

Convex analysis and Variational problems

Yiping Lu¹

¹School of Mathematical Science
Peking University

Seminar On Applied Analysis, PKU, 2017

- 1 Convex Analysis
 - Convex sets and Convex Functions
 - Conjugate
 - Subdifferentiable
 - Dual
 - Dual Flat Manifold
- 2 Convex Optimization
 - Proximal Operator
- 3 Variational Problems
 - Variational Problems
 - Hamilton-Jacobi Equation And Hopf-lax Formula
 - Application
- 4 Application: Optimal Transport
 - Introduction.

1 Convex Analysis

- Convex sets and Convex Functions
 - Conjugate
 - Subdifferentiable
 - Dual
 - Dual Flat Manifold

2 Convex Optimization

- Proximal Operator

3 Variational Problems

- Variational Problems
- Hamilton-Jacobi Equation And Hopf-lax Formula
- Application

4 Application: Optimal Transport

- Introduction.

Definition

Some definition about the spaces we discuss.

- t.v.s(topological vector space)
- l.c.s(locally convex space)

Other definition: interal

Theorem

(Hahn-Banach)

Let V be a real t.v.s. A is an open non-empty convex set and M is a non-empty affine subspace which does not intersect A . Then there exists a closed affine hyperplane H which contains M and doesn't intersect A .

We also have an analytical form of the Hahn-Banach Theorem, however, we will not introduce it in our course.

Corollary

A is an open non-empty convex set and B is a non-empty convex set which does not intersect A. Then there exists a closed affine hyperplane H which separates A and B.

Corollary

The boundary point is a supporting point for the non-empty closed convex set.

Corollary

Every closed convex set is the intersection of the closed half-space which contain it.

Theorem

(Hahn-Banach)

Let V be a real t.v.s. A is an open non-empty convex set and M is a non-empty affine subspace which does not intersect A . Then there exists a closed affine hyperplane H which contains M and doesn't intersect A .

Remark

There are also several interesting corollaries.

- every closed convex set is weakly closed.
- Mazur's lemma is also a corollary of the Hahn-Banach theorem.

Definition: convex functions, epigraph

Theorem

The a function is convex is equal to its epigraph is convex

Definition

The following definition of l.s.c(lower semi-continuous) functions is equal:

- $\forall a \in \mathbb{R}$ the set $\{u \in V | F(u) \leq a\}$ is closed
- $\forall \bar{u} \in V$ we have $\underline{\lim}_{u \rightarrow \bar{u}} F(u) \geq F(\bar{u})$
- the epigraph is closed

For every mapping $F : V \rightarrow \overline{\mathbb{R}}$, the largest I.s.c minorant of F will be called the I.s.c regularization of F and will denoted by \overline{F} , then we have

- $\text{epi} \overline{F} = \overline{\text{epi} F}$
- $\forall u \in V, \overline{F}(u) = \underline{\lim}_{v \rightarrow u} F(v)$

Definition

The set of function $F : V \rightarrow \overline{\mathbb{R}}$ which are pointwise supremum of a family of continuous affine functions is denoted by $\Gamma(V)$ and $\Gamma_0(V)$ denotes the functions in $\Gamma(V)$ other than the constants, $+\infty$ and $-\infty$

Remark

The functions in $\Gamma(V)$ is nothing other than the functions that is convex and l.s.c.

Definition

Γ Regularization

F and G are two functions of V into $\overline{\mathbb{R}}$. T.F.A.E.

- 1 G is a pointwise supremum of the continuous affine functions everywhere less than F
- 2 G is the largest minorant of F in $\Gamma(V)$. G is then called the Γ -regularization.
- 3 $\text{epi}G = \overline{\text{coepi}F}$

- 1 **Convex Analysis**
 - Convex sets and Convex Functions
 - **Conjugate**
 - Subdifferentiable
 - Dual
 - Dual Flat Manifold

- 2 **Convex Optimization**
 - Proximal Operator

- 3 **Variational Problems**
 - Variational Problems
 - Hamilton-Jacobi Equation And Hopf-lax Formula
 - Application

- 4 **Application: Optimal Transport**
 - Introduction.

Polar Functions

In this section we consider the topology $\sigma(V, V^*)$. Let F be a function $o V$ into $\overline{\mathbb{R}}$, if $u^* \in V^*$ and $\alpha \in \mathbb{R}$, the continuous affine function $u \rightarrow \langle u, u^* \rangle - \alpha$ is everywhere less than F if and only if

$$\forall u \in V \text{ we have } \alpha \geq \langle u, u^* \rangle - F(u)$$

In this motivation we can define the polar function as

$$F^*(u^*) = \sup_{u \in V} \{ \langle u, u^* \rangle - F(u) \} \quad (1)$$

Then we have

- $F^*(0) = -\inf_{u \in V} F(u)$
- $F \leq G$ then we have $F^* \geq G^*$

Then we consider the bipolar function

$$F^{**}(u) = \sup_{u^* \in V^*} \{ \langle u, u^* \rangle - F^*(u^*) \}$$

We will show that

Bipolar is nothing other than the Γ -regularization

Corollary

$$F^* = F^{***}$$

Corollary

$$F = F^{**} \text{ if and only if } f \in \Gamma(V)$$

This means the polarity establishes a bijection between $\Gamma(V)$ and $\Gamma(V^*)$

- 1 **Convex Analysis**
 - Convex sets and Convex Functions
 - Conjugate
 - **Subdifferentiable**
 - Dual
 - Dual Flat Manifold

- 2 **Convex Optimization**
 - Proximal Operator

- 3 **Variational Problems**
 - Variational Problems
 - Hamilton-Jacobi Equation And Hopf-lax Formula
 - Application

- 4 **Application: Optimal Transport**
 - Introduction.

Definition

A function F of V into $\overline{\mathbb{R}}$ is said to be subdifferentiable of point $u \in V$ if it has a continuous affine minorant which is exact at u . The slope $u^* \in V^*$ of such a minorant is called a subgradient of F at u and the set of subgradients at u is called the subdifferential at u and denoted $\partial F(u)$

Theorem

$u^* \in \partial F(u)$ if and only if $F(u)$ is finite and

$$\langle v - u, u^* \rangle + F(u) \leq F(v), \forall v \in V$$

Theorem

If $\partial F(u) \neq \emptyset$ then $F(u) = F^{**}(u)$

If $F(u) = F^{**}(u)$, then $\partial F(u) = \partial F^{**}(u)$

Theorem

F is a function of V into $\overline{\mathbb{R}}$ and F^* is its polar. Then $u^* \in \partial F(u)$ if and only if

$$F(u) + F^*(u^*) = \langle u, u^* \rangle$$

Proof.

- If $F(u) + F^*(u^*) = \langle u, u^* \rangle$ then the affine function $\langle \cdot, u^* \rangle + F(u) - \langle u, u^* \rangle$ is everywhere less than F
- For $l(v) = \langle v - u, u^* \rangle + F(u)$ is maximal means the constant term $F(u) - \langle u, u^* \rangle$ is the maximal.



Theorem

$$u^* \in \partial F(u) \Leftrightarrow u \in \partial F^*(u^*)$$

Definition

$$\forall v \in V, \lim_{\lambda \rightarrow 0^+} \frac{F(u + \lambda v) - F(u)}{\lambda} = \langle v, F'(u) \rangle$$

For convex functions, the Gateaux-differential is the subgradient

Theorem

For convex function F we have

$$F(v) \geq F(u) + \langle F'(u), v - u \rangle$$

Theorem

We assume that the function F is convex, l.s.c., proper and is Gateaux-differentiable with continuous derivative F' , then if $u \in C$ T.F.A.E

- *u is a solution of $\min F$*
- *$\langle F'(u), v - u \rangle \geq 0, \forall v \in C$*
- *$\langle F'(v), v - u \rangle \geq 0, \forall v \in C$*

- 1 **Convex Analysis**
 - Convex sets and Convex Functions
 - Conjugate
 - Subdifferentiable
 - **Dual**
 - Dual Flat Manifold

- 2 **Convex Optimization**
 - Proximal Operator

- 3 **Variational Problems**
 - Variational Problems
 - Hamilton-Jacobi Equation And Hopf-lax Formula
 - Application

- 4 **Application: Optimal Transport**
 - Introduction.

Duality In Convex Optimization

If we have a primal problem

$$(P) \inf_{u \in V} F(u)$$

we can define a perturbed problem

$$(P_\rho) \inf_{i \in V} \Phi(u, \rho)$$

and $\Phi(u, 0) = F(u)$

Next we can define the dual problem as

$$(P^*) \sup_{p \in Y^*} \{-\Phi^*(0, p^*)\}$$

Duality In Convex Optimization

Why $\sup_{p \in Y^*} \{-\Phi^*(0, p^*)\}$?

Now assume $\Phi(u, p) \in \Gamma(V \times Y)$ and let $h(p) = \inf_{u \in V} \Phi(u, p)$

Theorem

- h is convex
- $h^*(p^*) = \Phi^*(0, p)$

Proof.

$$\begin{aligned} h^*(p^*) &= \sup_{p \in Y} \{\langle p, p^* \rangle - h(p)\} \\ &= \sup_{p \in Y} \left\{ \langle p, p^* \rangle - \inf_{u \in V} \{\Phi(u, p)\} \right\} \\ &= \sup_{p \in Y, u \in V} \{\langle p, p^* \rangle - \Phi(u, p)\} = \Phi^*(0, p) \end{aligned}$$



Duality In Convex Optimization

Why $\sup_{p \in Y^*} \{-\Phi^*(0, p^*)\}$?

For $h^*(p^*) = \Phi^*(0, p)$ we have

Theorem

- $\sup P^* = h^{**}(0)$
- *The set of solutions of P^* is identical of $\partial h^{**}(0)$*

Proof.

For the solution $p^* \in Y^*$ we have

$$-h^*(p^*) = \sup_{q^* \in Y^*} \{\langle 0, q^* \rangle - h^*(q^*)\} = h^{**}(0)$$



Next we consider the bidual problem, we can naturally associate the perturbed problem of P^* with respect to the perturbed problems

$$(P_{u^*}^*) \sup_{p^* \in Y^*} \{-\Phi^*(u^*, p^*)\}$$

The the bidual problem is

$$(P^{**}) \inf_{u \in V} \{\Phi^{**}(u, 0)\}$$

So if we assume $\Phi \in \Gamma(V \times Y)$, it is easy to see that the bidual is nothing other than the primal problem

Example1. The problem P takes the form $F(u) = J(u, Au)$ and takes $\Phi(u, p) = J(u, Au - p)$

Example2. (Lagrangian Dual) we consider the problem

$$\inf_{u \in A, Bu \leq 0} F(u)$$

take $\Phi(u, p) = \hat{F}(u) + X_{E_p}(u)$

- 1 **Convex Analysis**
 - Convex sets and Convex Functions
 - Conjugate
 - Subdifferentiable
 - Dual
 - **Dual Flat Manifold**

- 2 **Convex Optimization**
 - Proximal Operator

- 3 **Variational Problems**
 - Variational Problems
 - Hamilton-Jacobi Equation And Hopf-lax Formula
 - Application

- 4 **Application: Optimal Transport**
 - Introduction.

An n -dimensional manifold M is a set of points such that each point has n -dimensional extensions in its neighborhood.

Example Of Manifolds

- Euclidean Space
- Sphere

Divergence

A divergence $D[P : Q]$ is a function of its coordinate ξ_P and ξ_Q satisfies certain criteria. We may also express as $D[\xi_P : \xi_Q]$

Definition

$D[P : Q]$ is called a divergence when it satisfies the following criteria:

- $D[P : Q] \geq 0$
- $D[P : Q] = 0$ if and only if $P = Q$
- $D[\xi_P : \xi_P + d\xi] = \frac{1}{2} \sum g_{ij}(\xi_P) d\xi_i d\xi_j + O(|d\xi|^3)$

Here $G = (g_{ij})$ is a positive-definite matrix depending on ξ_P

Remark

In general $D[P : Q] \neq D[Q : P]$ and it is possible to symmetrize a divergence by

$$D_s[P : Q] = \frac{1}{2}(D[P : Q] + D[Q : P])$$

Definition

When P and Q are sufficiently close, we define the square of infinitesimal distance ds between them by

$$ds^2 = 2D[\xi : \xi + d\xi] = \sum g_{ij} d\xi_i d\xi_j$$

This means a divergence D provides M with a Riemannian structure.

Example

Kullback-Leibler Divergence
Euclidean Divergence

Definition

Bregman divergence is defined as

$$D_{\psi}[\xi : \xi_0] = \psi(\xi) - \psi(\xi_0) - \nabla\psi(\xi_0) \cdot (\xi - \xi_0)$$

Euclidean Divergence and Kullback-Leibler Divergence are all Bregman Divergence.

Dual Structure

We let $\xi^* = \nabla\psi(\xi)$, we have

$$\begin{aligned}\psi^*(\xi^*) &= \xi \cdot \xi^* - \psi(\xi) \\ \xi &= \nabla\psi^*(\xi^*)\end{aligned}$$

We define the Hessian matrix as its Riemann structure, we have

$$G^*(\xi^*) = \nabla\nabla\phi^*x(\xi^*) = \frac{\partial\xi}{\partial\xi^*}$$

For convex, proper, l.s.c function ψ we have

$$D_\psi[P : Q] = \psi(\xi_P) + \psi^*(\xi_Q^*) - \xi_P \cdot \xi_Q^*$$

and as a result

$$D_\psi[\xi : \xi_0] = D_{\psi^*}[\xi^* : \xi_0^*]$$

Dual Flat Riemannian Structure Derived from Convex Function

Remember the the riemannian metrix

$$ds^2 = 2D_\psi[\theta : \theta + d\theta] = \sum g_{ij} d\theta^i d\theta^j, g_{ij}(\theta) = \frac{\partial^2}{\partial \theta^i \partial \theta^j} \psi(\theta)$$

A small line element $d\theta$ is a tangent vector expressed by

$$d\theta = \sum d\theta^i e_i$$

Dually, we also can introduce a set of basis vectors $\{e^{*i}\}$, $d\theta^* = \sum d\theta^{*i} e^{*i}$, and at the same time

$$ds^2 = \langle e_i, e_j \rangle d\theta^i d\theta^j$$

Dual Flat Riemannian Structure Derived from Convex Function

Remember the fact that $G^*(\xi^*) = \nabla \nabla \phi^* x(\xi^*) = \frac{\partial \xi}{\partial \xi^*}$, so that

$$d\theta^* = G d\theta, d\theta_i^* = g_{ij} d\theta^j$$

Theorem

$$\langle e_i, e^{*j} \rangle = \delta_i^j$$

Corollary

If $A = \sum A^j e_j = \sum A_i e^{*i}$ we have

$$A_i = \sum g_{ij} A^j, A^j = g^{*ij} A_i$$

Generalized Pythagorean Theorem

Theorem

When triangle PQR is orthogonal such that the dual geodesic connecting P and Q is orthogonal to the geodesic connecting Q and R , the following generalized Pythagorean relation holds

$$D_\psi[P : R] = D_\psi[P : Q] + D_\psi[Q : R]$$

Proof.

For $D_\psi[P : Q] = \psi(\theta_P) + \psi^*(\theta^*) - \theta_P \cdot \theta_Q^*$, we have

$$D_\psi[P : R] - D_\psi[P : Q] - D_\psi[Q : R] = (\theta_P^* - \theta_Q^*) \cdot (\theta_Q - \theta_R)$$



Projection Theorem

We can define the divergence from a point P to submanifold S is defined by

$$D_\psi[P : S] = \min_{R \in S} D_\psi[P : R]$$

Theorem

(Projection Theorem) Given $P \in M$ and a smooth submanifold $S \subset M$, the point \hat{P}_S^* that minimizes the divergence $D_\psi[P : R]$, $R \in S$ is the dual geodesic projection of P to S . The point \hat{P}_S that minimizes the dual divergence $D_{\psi^*}[P : R]$, $R \in S$ is the geodesic projection of P to S .

- 1 Convex Analysis
 - Convex sets and Convex Functions
 - Conjugate
 - Subdifferentiable
 - Dual
 - Dual Flat Manifold

- 2 Convex Optimization
 - Proximal Operator

- 3 Variational Problems
 - Variational Problems
 - Hamilton-Jacobi Equation And Hopf-lax Formula
 - Application

- 4 Application: Optimal Transport
 - Introduction.

Implicit Gradient Descent

Now we consider the gradient descent

$$x^{(k+1)} = x^k - \Delta t \nabla f(x^k)$$

as the forward Euler scheme of the dynamic system $\partial_t u = -\nabla f(u)$

Now we consider the backward euler scheme

$$x^{(k+1)} = x^k - \Delta t \nabla f(x^{k+1})$$

Now we define the proximal mapping as

$$\text{prox}_{cf}(z) = \arg \min_x \left\{ f(x) + \frac{1}{2c} \|x - z\|^2 \right\}$$

The f.o.c. can be written as

$$0 = c\nabla f(x^*) + (x^* - z)$$

which can form the implicit gradient introduced before.

$$\text{prox}_{cf} = (I + c\partial f)^{-1}$$

We define

$$M_{cf}(x) = \inf_y \left\{ f(y) + \frac{1}{2c} \|y - x\|^2 \right\}$$

Here are some properties:

- $M_{cf} \in C^1$ even f is not
- $M_{cf} = \left((cf)^* + (1/2) \|\cdot\|^2 \right)^*$
- $\nabla M_{cf}(x) = (1/c)(x - \text{prox}_{cf}(x))$
- $\text{prox}_f(x) = \nabla M_{f^*}(x)$
- $\text{prox}_{cf}(x) = x - c \nabla M_{cf}(x)$

Assume that f has a minimizer, then iterate

$$x^{k+1} = \text{prox}_{c_k f}(x^k) \left(\sum_{k=1}^{\infty} c^k = \infty \right)$$

or we can have a step size

$$x^{k+1} = \text{prox}_{c_k f}(x^k) + (1 - \alpha^k)x^k \quad (2)$$

Question: What will happen if we apply the proximal algorithm to the dual question?

Answer: augmented Lagrangian method or the method of multipliers

Theorem

$$x = \text{prox}_f(x) + \text{prox}_{f^*}(x)$$

Proof.

$$\begin{aligned} u = \text{prox}_f(x) &\Leftrightarrow x - u \in \partial f(x) \\ &\Leftrightarrow x \in \partial f^*(x - u) \Leftrightarrow x - u = \text{prox}_{f^*}(x) \end{aligned}$$



- 1 Convex Analysis
 - Convex sets and Convex Functions
 - Conjugate
 - Subdifferentiable
 - Dual
 - Dual Flat Manifold

- 2 Convex Optimization
 - Proximal Operator

- 3 Variational Problems
 - Variational Problems
 - Hamilton-Jacobi Equation And Hopf-lax Formula
 - Application

- 4 Application: Optimal Transport
 - Introduction.

Variational Problems

Assume that $L : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ is a given smooth function, hereafter called the **Lagrangian**. (We denote the lagrangian as $L(q, \dot{x})$)
Next we introduce the action functional as

$$I[w(\cdot)] = \int_0^t L(\dot{w}(s), w(s)) ds \quad (3)$$

defined for functions $w(\cdot) = (w^1(\cdot), w^2(\cdot), \dots, w^n(\cdot))$ belonging to the admissible class

$$A = \{w(\cdot) \in C^2([0, t]; \mathbb{R}^n) \mid w(0) = y, w(t) = x\}$$

The basic problem in the calculus of variations is then to find a curve $x(\cdot)$ that satisfying

$$I[x(\cdot)] = \min_{w(\cdot) \in A} I[w(\cdot)]$$

Euler-Lagrange Equation

Assume that there is a function $x(\cdot)$ that minimize the variational problem

Theorem

The function $x(\cdot)$ solves the system of Euler-Lagrange equation

$$-\frac{d}{ds}(D_q L(\dot{x}(s), x(s))) + D_x L(\dot{x}(s), x(s)) = 0 \quad (0 \leq s \leq t) \quad (4)$$

This is a simple corollary of the characterization of the variational problem described by Gateaux derivative.

- 1 Convex Analysis
 - Convex sets and Convex Functions
 - Conjugate
 - Subdifferentiable
 - Dual
 - Dual Flat Manifold

- 2 Convex Optimization
 - Proximal Operator

- 3 Variational Problems
 - Variational Problems
 - **Hamilton-Jacobi Equation And Hopf-lax Formula**
 - Application

- 4 Application: Optimal Transport
 - Introduction.

For Hamilton-Jacobi equation $u_t + H(\nabla u, x) = 0$ we can calculate its characteristic as

$$\dot{x}(s) = D_p H(p(s), x(s))$$

$$\dot{p}(s) = -D_x H(p(s), x(s))$$

Next we use a variational perspective to understand this problem.

Hamilton's ODE

We set $x(s)$ is the solution of the variational problem (that is to say which satisfies the E-L equation) and $p(s) = D_q L(\dot{x}(s), x(s))$, which is called the generalized momentum corresponding to the position $x(\cdot)$ and velocity $\dot{x}(\cdot)$. Besides, we will also make a hypothesis that for all $x, p \in \mathbb{R}^n$, the equation $p = D_q L(q, x)$ can be uniquely solved for q as a smooth function of x and p .

Definition

The hamiltonian H associated with the Lagrangian L is

$$H(p, x) := p \cdot q(p, x) - L(q(p, x), x) \quad (p, x \in \mathbb{R}^n)$$

where q is defined implicitly as before.

Example

If the Lagrangian $L(q, x) = \frac{1}{2}m|q|^2 - \phi(x)$, what is the Hamiltonian

$$H(p, x) = p \cdot q(p, x) - L(q(p, x), x) \quad (p, x \in \mathbb{R}^n)$$
$$p = D_q L(q, x)$$

Theorem

The functions $x(\cdot)$ and $p(\cdot)$ satisfy Hamilton's equation

$$\dot{x}(s) = D_p H(p(s), x(s))$$

$$\dot{p}(s) = -D_x H(p(s), x(s))$$

for $0 \leq s \leq t$. Furthermore the mapping $s \rightarrow H(p(s), x(s))$ is constant.

Definition

The **Legendre Transform** of L is defined as

$$L^*(p) = \sup_{q \in \mathbb{R}^n} \{p \cdot q - L(q)\}$$

Motivation: If the mapping $q \rightarrow p \cdot q - L(q)$ has a maximum at $q = q^*$, then $p = DL(q^*)$. Therefore $L^*(p) = p \cdot q(p) - L(q(p))$.

Theorem

If L is convex and at the same time $\lim_{q \rightarrow \infty} \frac{L(q)}{|q|} = +\infty$, then H is convex and at the same time $\lim_{q \rightarrow \infty} \frac{H(q)}{|q|} = +\infty$ and $H = L^*$, $L = H^*$

In order to take the $w(0)$ into condition, we modify the question into

$$u(x, t) = \inf \left\{ \int_0^t L(\dot{w}(s)) ds + g(y) \mid w(0) = y, w(t) = x \right\}$$

Hopf-lax lemma shows that $u(x, t)$ is the solution of the following PDE

$$u_t + H(Du) = 0$$

with initial value $u = g$ on $\mathbb{R}^n \times 0$

Hopf-lax Lemma

We assume that L is smooth, convex and at the same time

$$\lim_{q \rightarrow \infty} \frac{L(q)}{|q|} = +\infty$$

Theorem

If $x \in \mathbb{R}^n$ and $t > 0$, then the solution $u = u(x, t)$ of the minimization problem is

$$u(x, t) = \min_{y \in \mathbb{R}^n} \left\{ tL\left(\frac{x-y}{t}\right) + g(y) \right\}$$

We can also know that

$$u(x, t) = \min_{y \in \mathbb{R}^n} \left\{ (t-s)L\left(\frac{x-y}{t-s}\right) + u(y, s) \right\}$$

- 1 Convex Analysis
 - Convex sets and Convex Functions
 - Conjugate
 - Subdifferentiable
 - Dual
 - Dual Flat Manifold

- 2 Convex Optimization
 - Proximal Operator

- 3 Variational Problems
 - Variational Problems
 - Hamilton-Jacobi Equation And Hopf-lax Formula
 - Application

- 4 Application: Optimal Transport
 - Introduction.

If we consider a two dimension example. Let $\Omega \subset \mathbb{R}^2$, and we minimize

$$I[u] = \frac{1}{2} \int_{\Omega} |\nabla u(x)|^2 dx - \int_{\Omega} f(x)u(x)dx$$

If u subjects to the boundary condition $u = 0$, we can get the solution of a PDE: $-\Delta u^*(x) = f(x)$

Next we consider the energy functional $I[u] = \frac{\int_0^1 (u'(x))^2 dx}{\int_0^1 (u(x))^2 dx}$

The function u minimize the functional is the eigenvector of the operator
"

Isoperimetric Inequality

Problem 4. (Isoperimetric inequality). Consider a closed plane curve described by a parametric equation $(x(t), y(t)), 0 \leq t \leq T$ with parameter t oriented counterclockwise and $(x(0), y(0)) = (x(T), y(T))$.

(a): Show that the total length of the curve is given by

$$L = \int_0^T \sqrt{(x'(t))^2 + (y'(t))^2} dt$$

(b): Show that the total area enclosed by the curve is given by

$$A = \frac{1}{2} \int_0^T (x(t)y'(t) - y(t)x'(t)) dt$$

(c): The classical iso-perimetric inequality states that for closed plane curves with a fixed length L , circles have the largest enclosed area A . Formulate this question into a variational problem.

(d): Derive the Euler-Lagrange equation for the variational problem in (c).

(e): Show that there are two constants x_0 and y_0 such that

$$(x(t) - x_0)^2 + (y(t) - y_0)^2 \equiv r^2$$

where $r = L/(2\pi)$. Explain your result.

- 1 Convex Analysis
 - Convex sets and Convex Functions
 - Conjugate
 - Subdifferentiable
 - Dual
 - Dual Flat Manifold
- 2 Convex Optimization
 - Proximal Operator
- 3 Variational Problems
 - Variational Problems
 - Hamilton-Jacobi Equation And Hopf-lax Formula
 - Application
- 4 Application: Optimal Transport
 - Introduction.

Monge

Objective: Calculate a transport map $T_{\#}\mu = \nu$ which minimize the transport cost

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

Kantorovich

Objective: Calculate a transport plane minimize the transport cost

$$c(\Pi) = \int c(x, y)\Pi(x, y)$$

The optimal transport is a convex problem, which can be formulated as

$$\begin{aligned} \min & \langle C, F \rangle \\ \text{s.t.} & \sum_i F_{i,j} = q_j \\ & \sum_j F_{i,j} = q_i \end{aligned}$$

is a special case of the linear programming:

$$\begin{aligned} \min & c^T x \\ \text{s.t.} & Ax = b, x \geq 0 \end{aligned}$$

Consider the dual problem of the linear programming problem.

Primal:

$$\min c^T x$$

$$\text{s.t. } Ax = b, x \geq 0$$

and we have the relation:

Dual:

$$\min b^T y$$

$$\text{s.t. } A^T y \leq c$$

$$\begin{aligned} \inf_{Ax=b, x \geq 0} c^T x &= \inf_{x \geq 0} \sup_y c^T x + y^T (b - Ax) \\ &= ? \sup_y \inf_{x \geq 0} c^T x - y^T Ax + y^T b \\ &= \sup_{A^T y \leq c} y^T b \end{aligned}$$

First, let us express the constraint $\gamma \in \Pi(\mu, \nu)$ in the following way.

$$\sup_{\phi, \psi} \int_X \phi d\mu + \int_Y \psi d\nu - \int_{X \times Y} (\phi(x) + \psi(y)) d\gamma$$

so that the primal problem can be expressed by

$$\min_{\gamma} \int_{X \times Y} + \sup_{\phi, \psi} \int_X \phi d\mu + \int_Y \psi d\nu - \int_{X \times Y} (\phi(x) + \psi(y)) d\gamma$$

then consider interchanging sup and inf:

$$\sup_{\phi, \psi} \int_X \phi d\mu + \int_Y \psi d\nu + \inf_{\gamma} \int_{X \times Y} (c(x, y) - (\phi(x) + \psi(y))) d\gamma$$

Kantorovich Dual

If the central notion in the original Monge-Kantorovich problem is cost, in the dual problem it is price.

Imagine that a company offers to take care of all your transportation problem, buying bread at the bakeries and selling them to the cafes. Let $\psi(x)$ be the price at which a baker of bread at the bakery x and selling them to the cafe y at the price $\phi(y)$

Let us maximize the profit:

$$\sup \left\{ \int_Y \phi(y) d\nu(y) - \int_X \psi(x) d\mu(x) \mid \phi(y) - \psi(x) \leq c(x, y) \right\}$$

It is easy to prove that

$$\begin{aligned} \sup_{\phi - \psi \leq c} \left\{ \int_Y \phi(y) d\nu(y) - \int_X \psi(x) d\mu(x) \right\} \\ \leq \inf_{\pi \in \Pi(\mu, \nu)} \left\{ \int_{X \times Y} c(x, y) d\pi(x, y) \right\} \end{aligned}$$

If we describe a pair of prices (ϕ, ψ) as tight if

$$\begin{aligned} \phi(y) &= \inf_x (\psi(x) + c(x, y)) \\ \psi(x) &= \sup_y (\phi(y) - c(x, y)) \end{aligned}$$

The following formula can be seen as the definition of c -transform.

Definition

Once a function $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ is given, we say that a set $\Gamma \subset X \times Y$ is *c-cyclically monotone* if for every $k \in \mathbb{N}$, every permutation σ and every finite family of points $(x_1, y_1), \dots, (x_k, y_k) \in \Gamma$ we have

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

Theorem

If γ is an optimal transport plane for the cost c and c is continuous, then $\text{spt}(\gamma)$ is a CM-set.

Theorem

Rockafellar's Theorem

If $\Gamma \neq \emptyset$ is a c -CM set in $X \times Y$ and $c : X \times Y \rightarrow \mathbb{R}$, then there exists a c -concave function $\phi : X \rightarrow \mathbb{R} \cup \{-\infty\}$ such that

$$\Gamma \subset \{(x, y) \in X \times Y : \phi(x) + \phi^c(y) = c(x, y)\}$$

Proof.

The function ϕ can be defined as

$$\phi(x) = \inf \{ c(x, y_n) - c(x_n, y_n) + c(x_n, y_{n-1}) - c(x_{n-1}, y_{n-1}) + \cdots \\ + c(x_1, y_0) - c(x_0, y_0) : n \in \mathbb{N}, (x_i, y_i) \in \Gamma \}$$



As a result, we can get a theorem as below.

Theorem

If c is C^1 , ϕ is a Kantorovich potential for the cost c in the transport from μ to ν , and (x_0, y_0) belongs to the support of an optimal transport plane γ , then $\nabla\phi(x_0) = \nabla_x c(x_0, y_0)$, provided ϕ is differentiable at x_0 .

As an example, if the cost function has the following form $c(x, y) = h(x - y)$. Then there exists an optimal transport plan γ for the cost $c(x, y)$ and is unique of the form $(id, T)_{\#}\mu$.

Moreover, there exists a Kantorovich potential ϕ and T and the potentials ϕ are linked by

$$T(x) = x - (\nabla h)^{-1}(\nabla\phi(x))$$

Quadratic Case

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

$$T(x) = x - \nabla\phi(x) = \nabla\left(\frac{x^2}{2} - \phi(x)\right) = \nabla u(x)$$

Theorem

For function $X : \mathbb{R}^n \rightarrow \mathbb{R}$, let us define $u_X = \frac{1}{2}|x|^2 - X(x)$, then we have

$$u_{X^c} = (u_X)^*$$

Proof.

$$\begin{aligned}u_{X^c}(x) &= \sup_y \frac{1}{2}|x|^2 - \frac{1}{2}|x - y|^2 + X(y) \\ &= \sup_y x \cdot y - \left(\frac{1}{2}|y|^2 - X(y)\right)\end{aligned}$$

Quadratic Case

We go further more on the quadratic case, we only need to minimize the $\int x \cdot y d\gamma$ gives the same result.

We can give the same result easier, actually we have $\phi(x_0) + \phi^*(y_0) = x_0 \cdot y_0$ for $y_0 \in \partial\phi(x_0)$, which means

Theorem

For the quadratic case, there exists unique an optimal transport map T from μ to ν and it is of the form $T = \nabla u$ for a convex function u





Optimal Transport is one of the topics you can choose as one of your pre at this class:

- More on Kantorovich Dual
- Benamou and Brenier problem
- **Richard Jordan, David Kinderlehrer, Felix Otto** The variational formulation of the Fokker-planck equation

Other topics we will cover

- PDEs and Numerical PDEs (may introduce some methods in CFD)
- Stochastic Modeling (including SDEs, Uq and model reduction)
- Information Geometry.

For Further Reading I

-  L.Evans.
Partial Differential Equation.
-  Ivar Ekeland, Roger Teman
Convex analysis and variational problems
-  Cedric Villani
Optimal Transport: Old and New
-  Richard Jordan, David Kinderlehrer, Felix Otto
The variational formulation of the Fokker-planck equation
Physic Reivew D,1998, 2000.