Notes on Optimal Transport and Wasserstein Gradient Flow

Shuailong Zhu

Contents

1 Some Preliminaries for Optimal Transport and Wasserstein Gradient Flow

1.1 About Push-Forward

Proposition 1.1.1. *Let $T : X \to Y$, $\mu \in \mathcal{P}(X)$, and $v \in \mathcal{P}(X)$. Then*

$$v = T\mu$$

if and only if for anly $\varphi : Y \to \mathbb{R}$ Borel and bounded, we have that

$$\int_Y \varphi(y)dv(y) = \int_X \varphi(T(x)d\mu(x)$$

Proof. For any Borel set $A \subset Y$, it holds

$$\int_Y \mathbb{1}_A dv = \mu(T^{-1}(A)) = \int_Y \mathbb{1}_{T^{-1}(A)} d\mu = \int_Y \mathbb{1}_A \circ T d\mu$$

Thus, for any simple function $\varphi : Y \to \mathbb{R}$,

$$\int_Y \varphi dv = \int_Y \varphi \circ T d\mu$$

Consider a fixed bounded Borel function, we can have a sequence of simple functions $(\varphi_k)_{k\in\mathbb{N}}$ such that $|\varphi_k - \varphi| \to 0$ uniformly by Theorem 1.2.3 in my notes of Measure Theory. Then,

$$\int_Y \varphi dv = \lim_{k\to\infty} \int_X \varphi_k dv = \lim_{k\to\infty} \int_X \varphi_k \circ T d\mu = \int_X \varphi \circ T d\mu$$

where the last equality comes from dominated convergence theorem because $\varphi \circ T$ can still bound $\varphi_k \circ T$, but how to show it is measurable and absolutely integrable? Absolutely integrable can be deduced from the boundedness and finite measure μ; measurable should be related to borel measurability of φ, which is very similar to Example 1.2.1 in my notes of Measure Theory. □

Proposition 1.1.2 (Change of Variable, Theorem 6.1.7 [Coh13])**.** *Assume that T is a diffeomorphism between open sets X and Y of $\mathbb{R}^d$, and assume probability measures μ, v are absolutely continuous with respect to Lebesgue measure. Then,*

$$\int_Y \varphi(y)\sigma(y)dy = \int_X \varphi(T(x))\sigma(T(x))det(DT(x))dx$$

1.2 Weak-$*$ Topology and Narrow Topology

Proposition 1.2.1 (Riesz Representation Theorem for $C_c(X)$)**.**

$$\begin{aligned} \mathcal{M}(X) &:= \{\textit{finite signed measures on } X\} && (1.1)\\ &= C_c(X)^* := \{\textit{continuous compactly supported functions}\}^* && (1.2)\\ &= C_0(X)^* := \{\textit{continuous functions vanishing at } \infty\}^* && (1.3)\end{aligned}$$

Precisely, $(\mathcal{M}(X), \|\cdot\|_{TV})$ is the dual space of $(C_0(X), \|\cdot\|_{L^\infty})$ or $(C_c(X), \|\cdot\|_{L^\infty})$. Notice $C_c(X)$ not closed, $C_0(X)$ closed.

Weak-$\star$ convergence. By Banach-Alaoglu's Theorem, if $(\mu_k)_{k\in\mathbb{N}}$ is a sequence of probability measures, then $\exists$ a subsequence that weakly-* converges to a measure $\mu \in \mathcal{M}(X)$.

Narrow Topology. We say $\mu_k \rightharpoonup \mu$ if

$$\int \varphi d\mu_k \to \int \varphi d\mu, \text{ for all } \varphi \in C_b(X) \tag{1.4}$$

This is equivalent to: if

$$\liminf_{k\to\infty} \int \varphi d\mu_k \geq \int \varphi d\mu \tag{1.5}$$

for all φ that is lower semi-continuous and lower bounded.

2 Existence and Optimal Condition in (KP)

2.1 Existence and Optimal Condition

Proposition 2.1.1 (Existence of Optimal Plan/Coupling γ in (KP) problem). *Let $c : X \times Y \to [0, \infty]$ be lower semicontinuous, $\mu, v \in \mathcal{P}(X)$. Then there exists a coulping $\gamma \in \Gamma(\mu, v)$ which is a minimizer for (KP).*

The strategy we use is: (1) Compactness in a certain topology, i.e. narrow topology of $\Gamma(\mu, v)$ deduced from tightness; (2) lower semi-continuity. We might use similar strategy for the space of transport map (T_k) for (MP).

Remark. However, this strategy doesn't work for (MP). Using Proposition 3.8 and 3.9 in my notes of functional analysis, we can check the "admitting of a weakly convergent sequence" in L^2 and the weak limit of the operator.

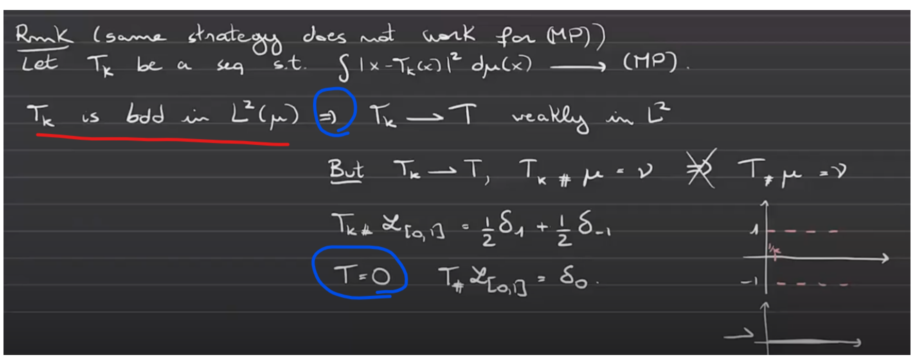


Figure 1: Same strategy doesn't work for (MP) [Col21]

Theorem 2.1.2 (Optimal Condition). *Let $\bar{\gamma}$ be optimal, and $c : X \times Y \to \mathbb{R}$ is continuous. Then supp($\bar{\gamma}$) is c-cyclically monotone. Note: assume $\mu = \sum \frac{1}{2^i}\delta_{q_i} \in \mathcal{P}(\mathbb{R})$, where q_i is a rational number. Then supp(μ) = $\mathbb{R}$.*

2.2 Some Convex Analysis Tools

A function $\varphi : \mathbb{R}^d \to \bar{\mathbb{R}}$ is convex if

$$\varphi(x) = \sup_{y\in\mathbb{R}^d} \{x \cdot y + \lambda_y\} \tag{2.1}$$

Definition 2.2.1 (c-convex). *Using the idea of the supremum of affine functions for convex, given X and Y metric spaces, $c : X \times Y \to \mathbb{R}$, we define that $\varphi : X \to \bar{\mathbb{R}}$ is c-convex if*

$$\varphi(x) = \sup_{y \in \mathbb{R}^d} \{-c(x, y) + \lambda_y\} \tag{2.2}$$

Theorem 2.2.1. *[Rockafellar, from convex analysis] A set $S \subset \mathbb{R}^d \times \mathbb{R}^d$ is (c-)cyclically monotone iff there exists a (c-)convex function $\varphi : \mathbb{R}^d \to \bar{\mathbb{R}}$ such that $S \subset \partial\varphi$. Rockafellar theorem provides a more clear description of c-cyclically monotone.*

2.3 General Kantorovich Duality

Definition 2.3.1 (c-Legendre transform). *Given a c-convex function $\varphi : X \to \bar{\mathbb{R}}$, we define its c-Legendre transform $\varphi^c : Y \to \bar{\mathbb{R}}$ as*

$$\varphi^c(y) = \sup_{x \in X} \{-c(x, y) - \varphi(x)\} \tag{2.3}$$

Properties.

$$\varphi(x) + \varphi^c(y) + c(x, y) \geq 0 \text{ for all } x \in X, y \in Y \tag{a}$$

$$\varphi(x) + \varphi^c(y) + c(x, y) = 0 \text{ iff } y \in \partial_c\varphi(x) \tag{b}$$

Theorem 2.3.1 (Kantorovich duality). *Let $c(\cdot, \cdot)$ be continuous and bounded from below, and assume that $\inf_{\gamma \in \Gamma(\mu,v)} \int_{X \times Y} c d\gamma < +\infty$, then*

$$\min_{\gamma \in \Gamma(\mu,v) c d\gamma} \int_{X \times Y} c d\gamma = \max_{\varphi(x)+\psi(y)+c(x,y) \geq 0} \int_X -\varphi d\mu + \int_Y -\psi dv$$

Proof. . (1) (KP)$\geq$ (DP).

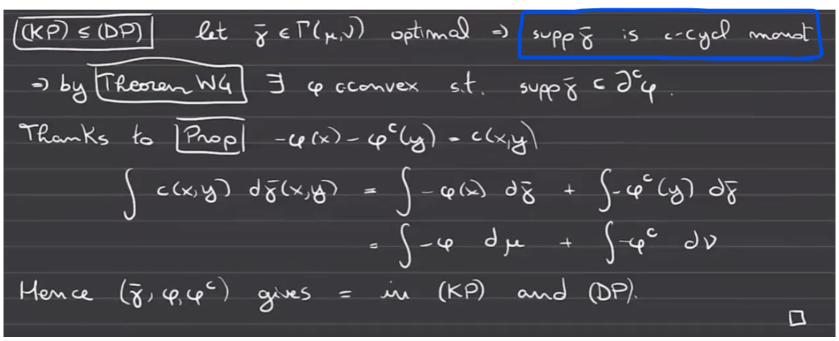


Figure 2: (KP) $\leq$ (DP)

$\square$

Remark. The existence of minimizer of LHS only requires $c(\cdot, \cdot)$ to be lower-continuous in Proposition 2.1.1. We can obtain some corollaries that are a bit tricky due to the proof's specialty, i.e. it only requires that supp(γ) to be c-cyclically monotone.

Corollary 2.3.2 (It might be inaccurate. We might need to put this corollary under the setting of c continuous.). *Given a γ with supp(γ) is c-cyclically monotone, then $\exists$ φ c-convex such that*

$$(KP) \leq \int c d\gamma = \int -\varphi d\mu + \int -\varphi^c dv \leq (DP)$$

Well, (KP)$\geq$(DP) is general. Thus, we can obtain (KP)=(DP) and γ is the optimal.

Interestingly, Theorem 2.1.2 is for continuous cost, while Theorem 2.2.1 is for "any" cost.

Corollary 2.3.3. *if $c(\cdot,\cdot)$ is continuous, the following are equivalent:*

- *γ is optimal;*
- *supp(γ) is c-cyclically monotone;*
- *there exists a convex map φ such that supp$(\gamma) \subset \partial_c\varphi$*

Corollary 2.3.4 (Theorem 2.3.2, [FG21])**.** *If $c(\cdot,\cdot)$ is lower semi-continuous, then we still have (KP)=(DP), which can be shown by approximation.*

We can find an excellent example that is c-cyclically monotone but not optimal, which is complementary to Corollary 5.2. However, this exmaple seems a bit controversial to Corollary 2.3.2.

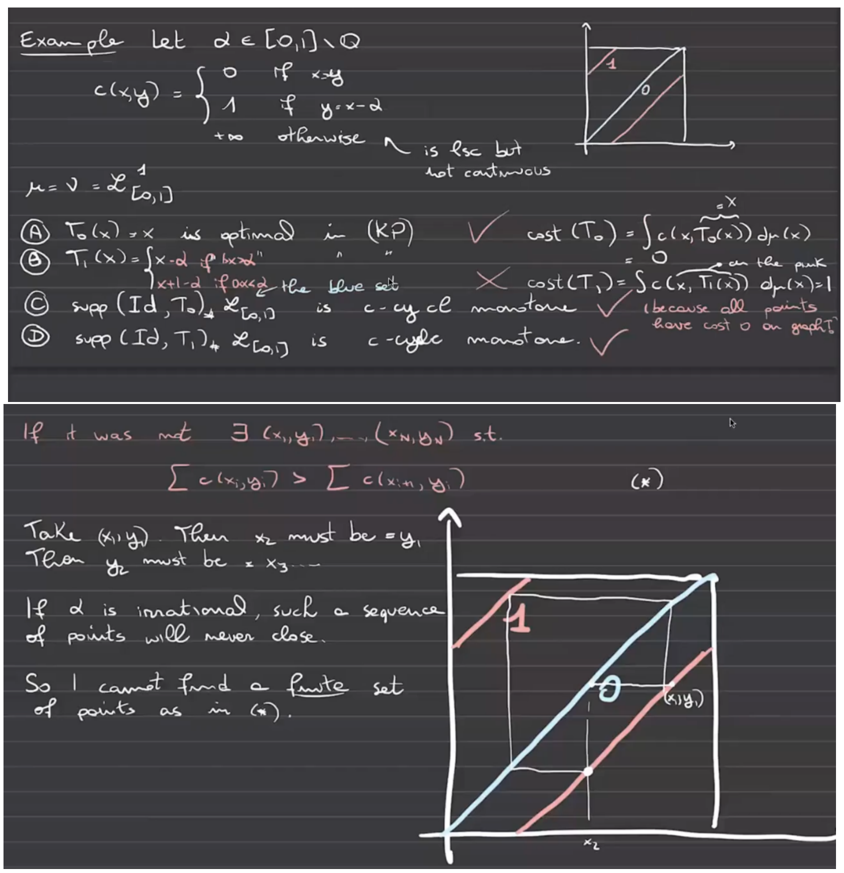


Figure 3: A beautiful counter-example [Col21].

2.4 From Convex Geometry to (KP)=(DP)

$$\begin{aligned}
\inf_{\gamma\in\Gamma(X,Y)} c(x,y)d\gamma &= \inf_{\gamma\geq 0}\left\{\int cd\gamma + \sup_{\varphi}\{\int_{X\times Y}\varphi(x)d\gamma - \int_X \varphi(x)d\mu\} + \sup_{\varphi}\{\int_{X\times Y}\psi(y)d\gamma - \int_Y \psi(y)dv\}\right\}\\
&= \inf_{\gamma\geq 0}\sup_{\varphi,\psi}\left\{\int cd\gamma + \int_{X\times Y}\varphi(x)d\gamma - \int_X \varphi(x)d\mu + \int_{X\times Y}\psi(y)d\gamma - \int_Y \psi(y)dv\right\}\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\text{(Simon Minimax Theorem)}\\
&= \sup_{\varphi,\psi}\left\{-\int_X \varphi(x)d\mu - \int_Y \psi(y)dv + \inf_{\gamma\geq 0}\int_{\times Y}[c+\varphi(x)+\psi(y)]d\gamma\right\}\\
&= \sup_{\varphi(x)+\psi(y)+c(x,y)\geq 0}\left\{-\int_X \varphi(x)d\mu - \int_Y \psi(y)dv\right\}
\end{aligned}$$

3 Existence and Characterization of Transport Maps in (MP)

3.1 Brenier's Theorem

Theorem 3.1.1 (Brenier's Theorem)**.** *Let $X = Y = \mathbb{R}^d$, $c(x,y) = \dfrac{|x-y|^2}{2}$. Suppose that*

$$\int_X |x|^2 dx + \int_Y |y|^2 dy < +\infty$$

and $\mu \ll dx$. Then there exists a unique optimal plan γ. In addition, $\gamma = (Id \times T)_{\#}\mu$ and $T = \nabla\varphi$ for some convex function φ.

Theorem 3.1.2 (General Brenier's Theorem)**.** *Let $X = Y = \mathbb{R}^d$, $\mu \ll dx$, and supp(v) compact. Let c be continuous and bounded from below, and assume that $inf_\gamma \int_{X\times Y} < \infty$. Also, suppose that:*

- *for every $y \in$ supp(v), the map $x \to c(x,y)$ is differentiable;*
- *for every $x \in \mathbb{R}^d$, the map $y \to \nabla_x c(x,y)$ is injective.*
- *for every $y \in$ supp(v) and $R > 0$, $|\nabla_x c(x,y)| \leq C_R$ for every $x \in \mathbb{B}_R$*

Then there exists a unique optimal plan γ with $\gamma = (Id \times T)_{\#}\mu$ and T satisfying

$$\nabla_x c(x,y)|_{y=T(x)} + \nabla\varphi(x) = \nabla_x c(x,T(x)) + \nabla\varphi(x) = 0$$

for some c-convex function φ.

3.2 Stability and Regularity of Optimal Transport Plans/Maps

Before going to any details, it seems that we refer to "the same γ" in the topology of $\Gamma(\mu,v)$, we are talking about almost everywhere equal.

Theorem 3.2.1 (Stability)**.** *Let $\{\mu_k\}, \{v_k\} \subset \mathcal{P}(X)$ with supp(μ_k), supp$(v_k) \subset K$ compact,*

$$\mu_k \rightharpoonup \mu, v_k \rightharpoonup v$$

Let $c : X \times X \to [0,\infty]$ Then,

- *Any weak limit point of π_k optimal in $\Gamma(\mu_k, v_k)$ is optimal.*

- *if* $X = \mathbb{R}^d$, $c(x,y) = \dfrac{\|x-y\|^2}{2}$, $\mu \ll \mathscr{L}^d$ *then*

$$\pi_h \to (Id, \nabla\varphi)_{\#}\mu$$

If $\mu_h \ll \mathscr{L}^d$, *then*

$$(Id, \nabla\varphi_h)_{\#}\mu_h \to (Id, \nabla\varphi)_{\#}\mu$$

The above is just a simple consequence of the first point and Brenier's theorem. More interestingly, if $\mu \equiv \mu_h$

$$\nabla\varphi_k \to \nabla\varphi \ in \ L^p(\mu)$$

Example 3.2.1 (Regularity)**.** *Even if the* $supp(\mu_h)$ *and* $supp(v_h)$ *is connected, there exists* h_0*, such that* T_h *for* $h \geq h_0$ *might not be continuous.*

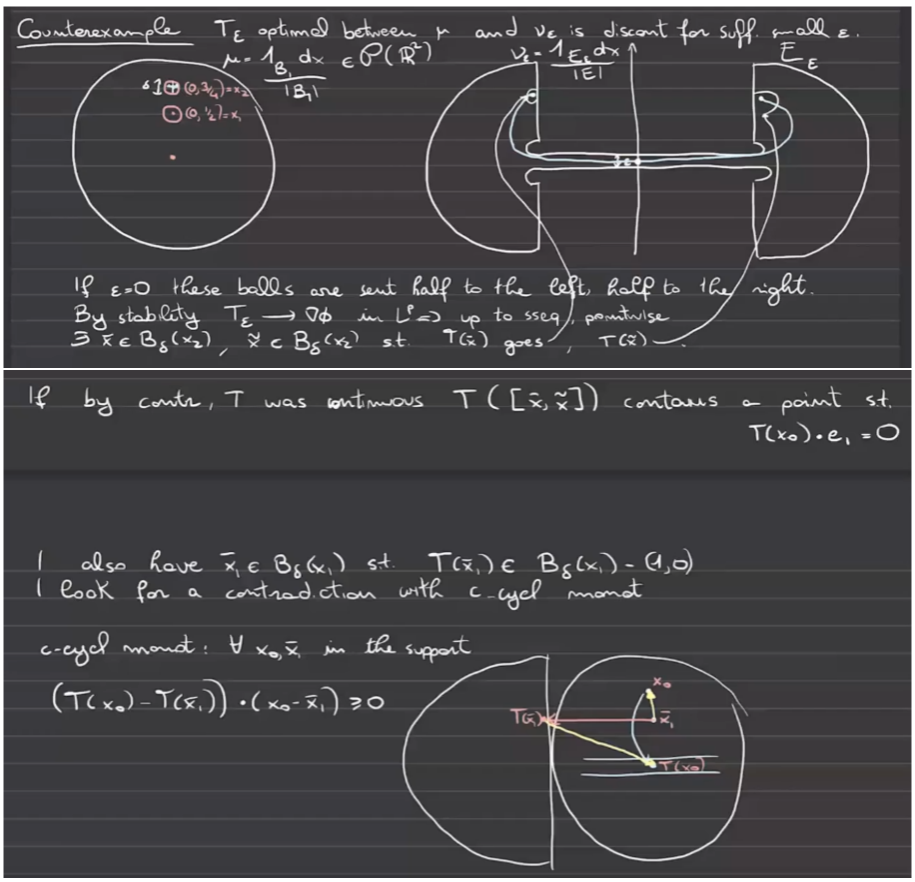


Figure 4: A example for regularity [Col21].

4 Wasserstein Space and Gradient Flows

4.1 Wasserstein Space and Geodesic

Theorem 4.1.1 (Wasserstein distance and narrow convergence (weak-$*$ convergence))**.** *Fix* $1 \leq p < \infty$ *and a base point* $x_0 \in X$*. Let* $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}_p(X)$ *be a sequence of probability measures and let* $\mu \in \mathcal{P}_p(X)$*. The following are equivalent:*

(1) $\mu_n \xrightharpoonup{*} \mu$ *and* $\int_X d(x_0,x)^p d\mu_n \to \int_X d(x_0,x)^p d\mu$
(2) $W_p(\mu_n,\mu) \to 0$

Proof. We only show (2)⇒(1) here [FG21] and leave the other direction to future.

- 2. ⇒ 1. Let $\gamma_n \in \Gamma(\mu_n,\mu)$ be an optimal transport plan with respect to the cost $c(x,y) = d(x,y)^p$. Applying the triangle inequality for the Wasserstein distance (recall Theorem 3.1.5) and using that $W_p(\mu_n,\mu) \to 0$, we have

$$\int_X d(x_0,x)^p \, d\mu_n = W_p(\delta_{x_0},\mu_n)^p \to W_p(\delta_{x_0},\mu)^p = \int_X d(x_0,x)^p \, d\mu \, .$$

It remains to show that $\mu_n \xrightharpoonup{*} \mu$. Let $\varphi \in C_c(X)$ be a compactly supported function, and let $\omega : [0,\infty) \to [0,\infty)$ be its modulus of continuity (i.e., $|\varphi(x) - \varphi(y)| \le \omega(d(x,y))$ for all $x,y \in X$). Given $\delta > 0$, we have

$$\begin{aligned}
\left| \int_X \varphi \, d\mu_n - \int_X \varphi \, d\mu \right| &\le \int_{X\times X} |\varphi(x) - \varphi(y)| \, d\gamma_n(x,y) \\
&\le \int_{\{d(x,y)\le\delta\}} \omega(\delta) \, d\gamma_n(x,y) + \int_{\{d(x,y)>\delta\}} 2\|\varphi\|_\infty \, d\gamma_n(x,y) \\
&\le \omega(\delta) + 2\|\varphi\|_\infty \int_{\{d(x,y)>\delta\}} \frac{d(x,y)^p}{\delta^p} \, d\gamma_n(x,y) \\
&\le \omega(\delta) + \frac{2\|\varphi\|_\infty}{\delta^p} \int_{X\times X} d(x,y)^p \, d\gamma_n(x,y) \\
&= \omega(\delta) + \frac{2\|\varphi\|_\infty}{\delta^p} W_p(\mu_n,\mu)^p.
\end{aligned}$$

By first letting $n \to \infty$ and then $\delta \to 0$, the last inequality implies that $\int_X \varphi \, d\mu_n \to \int_X \varphi \, d\mu$, concluding the proof.

I want to mention one thing here: $C_c(X) \subset C_b(X)$ because any continuous mapping of a compact set is compact, and thus the image is closed and bounded in $\mathbb{R}$. Therefore $\|\varphi\|_\infty$ is finite. Actually, we can also prove narrow convergence. □

Corollary 4.1.2 (Wasserstein distance metricize weak-$*$ topology)**.** *Let X be compact, $(\mu)_{n\in\mathbb{N}}, \mu \subset \mathcal{P}_p(X)$. Then*

$$\mu_n \xrightharpoonup{*} \mu \iff W_p(\mu_n,\mu) \to 0$$

Proof. Instead of using Theorem 4.1.1. We can easily use Stability result to prove $\mu_n \xrightharpoonup{*} \mu \Rightarrow W_p(\mu_n,\mu) \to 0$ [Col21]. □

Definition 4.1.1 (Geodesic and Minimizing geodesic)**.** *Geodesic is not necessarily minimizing geodesic. We can find two points A, B on the largest circle on a sphere, then we will have two curves. Both AB and $\tilde{AB}$ are geodesics, but Only One curve will be a minimizing geodesic if AB is not diameter.*

However, in lots of book like [San15], we only take minimizing geodesic as geodesic. And in the context of Wasserstein space, it is also the case.

Theorem 4.1.3 (Construction of geodesic in Wasserstein space)**.** *Every optimal coupling induces a minimizing geodesic.*

4.2 Displacement Convexity and Functionals on $\mathcal{P}_2(X)$

Let us start with some basic definitions of a particle system. And in this section $X = \mathbb{R}^d$.

Internal Energy. Given $U : \mathbb{R} \to \mathbb{R}$,

$$\mathcal{U}(\mu) = \int U(\rho)dx, \text{ if } \mu = \rho(x)dx$$

Potential energy. Given $V : \mathbb{R}^d \to \mathbb{R}$,

$$\mathcal{V}(\mu) = \int V(x)d\mu$$

Interaction Energy. Given $W : \mathbb{R}^{2d} \to \mathbb{R}$

$$\mathcal{W}(\mu) = \int W(x,y)d\mu(x)d\mu(y)$$

Definition 4.2.1 (Displacement-convexity)**.** *We say a functional $F : \mathcal{P}_2(X) \to \bar{\mathbb{R}}$ is λ-displacement convex if for any minimizing geodesic $\eta : [0,1] \to \mathcal{P}(X)$ such that $F \circ \eta : [0,1] \to \bar{\mathbb{R}}$ is λ-convex,*

$$F(\eta(t)) \le (1-t)F(x) + tF(y) - \frac{\lambda}{2}t(1-t)d^2(x,y), \forall t \in [0,1]$$

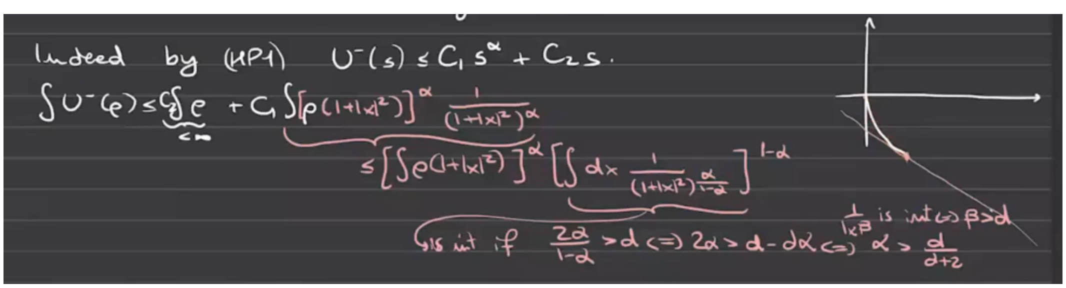
If we replace the definition with the exsitence of one minimizing geodesic, then it is weak displacement convexity, which is used in [San15].

Proposition 4.2.1 (Displacement-convexity of potential energy and interaction energy)**.** *If V is convex, then $\mathcal{V}$ is W_2-displacement convex. The functional $\mathcal{W} : \mathcal{P}_2(X) \to \bar{\mathbb{R}}$ is displacement convex if W is convex. Note that the proof for "V convex is a necessary condition" in [San15] is not completely correct.*

Proposition 4.2.2 (Displacement convexity of internal energy)**.** *Let $U : [0,\infty) \to \bar{\mathbb{R}}$, lower semicontinuous, and $U(0) = 0$, and satisfies the following properties,*
(1) $\limsup_{s\to 0^+} \dfrac{U^-(s)}{s} < \infty$ for some $\alpha > \frac{d}{d+2}$
(2) $s \to s^d \cdot U(s^{-\alpha})$ is convex and decreasing on $(0,\infty)$

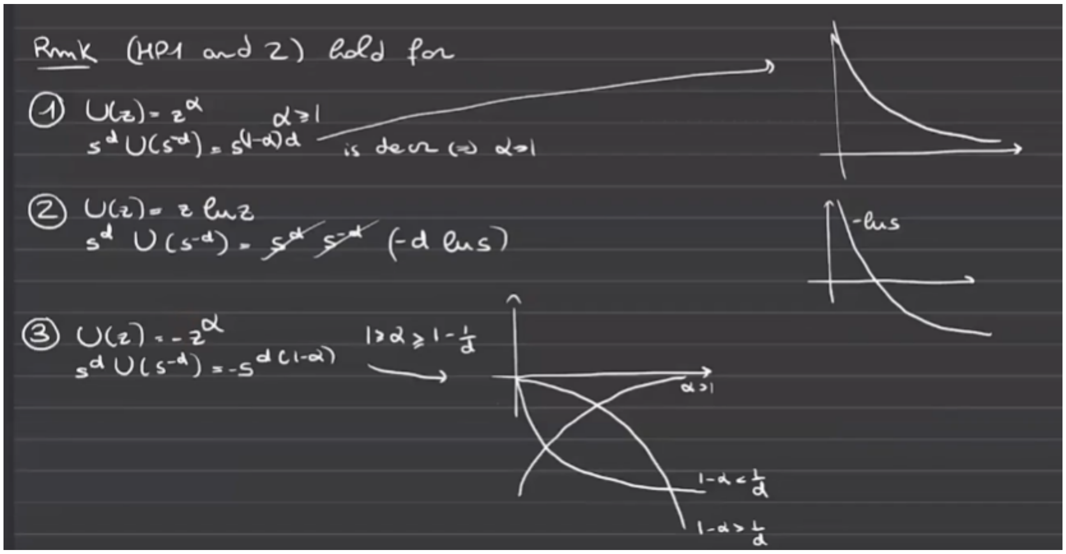
Then $\mathcal{U}$ is W_2-displacement convex in $\mathcal{P}_2^{ac}(X)$

Proof. In the main proof (2), it is enough to only consider the displacement interpolation as it ensures the displacement convex of other minimizing geodeiscs [FG21]. We can refer to [FG21] for details. For condition (1), if it holds, then $\int U^-(\rho)dx < +\infty, \forall \rho \in L^1(\mathbb{R}^d)$, which can guarantee the integrability of $\int U(\rho)dx$.



∎

Example 4.2.1 (Some examples of $U(\cdot)$).



Corollary 4.2.3 (Brunn-Minkowski inequality). *Assume $A, B \subset \mathbb{R}^d$ compact, then*

$$\mathscr{L}^d(A+B)^{1/d} \geq \mathscr{L}^d(A)^{1/d} + \mathscr{L}^d(B)^{1/d}$$

Proof. Consider $\mathscr{L}(A), \mathscr{L}(B) > 0$. Define

$$\mu_A = \frac{1}{\mathscr{L}(A)}\mathcal{X}_A dx, \mu_B = \frac{1}{\mathscr{L}(B)}\mathcal{X}_B dx$$

Let $U(z) = -z^{1-1/d}$, and μ_t be the geodesic between μ_A, μ_B. Since $\mathcal{U}$ is displacement-convex,

$$\begin{aligned}\mathscr{L}^d(A)^{1/d} + \mathscr{L}^d(B)^{1/d} &= 2 \times \frac{1}{2}(-\mathcal{U}(\mu_A) - \mathcal{U}(\mu_B)) \\ &\leq 2[-\mathcal{U}(\mu_{\frac{1}{2}})]\end{aligned}$$

Since $\mu_{1/2} = (\frac{x + T_{A\to B}(x)}{2})_{\#}\mu_A$ is concentrated on $\frac{A+B}{2}$ Implicitly, we will need $\mu_{1/2}$ is absolutely continuous with respect to Lebesgue measure, which can be shown easily Proposition 5.9 in [Vil21]. We are only left to show that if $\mu_{1/2} = \rho dx$ is concentrated on $\frac{A+B}{2}$,

$$\mathcal{U}(\rho) = \int \rho^{1-1/d} dx \leq \mathscr{L}(\frac{A+B}{2})\left(\int_{\frac{A+B}{2}} \rho \frac{dx}{\mathscr{L}(\frac{A+B}{2})}\right)^{1-\frac{1}{d}} = \mathscr{L}(\frac{A+B}{2})^{1/d}$$

which is by the concave property of $\rho^{1-1/d}$. □

We can utilize Brunn-Minkowski inequality to prove isoperimetric inequality, which is equivalent to Sobolev inequality. Using the displacement convexity of internal energy, we can also show Talagrand transportation inequality among four famous "Talagrand inequalities".

4.3 Gradient Flows

The most interesting part is there are two equivalent "solutions" for heat equation

$$\partial_t u(t,x) = \Delta u(t,x) \tag{4.1}$$

(1) Solve the following implicit Euler scheme and then let $\tau \to 0$

$$u_{k+1}^{\tau} \in argmin \frac{\|u - u_k^{\tau}\|_{L^2}}{2\tau} + \phi(u) \tag{4.2}$$

where

$$\phi(u) = \begin{cases} \frac{1}{2}\int_{\mathbb{R}^d} |\nabla u|^2 dx, \ if \ u \in W^{1,2}(\mathbb{R}^d) \\ +\infty, otherwise \end{cases}$$

(2) Solve the following and let $\tau \to 0$,

$$\rho_{k+1}^{\tau} = argmin \frac{W_2^2(\rho, \rho_k^{\tau})}{2\tau} + \int \rho log(\rho) dx \tag{4.3}$$

This is called JKO scheme. For the more detailed statement of the theorem and proof, we refer to [FG21].

4.4 Benamou-Brenier Formula

5 Wasserstein Gradient Flow and its Applications

In this section, we focus on the relationship between Wasserstein gradient flow, Langevin dynamics, and the Fokker-Planck equation.

- There is an equivalence between SDE and Fokker-Planck equation.
- We can also build some equivalence between Wasserstein gradient flow and some PDEs, including heat equation, continuity equation, and linear Fokker-Planck equation. There are two lines of style (method) to introduce Wasserstein gradient [San15; AGS08] and [FG21] basically follows [AGS08].

5.1 Wasserstein Gradient Flow

For the gradient flow in Hilbert space, we can establish the equivalence between the gradient flow and the limit curve of implicit Euler scheme, We also present the "Euler scheme" in Wasserstein space in Equation 4.3, for which we omit the details and refer to [FG21] for a thorough analysis.

5.1.1 From Variational Formula

The equivalence between the heat equation and limit curve of the Formula 4.3 is already built in Section 4.3. Now, we want to consider a more general problem. Assume $F : \mathcal{P}_2(X) \to \mathbb{R}$, we want to explore what the limit curve of the following iterate scheme is.

$$\rho_{k+1}^{\tau} = argmin \frac{W_2^2(\rho, \rho_k^{\tau})}{2\tau} + F(\rho) \tag{5.1}$$

Specifically, we want to find the correct continuity equation of the limit curve.

$$\partial_t \rho_t + div(v_t \rho_t) = 0 \tag{5.2}$$

Definition 5.1.1 (First L^2-variation)**.** *Assume $F : \mathcal{P}_2(X) \to \mathbb{R}$, we call $\frac{\delta F}{\delta \rho}$, if it exists, any measurable function such that*

$$\frac{d}{d\varepsilon}F(\rho+\varepsilon\chi)|_{\varepsilon=0} = \int \frac{\delta F}{\delta \rho}(\rho)d\chi \tag{5.3}$$

for every pertubation $\chi = \tilde{\rho} - \rho$ with $\tilde{\rho} \in L^{\infty}_c(X) \cap \mathcal{P}(X)$. Specially, if $\frac{\delta F}{\delta \rho}(\rho) = Const$, then

$$\frac{d}{d\varepsilon}F(\rho+\varepsilon\chi) = 0, \textit{for every } \chi$$

We define $G(\rho) = \frac{W_2^2(\rho, \rho_k^\tau)}{2\tau} + F(\rho)$. Following the definition above, we can show

$$\frac{\delta W_2^2(\rho, \rho_k^\tau)}{\delta \rho}(x) = -2\varphi(x) + x^2 \tag{5.4}$$

where $\nabla\varphi$ is the optimal mapping from ρ to ρ_k^τ. To minimize $G(\rho)$,

$$\frac{\delta G}{\delta \rho}(\rho)(x) = Const, \text{for any } x \tag{5.5}$$

$$\frac{-\nabla\varphi(x) + x}{\tau} + \nabla(\frac{\delta F}{\delta \rho}(\rho))(x) = 0 \tag{5.6}$$

$$\frac{T(x) - x}{\tau} = \nabla(\frac{\delta F}{\delta \rho}(\rho))(x) \tag{5.7}$$

Intuitively, when $\tau \to 0$, $\rho \to \rho_k^\tau$. Also, the move from x to $T(x)$ is like the "steepest" gradient descent. Thus, we can define the velocity field as $v = -\frac{T(x)-x}{\tau}$, which is from ρ_k^τ to ρ. Thus, the continuity equation can be written as,

$$\partial_t\rho_t - \nabla \cdot \left(\rho_t \nabla(\frac{\delta F}{\delta \rho}(\rho_t))\right) = 0 \tag{5.8}$$

We can prove the equivalence between the variational formula will converge to Formula 5.8 when $\tau \to 0$ [San15].

5.1.2 From Wasserstein Scalar Product Structure

Definition 5.1.2. *Given two functions $h_1, h_2 : \Omega \to \mathbb{R}$*[1] *with $\int_X h_1 = \int_X h_2 = 0$, one can define the Wasserstein scale product* [2] *at ρ as*

$$\langle h_1, h_2\rangle_\rho = \int \nabla\psi_1 \cdot \nabla\psi_2 \rho dx, \ \textit{where} \begin{cases} div(\rho\nabla\psi_i) = -h_i & \textit{in } \Omega \\ \frac{\partial\psi_i}{\partial v} = 0 & \textit{on } \partial\Omega \end{cases}$$

Definition 5.1.3. *Similar to gradient in Euclidean space, given a functional $F : \mathcal{P}_2(\Omega) \to \mathbb{R}$, its gradient with respect to Wasserstein scalar product at $\tilde{\rho}$ is the unique function $grad_{\mathcal{W}_2}F(\tilde{\rho})$ such that*

$$\langle grad_{\mathcal{W}_2}F(\tilde{\rho}), \frac{\partial\rho_t}{\partial t}|_{t=0}\rangle_{\tilde{\rho}} = \frac{\partial}{\partial t}|_{t=0}F(\rho_t)$$

for any smooth curve $\rho_t : (-1, 1) \to \mathcal{P}(\Omega)$ with $\rho_= \tilde{\rho}$.

[1] $\Omega \subseteq X$ is a convex set

[2] I'm not sure whether the second constraint should be understood distributionally: $\int_{\partial\Omega} \varphi\nabla\psi \cdot v = 0$, for any φ like ρ.

Actually, we defined some Riemannian structure on $\mathcal{P}_2(\Omega)$, and should treat $\frac{\partial \rho_t}{\partial t} \in T_\rho \mathcal{P}_2(\Omega) = \{curves : I \to \mathcal{P}_2(\Omega)\}$. In some sense, we could recogonize $h_1, h_2 \in \mathcal{P}_2(\Omega)$. Finally we can obtain the same formula,

$$grad_{\mathcal{W}_2} F(\tilde{\rho}) = -\nabla \cdot \left(\tilde{\rho} \nabla (\frac{\delta F}{\delta \rho}(\tilde{\rho})) \right) \tag{5.9}$$

Assume we have

$$F(\rho) = \int \rho ln \frac{\rho}{\pi} = \underbrace{\int \rho ln \rho dx}_{internal\ energy} + \underbrace{\int V \rho dx}_{potential\ energy}$$

with $\pi = e^{-V}$. We can get,

$$\frac{\delta F}{\delta \rho}(\rho) = V + ln\rho + 1$$

The continuity equation with respect to the Wasserstein gradient of F is

$$\partial_t \rho_t = -\nabla \cdot (\rho \nabla V + \rho \nabla ln\rho) = -\nabla \cdot (\rho \nabla V) - \Delta \rho$$

This is called (linear) Fokker-Planck equation, which can be also deduced from a SDE using the Markov Semigroup operator in the next section.

5.2 Fokker-Planck Equation and Langevin Dynamics

5.2.1 Kolmogorov Equation

Definition 5.2.1 (Markov Semigroup)**.** *Let X_t be a time-homogeneous Markov process. We can its associated Markov semigroup $(P_t)_t$ acting on the functions*

$$P_t f(x) = \mathbb{E}(f(X_t)|X_0 = x)$$

And notice that $P_{t+s} = P_t \circ P_s$.

To have some differentiable structure on $(P_t)_t$, we define the generator of the operator semigroup as

$$\mathscr{L} f := \lim_{t \to 0} \frac{P_t f - f}{t}$$

for all bounded continuous functions f such that the limit exists. We explore the dynamics of $P_t f$ below.

Definition 5.2.2 (Kolmogorov backward equation)**.**

$$\partial_t (P_t f) = \lim_{h \to 0} \frac{P_{t+h} f - P_t f}{h} = \lim_{h \to 0} \frac{P_h - Id}{h} P_t f = \mathscr{L} P_t f$$

Coarsely, we regard the space of all bounded continuous functions and the probability measures as L^2. Then, we can define the adjoint operator for this Hilbert space L^2. Assume $X_0 \sim \mu_0 = \rho_0 dx$

$$\int P_t f d\mu_0 = \int f d(P_t^* \mu_0) \tag{5.10}$$

Definition 5.2.3 (Kolmogorov forward equation)**.** *Similarly, we can obtain,*

$$\partial_t P_t^* \pi_0 = \mathscr{L}^* P_t^* \pi_0$$

where $\mathscr{L}^*$ is the L^2-adjoint of $\mathscr{L}$ such that $\int \mathscr{L} f h dx = \int f \mathscr{L} h dx$, for "suitable" f, g.

Now we have two equations,

$$\partial_t u_t = \mathscr{L} u_t \tag{Kolmogorov backward equation}$$

$$\partial_t \mu_t = \mathscr{L}^* \mu_t \tag{Fokker-Planck equation, Kolmogorov forward equation}$$

where $u_t = P_t f$ and $\mu_t = P_t^* \mu_0$.

5.2.2 Fokker-Planck Equation of Langevin Dynamics

Consider a Langevin Diffusion Process,

$$dX_t = \nabla V dt + \sqrt{2} dW_t$$

Using Ito's formula, $df(X_t) = \nabla f(X_t) dX_t + \frac{1}{2}\nabla^2 f(X_t) d[X,X]_t = -\nabla f \cdot \nabla V dt + \nabla^2 f dt + (...)dW_t$

$$\mathscr{L} f(x) = \mathbb{E}[\frac{df(X_t)}{dt} | \mathcal{F}_0, X_0 = x] = -\nabla f(x) \cdot \nabla V(x) + \nabla^2 f(x)$$

Assume f, g are functions that vanish at infinity[3], we can have that

$$\int_\Omega (\nabla^2 f) g dx = \underbrace{\int_{\partial\Omega} (\nabla f) g \cdot dx}_{0} - \int_\Omega \nabla f \cdot \nabla g dx \tag{5.11}$$

$$= \underbrace{-\int_{\partial\Omega} f \nabla g \cdot dx}_{0} + \int_\Omega f \nabla^2 g dx \tag{5.12}$$

$$\int_\Omega \nabla f \cdot \nabla V g dx = \underbrace{\int_{\partial\Omega} f(g\nabla V) \cdot dx}_{0} - \int_\Omega \nabla V \nabla g \cdot dx \tag{5.13}$$

$$\mathscr{L}^* g = \Delta g + div(g \nabla V)$$

through integration by parts. So now we get the Fokker-Planck equation,

$$\partial_t \mu_t = \Delta \mu_t + div(\mu_t \nabla V)$$

[3] Coarsely, it is the case for Fokker-Planck equation of Langevin dynamics, as f is usually the test function, g can be regarded as the probability density

References

[AGS08] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

[Coh13] Donald L Cohn. *Measure theory*. Vol. 5. Springer, 2013.

[Col21] Maria Colombo. *Optimal Transport, MATH-476*. 2021.

[FG21] Alessio Figalli and Federico Glaudo. *An invitation to optimal transport, Wasserstein distances, and gradient flows*. 2021.

[San15] Filippo Santambrogio. "Optimal transport for applied mathematicians". In: (2015).

[Vil21] Cédric Villani. *Topics in optimal transportation*. Vol. 58. American Mathematical Soc., 2021.