π

$-\nabla \dfrac{\delta E}{\delta f}$

This is a continuity eq.

② Rosenbluth form:

$$\partial_t f = \mathcal{C}(f,f)$$

diffusion $\quad$ drift

$$= \nabla_v \cdot (A_f \nabla f - a_f f),$$

where

$$a_f = \int_{\mathbb{R}^3} A(v-v_*)\, \nabla_{v_*} f(v_*)\, dv_*,$$

$$A_f = \int_{\mathbb{R}^3} A(v-v_*)\, f(v_*)\, dv_*,$$

and $\quad A_f = D^2 G, \quad a_f = \nabla H,$

$$-\Delta H = f, \qquad \Delta G = H$$

Writing Landau as Gradient Flow in Wasserstein Space:

$$Q(f,f) = \nabla_v \cdot \left(\underbrace{\int_{\mathbb{R}^3} A(v-v_*)\left(\nabla_v \log f - \nabla_{v_*} \log f(v_*)\right) f_*\, dv_*} \; f(v) \right)$$

$$V[f](t,v) = \int_{\mathbb{R}^3} A(v-v_*)\left[\underbrace{\nabla_v \log f}_{\Vert} - \underbrace{\nabla_{v_*} \log f}_{\Vert}\right] f_*\, dv_*$$

$$\underbrace{\nabla_v \frac{\delta E}{\delta f}}_{\pi} \qquad \underbrace{\nabla_{v_*} \frac{\delta E_*}{\delta f_*}}_{\pi}$$

$$E(f) = \int \underbrace{f \log f}\, dv$$

$$\frac{\delta E}{\delta f} = \log f$$

$$\partial_t u = \Delta u = \operatorname{div}\left(u \cdot \nabla \underbrace{\log u}_{\frac{\delta E}{\delta u}} \right)$$

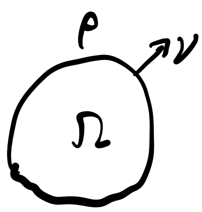
$\downarrow$

$$V = -\nabla \frac{\delta E}{\delta u}$$

Gradient Flows in the Probability Space (endowed with Wasserstein metric)

We here focus on the heat eq. and JKO scheme.

$$\star\begin{cases} \partial_t \rho = \Delta \rho & (t,x) \in [0,+\infty) \times \Omega \leftarrow \text{bounded convex domain} \\ \rho|_{t=0} = \rho_0(x) \\ \frac{\partial \rho}{\partial \nu}\Big|_{\partial\Omega} = 0 \end{cases}$$



Implicit Euler Scheme: Fix any $\tau > 0$, $\rho_0^\tau = \rho_0$,

given ρ_k^τ, ρ_{k+1}^τ is defined as

JKO: $$\rho_{k+1}^\tau \in \underset{\rho \in \mathcal{P}(\Omega)}{\operatorname{argmin}} \left(\int_\Omega \rho \log \rho \, dx + \frac{W_2^2(\rho, \rho_k^\tau)}{2\tau} \right).$$

(The following discussion also works for Fokker-Planck or general evolution PDEs)

Assume $\begin{cases} \bullet \int_\Omega \rho_0(x)\, dx = 1, \quad \rho_0(x) \geq 1 \quad (\text{i.e. } \rho_0 \in \mathcal{P}(\Omega)) \\ \bullet \int_\Omega \rho_0 \log \rho_0 \, dx < +\infty \end{cases}$

We define $(\rho_k^\tau)_k$ according to JKO scheme

Goal: Show that as $\tau \to 0$, the scheme will converge to the solution of the heat eq.

Step I: Existence of Discrete Solutions

Lemma: $\forall k \geq 0$, ρ_{k+1}^{τ} exists.

(i.e. $\mathcal{P}(\Omega) \ni \rho \longmapsto \int_{\Omega} \rho \log \rho \, dx + \frac{1}{2\tau} W_2^2(\rho, \rho_k^{\tau})$ has a minimizer)

Pf: Fix $k \geq 0$.

Choose $(\rho_m)_{m \in \mathbb{N}} \subset \mathcal{P}(\Omega)$ as a minimizing sequence, i.e.

$$\int \rho_m \log \rho_m + \underbrace{\frac{1}{2\tau} W_2^2(\rho_m, \rho_k^{\tau})} \longrightarrow \inf.$$

$\forall M = 1, 2, 3, \cdots$, $\{\rho_m \wedge M\}_{m=1}^{\infty}$ is bounded in $L^{\infty}(\Omega)$, thus by Banach-Alaoglu theorem, it is weak-$*$ compact in L^{∞}.

$\exists$ subsequence m_i (independent of M), s.t.

$$\rho_{m_i} \wedge M \xrightarrow{*} \rho_M \quad \text{in } L^{\infty}(\Omega)$$

Also $\forall s \geq 0$, $s \log s + 1 \geq 0$, we have

$$\int_{\Omega} (\rho_m - \rho_m \wedge M) \, dx$$

$$= \int_{\{\rho_m \geq M\}} (\rho_m - M) \, dx \qquad \underline{(s \log s + 1) \geq 0}$$

$$\leq \frac{1}{\log M} \int_{\{\rho_m \geq M\}} \rho_m \log \rho_m \, dx$$

$$\leq \frac{1}{\log M} \int_{\Omega} (\rho_m \log \rho_m + 1) \, dx \leq \frac{C}{\log M}$$

Set $\rho_\infty := \underset{\text{def.}}{\overset{\nearrow}{\sup_M}} \rho_M$ (This is $\underset{\rho_\infty}{\overset{\shortparallel}{\rho_{k+1}^{\tau}}}$)

We have obtained:

$$\begin{cases} \cdot\ \rho_{m_k} \wedge M \xrightarrow{*} \rho_M \\ \cdot\ \rho_M \xrightarrow{L^1} \rho_\infty \text{ (Monotone convergence)} \end{cases}$$

$$\cdot\ \|\rho_{m_k} \wedge M - \rho_{m_k}\|_{L^1} \leq \frac{C}{\log M}.$$

Combining all three facts, we have

$$\rho_{m_{k_M}} \longrightarrow \rho_\infty \quad \text{in } L^1(\Omega)$$

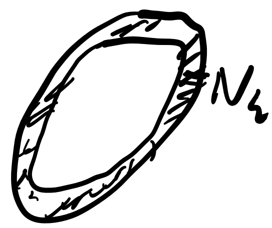
Then the remaining is to show

$$\begin{cases} \cdot\ \rho_\infty \in \mathcal{P}(\Omega) \text{ (no mass loss at the boundary } \partial\Omega) \\ +\ \text{weak}^* \text{ convergence} \end{cases}$$

$\Downarrow$

then the convergence is narrow convergence.

$\Big(\Leftarrow$ idea:



$$N_\varepsilon = \{x \in \Omega \mid dist(x, \partial\Omega) < \varepsilon\}$$

$$|N_\varepsilon| \leq C\varepsilon$$

$$L = \frac{1}{\varepsilon |\log \varepsilon|}$$

$$\int_{N_\varepsilon} \rho_m \leq \int_{N_\varepsilon \cap \{\rho_m \leq L\}} \rho_m$$

$$+ \int_{N_\varepsilon \cap \{\rho_m \geq L\}} \rho_m \frac{\log \rho_m}{\log L}$$

$$\leq L |N_\varepsilon| + \frac{C}{\log L} \leq \frac{C}{|\log \varepsilon|}$$

$$\Rightarrow \int_{\Omega \setminus N_\varepsilon} \rho_m \geq 1 - \frac{C}{|\log \varepsilon|}$$

$$\overset{M \to \infty}{\Rightarrow} \int_{\Omega \setminus N_\varepsilon} \rho_\infty \geq 1 - \frac{C}{|\log \varepsilon|} \Big)$$

Now $\rho_{m_{l_M}} \xrightarrow{M \to \infty} \rho_\infty$ narrowly

Prokhorov's theorem: $\{\rho_{m_{l_M}}\}$ is tight.

Further, $\varphi(s) = s\log s$ is convex, we have

$$(*)\ \int_\Omega \rho_\infty \log \rho_\infty \le \liminf_{M\to\infty} \int_\Omega \rho_{m_{l_M}} \log \rho_{m_{l_M}}.$$

(We use l.s.c. of the entropy w.r.t. the narrow topology).

Similarly

$$(*)\ W_2^2(\rho_\infty, \rho_k^\tau) \le \liminf_{M\to\infty} \frac{1}{2\tau} W_2^2(\rho_{m_{l_M}}, \rho_k^\tau)$$

Hence

$$\inf = \liminf_{M\to\infty} \left(\frac{W_2^2(\rho_{m_{l_M}}, \rho_k^\tau)}{2\tau} + \int_\Omega \rho_{m_{l_M}} \log \rho_{m_{l_M}} \right)$$

$$\ge \frac{W_2^2(\rho_\infty, \rho_k^\tau)}{2\tau} + \int_\Omega \rho_\infty \log \rho_\infty$$

i.e. ρ_∞ is a minimizer,

we can set $\rho_{k+1}^\tau = \rho_\infty$

Step II: $\underbrace{\rho_{k+1}^{\tau} \text{ satisfy some optimality condition}}$

(Minimality eq.)

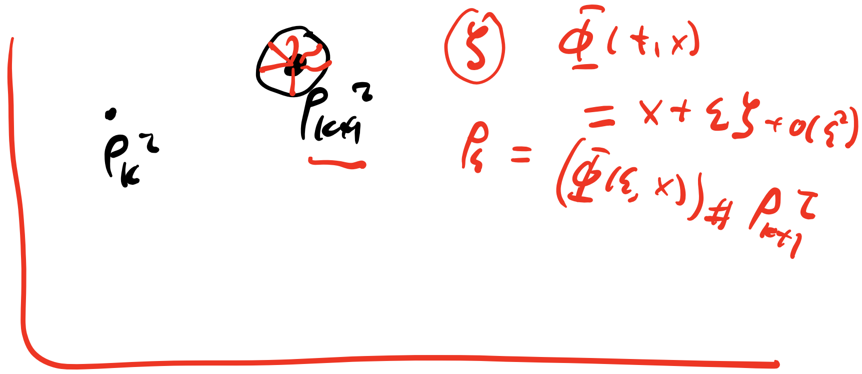
Lemma:

$\forall\, \xi \in C^{\infty}(\Omega, \mathbb{R}^d)$, tangent to $\partial\Omega$,

it holds that

$$\int_{\Omega} \rho_{k+1}^{\tau} \operatorname{div}(\xi)\, dx = \frac{1}{\tau} \int_{\Omega} \langle \xi \circ T_{k+1},\, T_{k+1} - x \rangle \rho_k^{\tau}\, dx.$$

where $T_{k+1} : \Omega \to \Omega$ is the optimal map from ρ_k^{τ} to ρ_{k+1}^{τ}.



Pf: Consider the flow of ξ:

$$\begin{cases} \dot{\Phi}(t,x) = \xi(\Phi(t,x)) \\ \Phi(0,x) = x \end{cases}$$

Since $\xi \parallel \partial\Omega$, $\Phi(t):\Omega \to \Omega$ is a diffeomorphism.

Define

$$\rho_\varepsilon := \Phi(\varepsilon)_\# \rho^\tau_{k+1} \in \mathcal{P}(\Omega)$$

and we have

$$\rho^\tau_{k+1}(x) = \rho_\varepsilon(\Phi(\varepsilon, x)) \det(\nabla\Phi(\varepsilon, x))$$

then

$$\int \rho_\varepsilon(y) \log \rho_\varepsilon(y)\, dy \underset{(\Phi(\varepsilon))_\# \rho^\tau_{k+1}}{=} \int_\Omega \rho^\tau_{k+1}(x) \log \rho_\varepsilon(\Phi(\varepsilon, x))\, dx.$$

$$= \int_\Omega \rho^\tau_{k+1}(x) \log\left(\frac{\rho^\tau_{k+1}(x)}{\det \nabla\Phi(\varepsilon, x)}\right) dx$$

$\Phi = x + \varepsilon\xi + \cdots$

Then $(\det \nabla\Phi(\varepsilon, x) = 1 + \varepsilon \operatorname{div}\xi + O(\varepsilon^2))$

$$\text{(A)}: \int_\Omega \rho_\varepsilon \log \rho_\varepsilon = \int_\Omega \rho^\tau_{k+1} \log \rho^\tau_{k+1} - \varepsilon \int_\Omega \rho^\tau_{k+1} \operatorname{div}\xi\, dx + o(\varepsilon)$$

Now take an optimal coupling $\gamma \in \Pi(\rho^\tau_{k+1}, \rho^\tau_k)$ for W_2 - distance,

define $\gamma_\varepsilon := (\Phi(\varepsilon, \cdot) \times Id)_\# \gamma$

then $(\pi_1)_\# \gamma_\varepsilon = \rho_\varepsilon = (\Phi(\varepsilon,\cdot))_\# \rho_{k+1}^\tau$

$(\pi_2)_\# \gamma_\varepsilon = \rho_k^\tau$

$\Phi(\varepsilon, x) = x + \varepsilon \zeta(x) + o(\varepsilon)$

then

$$W_2^2(\rho_\varepsilon, \rho_k^\tau) \le \int_{\Omega\times\Omega} |x-y|^2 \, d\gamma_\varepsilon$$

$$= \int_{\Omega\times\Omega} |\underbrace{\Phi(\varepsilon, x)}_{x+\varepsilon\zeta(x)+o(\varepsilon)} - y|^2 \, d\gamma(x,y)$$

$$= \underbrace{\int_{\Omega\times\Omega} |x-y|^2 d\gamma}_{= W_2^2(\rho_{k+1}^\tau, \rho_k^\tau)} + 2\varepsilon \int_{\Omega\times\Omega} \langle \zeta(x), x-y\rangle \, d\gamma + o(\varepsilon)$$

So

B) $$W_2^2(\rho_\varepsilon, \rho_k^\tau) \le W_2^2(\rho_{k+1}^\tau, \rho_k^\tau) + 2\varepsilon \int_{\Omega\times\Omega} \langle \zeta(x), x-y\rangle \, d\gamma + o(\varepsilon)$$

Hence

$$\inf = \int_\Omega \rho_{k+1}^\tau \log \rho_{k+1}^\tau \, dx + \frac{1}{2\tau} W_2^2(\rho_{k+1}^\tau, \rho_k^\tau)$$

$$\leq \int_\Omega \rho_\varepsilon \log \rho_\varepsilon \, dx + \frac{1}{2\tau} W_2^2(\rho_\varepsilon, \rho_k^\tau)$$

$$\leq \inf + \frac{\varepsilon}{\tau} \int_{\Omega\times\Omega} \langle \xi(x), x-y \rangle \, d\gamma$$

$$- \varepsilon \int_\Omega \rho_{k+1}^\tau \operatorname{div} \xi \, dx + o(\varepsilon)$$

ε can be $\pm$

$\Downarrow$

$$\int_\Omega \rho_{k+1}^\tau \operatorname{div} \xi \, dx = \frac{1}{\tau} \int_{\Omega\times\Omega} \langle \xi(x), x-y \rangle \, d\underbrace{\gamma}_{(T_{k+1} \times Id)_\# \rho_k^\tau}$$

$$= \frac{1}{\tau} \int \langle \xi \circ T_{k+1}(x), T_{k+1}(x) - x \rangle \, d\rho_k^\tau(x)$$

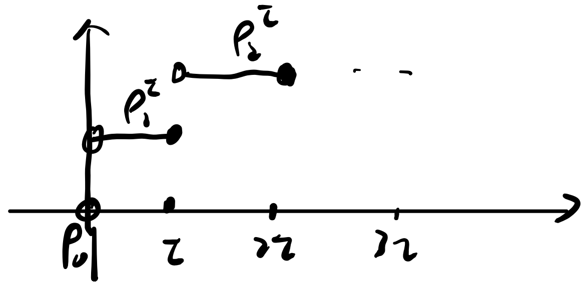
$\boxtimes$

<u>Thm</u>: Given $\tau > 0$,

let $\rho^\tau : [0, \infty) \to \mathcal{P}(\Omega)$

be the curve of probability densities given by

$$\rho^\tau(t) := \begin{cases} \rho_0, & \text{for } t = 0, \\ \rho_k^\tau & \text{for } t \in ((k-1)\tau, k\tau], \quad k \geq 1 \end{cases}$$



Conclusion: $\exists$ a curve of probabilities measures $\rho \in L^1_{loc}([0,\infty) \times \Omega)$, s.t. up to a subsequence in τ,

$$\rho^\tau \rightharpoonup \rho \quad \text{weakly in } L^1_{loc}([0,\infty)\times\Omega).$$

Further, ρ satisfies the heat eq. in $\mathcal{D}'$ with initial data ρ_0 and zero Neumann boundary condition.

Pf: By the JKO scheme,

$$\frac{W_2^2(\rho_k^\tau, \rho_{k-1}^\tau)}{2\tau} + \int_\Omega \rho_k^\tau \log \rho_k^\tau \le \int_\Omega \rho_{k-1}^\tau \log \rho_{k-1}^\tau \, dx$$

Taking sum over $k = 1, 2, \dots, k_0$,

c) $$\sum_{k=1}^{k_0} \frac{W_2^2(\rho_k^\tau, \rho_{k-1}^\tau)}{2\tau} + \int_\Omega \rho_{k_0}^\tau \log \rho_{k_0}^\tau \, dx \le \int_\Omega \rho_0 \log \rho_0 \, dx$$

It implies: $\int_\Omega \rho_k^\tau \log \rho_k^\tau$ decreases in k.

or

$$\int_\Omega \rho^\tau(t,x) \log \rho^\tau(t,x) \, dx \le \int_\Omega \rho_0 \log \rho_0 \, dx.$$

For any $k_0 \ge 1$,

$$\sum_{k=1}^{k_0} \frac{W_2^2(\rho^\tau(k\tau), \rho^\tau((k-1)\tau))}{2\tau}$$

$$\le \int_\Omega \rho_0 \log \rho_0 - \int_\Omega \underbrace{\rho_{k_0} \log \rho_{k_0}}_{\ge -1} \, dx$$

$$\le \int_\Omega (\rho_0 \log \rho_0 + 1) \, dx < \infty$$

Further, $$\int_\Omega \underbrace{\rho^\tau(t)}_{(\rho_k^\tau)_k} \, dx = \underbrace{1}_{(\rho_{k+1}^\tau)}$$

$$(*) \int_{t_1}^{t_2} \int_\Omega \rho^\tau(t,x)\,dx\,dt = t_2 - t_1$$

$$\forall\ 0 \le t_1 \le t_2$$

Hence, up to a subsequence,

$$\rho^\tau(t,x) \rightharpoonup \rho(t,x)$$

weakly in $L^1_{loc}([0,\infty)\times\Omega)$.

Passing limit for $(*)$ $(\tau \to 0)$

$$\int_\Omega \rho(t,x)\,dx = 1, \quad \text{a.e. } t \in [0,T].$$

Next we verify that ρ satisfies the heat eq in $\mathcal{D}'$.

Idea: test the heat eq. against a test function of the form $\psi(x)\xi(t)$

FACT: take $\psi \in C^\infty(\Omega)$,

$$\frac{\partial \psi}{\partial \nu}\Big|_{\partial\Omega} = 0 \quad (\nabla\psi\cdot\nu = 0)$$

Taylor's expansion:

$$\psi(x) - \psi(y) = \langle \nabla\psi(y), x-y\rangle + \frac{1}{2}\int_0^1 \nabla^2\psi(tx+(1-t)y)[x-y, x-y]\,dt$$

$$\Rightarrow \left|\underbrace{\psi(x) - \psi(y) - \langle \nabla\psi(y), x-y\rangle}\right| \le \frac{1}{2}\|\nabla^2\psi\|_{L^\infty}|x-y|^2$$

Thus we can estimate

$x = x$

$y = T_k(x)$

$$\int_\Omega \left|\underbrace{\langle \nabla\psi\circ T_k, T_k - x\rangle + \psi(x) - \psi(T_k x)}\right| \rho_{k-1}^\tau\,dx$$

$$\le \frac{1}{2}\|\nabla^2\psi\|_{L^\infty}\int_\Omega |T_k(x) - x|^2\,\rho_{k-1}^\tau(dx)$$

$$= \frac{1}{2}\|\nabla^2\psi\|_\infty W_2^2(\rho_k^\tau, \rho_{k-1}^\tau)$$

Applying previous lemma with $\xi = \nabla\psi$

$\xi \,//\, \partial\Omega$ since $\frac{\partial\psi}{\partial\nu}\Big|_{\partial\Omega} = 0$.

$$\int_\Omega \rho_k^\tau \underbrace{\operatorname{div}\xi}_{\Delta\psi} = \underbrace{\frac{1}{\tau}\int_\Omega \langle \overset{\nabla\psi}{\xi} \circ T_k, T_k - x\rangle \, d\rho_{k-1}^\tau(x)}_{\star}$$

i.e.

i.e.

$$\left| -\int \Delta\psi \, \rho_k^\tau \, dx + \underbrace{\frac{1}{\tau}\int_\Omega (\psi(T_k(x)) - \psi(x)) \, \rho_{k-1}^\tau \, dx}_{} \right|$$

$$\leq \frac{1}{2} \|D^2\psi\|_\infty \frac{W_2^2(\rho_k^\tau, \rho_{k-1}^\tau)}{\tau}$$

Or

$$\left| -\int \Delta\psi \, \rho_k^\tau \, dx + \frac{1}{\tau}\left(\int_\Omega \psi \, d\rho_k^\tau - \int_\Omega \psi \, d\rho_{k-1}^\tau\right) \right|$$

$$\leq \frac{1}{2} \|D^2\psi\|_\infty \frac{W_2^2(\rho_k^\tau, \rho_{k-1}^\tau)}{\tau}$$

Now take $\xi \in C_c^\infty([0, +\infty))$,

Multiply by $\tau \xi((k-1)\tau)$,

$$\Big| \int_\Omega \psi(x) \, \rho^\tau(k\tau, x) \, \xi((k-1)\tau) \, dx$$

$$- \int_\Omega \psi(x) \, \rho^\tau((k-1)\tau, x) \, \xi((k-1)\tau) \, dx$$

$$- \tau \int_\Omega \Delta \psi(x) \rho^\tau(k\tau, x) \zeta((k-1)\tau)\, dx \Big|$$

$$\le \frac{1}{2} \|\nabla^2 \psi\|_\infty \|\zeta\|_\infty W_2^2(\rho^\tau(k\tau), \rho^\tau((k-1)\tau)).$$

Taking summation over $k = 1, 2, \cdots$ yields.

$$\Big\| \underbrace{-\zeta(0) \int_\Omega \psi(x) \rho_0(x)\, dx}_{I_0}$$

$$\left.\begin{aligned} &+ \sum_{k=1}^\infty \int_\Omega \psi(x) \rho^\tau(k\tau, x) \zeta(\underline{(k-1)\tau})\, dx \\ &- \sum_{k=1}^\infty \int_\Omega \psi(x) \rho^\tau(k\tau, x) \zeta(\underline{k\tau})\, dx \end{aligned}\right\} \mathrm{I}.$$

$$\underbrace{- \sum_{k=1}^\infty \tau \int_\Omega \Delta\psi(x) \rho^\tau(k\tau, x) \zeta(\underline{(k-1)}\tau)\, dx}_{\mathrm{II}} \Big\|$$

$$\le C \sum_{k=1}^\infty W_2^2(\rho^\tau(k\tau), \rho^\tau((k-1)\tau)) \le C\tau.$$

$$\text{Term I} = -\int_0^\infty \int_\Omega \psi(x) \rho^\tau(t, x)\, \partial_t \zeta\, dx\, dt$$

$$\text{Term II} = -\int_0^\infty \int_\Omega \Delta\psi(x) \rho^\tau(t, x) \zeta(t)\, dx\, dt$$

$$+ \quad \underset{\uparrow}{O}(\tau) \quad \zeta \in C_c^\infty([0,\infty)\times\Omega).$$

Hence:

$$\Big| - \zeta(0)\int_\Omega \psi(x)\,\rho_0(x)\,dx$$

$$- \int_0^\infty \int_\Omega \psi(x)\,\underbrace{\rho^\tau(t,x)}\,\partial_t\zeta\,dx\,dt$$

$$- \int_0^\infty \int_\Omega \Delta\psi(x)\,\underbrace{\rho^\tau(t,x)}\,\zeta(t)\,dx\,dt\Big|$$

$$\le O(\tau) \to 0$$

Since $\rho^\tau \to \rho$ in $L^1_{loc}([0,\infty)\times\Omega)$,

we conclude

$$\left\{\begin{aligned} &-\zeta(0)\int_\Omega \psi(x)\,\rho_0(x)\,dx \\ &- \int_0^\infty\int_\Omega \psi(x)\,\rho(t,x)\,\partial_t\zeta(t)\,dx\,dt \\ &- \int_0^\infty\int_\Omega \Delta\psi(x)\,\rho(t,x)\,\zeta(t)\,dx\,dt = 0 \end{aligned}\right.$$

$\forall \xi, \psi$ compactly support. C^∞

* This is the weak formulation of heat equation with 0 Neumann boundary condition.

$$\boxed{\begin{array}{l}\text{See Figalli et al.} \\ \text{A Invitation to Optimal Transport.} \\ \text{§3.3. for details.}\end{array}}$$

The end.