



Chapter 1

Convex Analysis

Convex Optimization for Machine Learning & Computer Vision
SS 2018

[Convex Set](#)
[Convex Function](#)
[Existence of Minimizer](#)
[Subdifferential](#)
[Convex Conjugate](#)
[Duality Theory](#)
[Proximal Operator](#)

Tao Wu
Emanuel Laude
Zhenzhang Ye

Computer Vision Group
Department of Informatics
TU Munich



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Convex Set

Notations

- \mathbb{E} is a Euclidean space (finite dimensional vector space), equipped with the inner product $\langle \cdot, \cdot \rangle$, e.g. $\langle u, v \rangle = u^\top v$.
- C is a closed, convex subset of \mathbb{E} .
- J is a convex objective function.

Convex optimization

$$\text{minimize } J(u) \quad \text{over } u \in C.$$

First questions:

- What is a convex set?
- What is a convex function?



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Definition

A set C is said to be **convex** if

$$\alpha u + (1 - \alpha)v \in C, \quad \forall u, v \in C, \forall \alpha \in [0, 1].$$



Convex Set

Convex Function

Existence of Minimizer

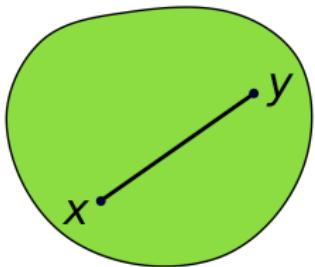
Subdifferential

Convex Conjugate

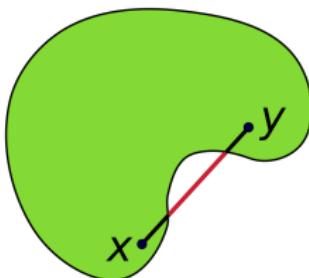
Duality Theory

Proximal Operator

convex



non-convex



Recall basic concepts in analysis

Convex Analysis

Tao Wu
Emmanuel Laude
Zhenzhang Ye

Definition

- A set $C \subset \mathbb{E}$ is **open** if $\forall u \in C, \exists \epsilon > 0$ s.t. $B_\epsilon(u) \subset C$, where $B_\epsilon(u) := \{v \in \mathbb{E} : \|v - u\| < \epsilon\}$.
- A set $C \subset \mathbb{E}$ is **closed** if its complement $\mathbb{E} \setminus C$ is open.
- The **closure** of a set $C \subset \mathbb{E}$ is

$$\text{cl } C = \{u \in \mathbb{E} : \exists \{u^k\} \subset C \text{ s.t. } \lim_{k \rightarrow \infty} u^k = u\}.$$

- The **interior** of a set $C \subset \mathbb{E}$ is

$$\text{int } C = \{u \in C : \exists \epsilon > 0 \text{ s.t. } B_\epsilon(u) \subset C\}.$$



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Recall basic concepts in analysis

Definition

- A set $C \subset \mathbb{E}$ is **open** if $\forall u \in C, \exists \epsilon > 0$ s.t. $B_\epsilon(u) \subset C$, where $B_\epsilon(u) := \{v \in \mathbb{E} : \|v - u\| < \epsilon\}$.
- A set $C \subset \mathbb{E}$ is **closed** if its complement $\mathbb{E} \setminus C$ is open.
- The **closure** of a set $C \subset \mathbb{E}$ is

$$\text{cl } C = \{u \in \mathbb{E} : \exists \{u^k\} \subset C \text{ s.t. } \lim_{k \rightarrow \infty} u^k = u\}.$$

- The **interior** of a set $C \subset \mathbb{E}$ is

$$\text{int } C = \{u \in C : \exists \epsilon > 0 \text{ s.t. } B_\epsilon(u) \subset C\}.$$

- The **relative interior** of a set $C \subset \mathbb{E}$ is

$$\begin{aligned} \text{rint } C &:= \{u \in C : \exists \epsilon > 0 \text{ s.t. } B_\epsilon(u) \cap \text{aff } C \subset C\} \\ &= \{u \in C : \forall v \in C, \exists \alpha > 1 \text{ s.t. } v + \alpha(u - v) \in C\} \end{aligned}$$

if C is convex. Here $\text{aff } C$ stands for the **affine hull** of C .



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Basic properties

The following operations preserve the convexity:

- Intersection: $C_1 \cap C_2$
 - Summation: $C_1 + C_2 := \{u^1 + u^2 : u^1 \in C_1, u^2 \in C_2\}$
 - Closure: $\text{cl } C$
 - Interior: $\text{int } C$
- The union of convex sets is not convex in general.



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Basic properties

The following operations preserve the convexity:

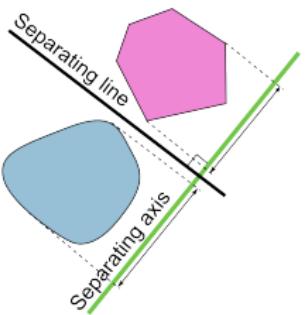
- Intersection: $C_1 \cap C_2$
- Summation: $C_1 + C_2 := \{u^1 + u^2 : u^1 \in C_1, u^2 \in C_2\}$

- Closure: $\text{cl } C$
- Interior: $\text{int } C$

- The union of convex sets is not convex in general.
- *Polyhedral sets* are always convex; *cones* are not necessarily convex.

Convex cone

C is a **cone** if $C = \alpha C$ for any $\alpha > 0$. C is a **convex cone** if C is a cone and is convex as well.



Source: Wikipedia.

Theorem (separation of convex sets)

Let C_1, C_2 be nonempty convex subsets in \mathbb{E} s.t. $C_1 \cap C_2 = \emptyset$ and C_1 is open. Then there exists a hyperplane separating C_1 and C_2 , i.e. $\exists v \in \mathbb{E}, v \neq 0, \alpha \in \mathbb{R}$ s.t.

$$\langle v, u^1 \rangle \geq \alpha \geq \langle v, u^2 \rangle, \quad \forall u^1 \in C_1, u^2 \in C_2.$$

Proof: on board.

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Separation of convex sets

Theorem (separation of convex sets)

Let C_1, C_2 be nonempty convex subsets in \mathbb{E} s.t. $C_1 \cap C_2 = \emptyset$ and C_1 is open. Then there exists a hyperplane separating C_1 and C_2 , i.e. $\exists v \in \mathbb{E}, v \neq 0, \alpha \in \mathbb{R}$ s.t.

$$\langle v, u^1 \rangle \geq \alpha \geq \langle v, u^2 \rangle, \quad \forall u^1 \in C_1, u^2 \in C_2.$$

Proof: on board.

Remarks

- ① The proof works in any Hilbert space.
- ② Corollary: In a Hilbert space, any (strongly) closed convex subset C is weakly closed.
- ③ The above theorem generalizes to any topological vector space, known as the *Hahn-Banach theorem*.



Convex Function

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



- An **extended real-valued function** J maps from \mathbb{E} to $\overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$.
- The **domain** of $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ is

$$\text{dom } J = \{u \in \mathbb{E} : J(u) < \infty\}.$$

- The function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ is **proper** if $\text{dom } J \neq \emptyset$.

Definition

We say $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ is a **convex function** if

- ① $\text{dom } J$ is a convex set.
- ② For all $u, v \in \text{dom } J$ and $\alpha \in [0, 1]$ it holds that

$$J(\alpha u + (1 - \alpha)v) \leq \alpha J(u) + (1 - \alpha)J(v).$$

We say J is **strictly convex** if the above inequality is strict for all $\alpha \in (0, 1)$ and $u \neq v$.

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



- $J_{data}(u) = \|u - f\|_q^q$, where $q \geq 1$ and $\|\cdot\|_q$ is ℓ^q -norm.
- $J_{regu}(u) = \|Ku\|_{q'}^{q'}$, where K is linear transform and $q' \geq 1$.
- $J(u) = J_{data}(u) + \alpha J_{regu}(u)$.

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Examples

- $J_{data}(u) = \|u - f\|_q^q$, where $q \geq 1$ and $\|\cdot\|_q$ is ℓ^q -norm.
- $J_{regu}(u) = \|Ku\|_{q'}^{q'}$, where K is linear transform and $q' \geq 1$.
- $J(u) = J_{data}(u) + \alpha J_{regu}(u)$.
- (Binary) entropy: $J_\epsilon(u) = \epsilon(u \log(u) + (1 - u) \log(1 - u))$.
- Soft max: $J_\epsilon^*(v) = \epsilon \log(1 + \exp(v/\epsilon)) \approx \max(v, 0)$.

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Examples

- $J_{data}(u) = \|u - f\|_q^q$, where $q \geq 1$ and $\|\cdot\|_q$ is ℓ^q -norm.
- $J_{regu}(u) = \|Ku\|_{q'}^{q'}$, where K is linear transform and $q' \geq 1$.
- $J(u) = J_{data}(u) + \alpha J_{regu}(u)$.
- (Binary) entropy: $J_\epsilon(u) = \epsilon(u \log(u) + (1-u) \log(1-u))$.
- Soft max: $J_\epsilon^*(v) = \epsilon \log(1 + \exp(v/\epsilon)) \approx \max(v, 0)$.
- **Indicator function** ($C \subset \mathbb{E}$ is closed and convex):

$$\delta_C(u) = \begin{cases} 0 & \text{if } u \in C, \\ \infty & \text{otherwise.} \end{cases}$$

- Formulate constrained optimization with indicator function:

$$\min J(u) \text{ over } u \in C \Leftrightarrow \min J(u) + \delta_C(u) \text{ over } u \in \mathbb{E}.$$



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Basic facts

(As exercises)

- Any norm (over a normed vector space) is a convex function.
- J is a convex function and K is a linear transform
 $\Rightarrow J(K \cdot)$ is convex function.
- (Jensen's inequality) $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ is convex iff

$$J\left(\sum_{i=1}^n \alpha_i u^i\right) \leq \sum_{i=1}^n \alpha_i J(u^i),$$

whenever $\{u^i\}_{i=1}^n \subset \mathbb{E}$, $\{\alpha_i\}_{i=1}^n \subset [0, 1]$, $\sum_{i=1}^n \alpha_i = 1$.

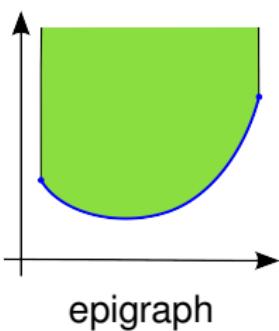


Epigraph

Definition

The **epigraph** of a proper function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ is

$$\text{epi } J = \{(u, \alpha) \in \mathbb{E} \times \mathbb{R} : J(u) \leq \alpha\}.$$



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Theorem

A proper function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ is convex (resp. strictly convex) iff $\text{epi } J$ is a convex (resp. strictly convex) set.

Proof: as exercise.



Definition

Assume $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ with $\text{rint dom } J \neq \emptyset$. We say J is **locally Lipschitz** at $u \in \text{rint dom } J$ with modulus $L_u > 0$ if there exists $\epsilon > 0$ s.t.

$$|J(u^1) - J(u^2)| \leq L_u \|u^1 - u^2\| \quad \forall u^1, u^2 \in B_\epsilon(u) \cap \text{rint dom } J.$$

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Definition

Assume $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ with $\text{rint dom } J \neq \emptyset$. We say J is **locally Lipschitz** at $u \in \text{rint dom } J$ with modulus $L_u > 0$ if there exists $\epsilon > 0$ s.t.

$$|J(u^1) - J(u^2)| \leq L_u \|u^1 - u^2\| \quad \forall u^1, u^2 \in B_\epsilon(u) \cap \text{rint dom } J.$$

Theorem

A proper convex function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ is locally Lipschitz at any $u \in \text{rint dom } J$.

Proof: on board.

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Existence of Minimizer

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Recall the optimization of $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$:

$$\text{minimize } J(u) \quad \text{over } u \in \mathbb{E}.$$

Definition

- ① $u^* \in \mathbb{E}$ is a **global minimizer** if $J(u^*) \leq J(u)$ for all $u \in \mathbb{E}$.
- ② u^* is a **local minimizer** if $\exists \epsilon > 0$ s.t. $J(u^*) \leq J(u)$ for all $u \in B_\epsilon(u^*)$.
- ③ In the above definitions, a global/local minimizer is **strict** if $J(u^*) \leq J(u)$ is replaced by $J(u^*) < J(u)$.



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Recall the optimization of $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$:

$$\text{minimize } J(u) \quad \text{over } u \in \mathbb{E}.$$

Definition

- ① $u^* \in \mathbb{E}$ is a **global minimizer** if $J(u^*) \leq J(u)$ for all $u \in \mathbb{E}$.
- ② u^* is a **local minimizer** if $\exists \epsilon > 0$ s.t. $J(u^*) \leq J(u)$ for all $u \in B_\epsilon(u^*)$.
- ③ In the above definitions, a global/local minimizer is **strict** if $J(u^*) \leq J(u)$ is replaced by $J(u^*) < J(u)$.

Theorem

For any proper convex function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, if $u^* \in \text{dom } J$ is a local minimizer of J , then it is also a global minimizer.

Proof: on board.



Convex Set
Convex Function
Existence of Minimizer
Subdifferential
Convex Conjugate
Duality Theory
Proximal Operator



Does a minimizer always exist?

- Consider

$$\text{minimize } J(u) \quad \text{over } u \in \mathbb{E},$$

where $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ is a proper, convex function.

Convex Set

Convex Function

Existence of Minimizer

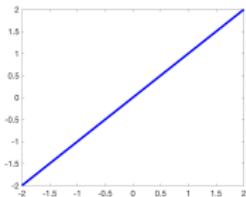
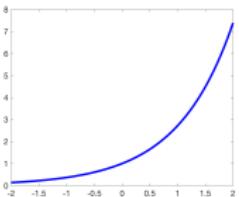
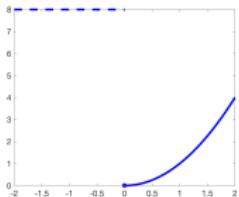
Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

- Some counterexamples for $J : \mathbb{R} \rightarrow \overline{\mathbb{R}}$:

 u  $\exp u$  $u^2 + \delta\{u > 0\}$ 

- We shall formalize our observations and derive sufficient conditions for existence.

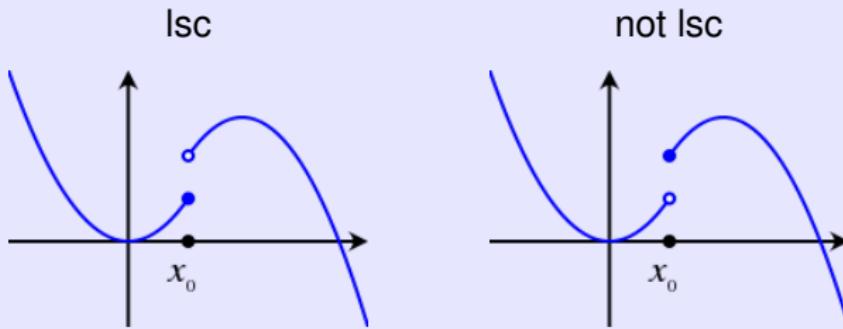


Sufficient conditions for existence

Definition

- ① J is **bounded from below** if $J(\cdot) \geq C$ for some $C \in \mathbb{R}$.
- ② J is **coercive** if $J(u) \rightarrow \infty$ whenever $\|u\| \rightarrow \infty$.
 - Proposition: J is coercive if $\text{dom } J$ is bounded.
- ③ J is **lower semi-continuous** (lsc) at u^* if

$$J(u^*) \leq \liminf_{u \rightarrow u^*} J(u).$$



- Proposition: J is lsc iff $\text{epi } J$ is closed.

Definition

- ① J is **bounded from below** if $J(\cdot) \geq C$ for some $C \in \mathbb{R}$.
- ② J is **coercive** if $J(u) \rightarrow \infty$ whenever $\|u\| \rightarrow \infty$.
- ③ J is **lower semi-continuous** (lsc) at u^* if

$$J(u^*) \leq \liminf_{u \rightarrow u^*} J(u).$$

[Convex Set](#)[Convex Function](#)[Existence of Minimizer](#)[Subdifferential](#)[Convex Conjugate](#)[Duality Theory](#)[Proximal Operator](#)

Theorem

Any proper function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, which is bounded from below, coercive, and lsc (everywhere), has a (global) minimizer.

Proof: on board.

Definition

- ① J is **bounded from below** if $J(\cdot) \geq C$ for some $C \in \mathbb{R}$.
- ② J is **coercive** if $J(u) \rightarrow \infty$ whenever $\|u\| \rightarrow \infty$.
- ③ J is **lower semi-continuous** (lsc) at u^* if

$$J(u^*) \leq \liminf_{u \rightarrow u^*} J(u).$$



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Theorem

Any proper function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, which is bounded from below, coercive, and lsc (everywhere), has a (global) minimizer.

Proof: on board.

Remarks for infinite dimensions

- ① Weak compactness in reflexive Banach (e.g. Hilbert) sp.
- ② J is convex and strongly lsc $\Rightarrow J$ is weakly lsc.



- Recall that a function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ is strictly convex if

$$J(\alpha u + (1 - \alpha)v) < \alpha J(u) + (1 - \alpha)J(v),$$

for all $u, v \in \text{dom } J$, $u \neq v$, $\alpha \in (0, 1)$.

Theorem

The minimizer of a strictly convex function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ is unique.

Proof: on board.

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Subdifferential

Convex Set

Convex Function

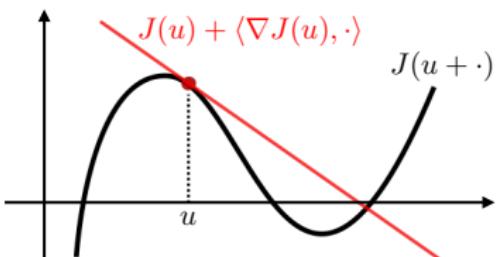
Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

**Definition**

$J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ is called (Fréchet) **differentiable** at $u \in \text{int dom } J$ and $\nabla J(u) \in \mathbb{E}$ is the (Fréchet) **differential** of J at u if

$$\lim_{h \rightarrow 0} \frac{|J(u+h) - J(u) - \langle \nabla J(u), h \rangle|}{\|h\|} = 0.$$

J is **continuously differentiable** at $u \in \text{int dom } J$ if $\nabla J(\cdot)$ is continuous on $(\text{dom } J) \cap B_\epsilon(u)$ for some $\epsilon > 0$.



Convex Set

Convex Function

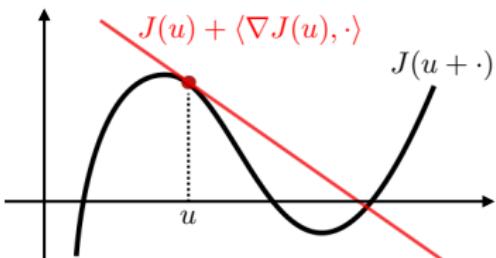
Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

**Definition**

$J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ is called (Fréchet) **differentiable** at $u \in \text{int dom } J$ and $\nabla J(u) \in \mathbb{E}$ is the (Fréchet) **differential** of J at u if

$$\lim_{h \rightarrow 0} \frac{|J(u + h) - J(u) - \langle \nabla J(u), h \rangle|}{\|h\|} = 0.$$

J is **continuously differentiable** at $u \in \text{int dom } J$ if $\nabla J(\cdot)$ is continuous on $(\text{dom } J) \cap B_\epsilon(u)$ for some $\epsilon > 0$.

Remark

If \mathbb{E} is a topological vector space, $\nabla J(u)$ is treated as a *dual* object in \mathbb{E}^* , and $\langle \nabla J(u), h \rangle_{\mathbb{E}^*, \mathbb{E}}$ as *duality pairing*.



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

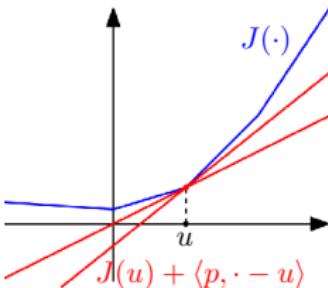
Convex Conjugate

Duality Theory

Proximal Operator

Subdifferential

Now we generalize differentiability from differentiable functions to non-differentiable convex functions.



Definition

The **subdifferential** of a convex function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ at $u \in \text{dom } J$ is defined by

$$\partial J(u) = \{p \in \mathbb{E} : J(v) \geq J(u) + \langle p, v - u \rangle \quad \forall v \in \mathbb{E}\}.$$



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

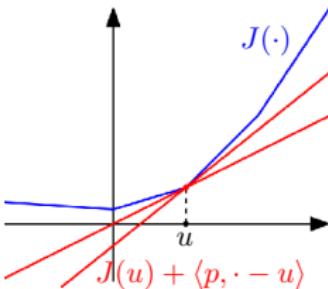
Convex Conjugate

Duality Theory

Proximal Operator

Subdifferential

Now we generalize differentiability from differentiable functions to non-differentiable convex functions.



Definition

The **subdifferential** of a convex function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ at $u \in \text{dom } J$ is defined by

$$\partial J(u) = \{p \in \mathbb{E} : J(v) \geq J(u) + \langle p, v - u \rangle \quad \forall v \in \mathbb{E}\}.$$

Geometric interpretation

$p \in \partial J(u)$ iff $(p, -1)$ is a normal vector for the supporting hyperplane of $\text{epi } J$ at $(u, J(u))$.

Basic facts

- ① $\partial J(\cdot)$ is a *set-valued* map.
- ② If J is cont. differentiable at u , then $\partial J(u) = \{\nabla J(u)\}$.



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Basic facts

- ① $\partial J(\cdot)$ is a *set-valued* map.
- ② If J is cont. differentiable at u , then $\partial J(u) = \{\nabla J(u)\}$.



Examples

- ① $J(u) = \|u\| \Rightarrow \partial J(0) = \{p : \|p\|_* \leq 1\}$, where $\|\cdot\|_*$ is the **dual norm** of $\|\cdot\|$, i.e., $\|p\|_* = \sup\{\langle p, u \rangle : \|u\| \leq 1\}$.
- ② Given any closed, convex subset $C \subset \mathbb{E}$ and $u \in C$,

$$\partial \delta_C(u) = \{p \in \mathbb{E} : \langle p, v - u \rangle \leq 0 \ \forall v \in C\} =: N_C(u),$$

known as the **normal cone** of C at u .

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Examples and basic facts

Basic facts

- ① $\partial J(\cdot)$ is a *set-valued* map.
- ② If J is cont. differentiable at u , then $\partial J(u) = \{\nabla J(u)\}$.

Examples

- ① $J(u) = \|u\| \Rightarrow \partial J(0) = \{p : \|p\|_* \leq 1\}$, where $\|\cdot\|_*$ is the **dual norm** of $\|\cdot\|$, i.e., $\|p\|_* = \sup\{\langle p, u \rangle : \|u\| \leq 1\}$.
- ② Given any closed, convex subset $C \subset \mathbb{E}$ and $u \in C$,

$$\partial \delta_C(u) = \{p \in \mathbb{E} : \langle p, v - u \rangle \leq 0 \ \forall v \in C\} =: N_C(u),$$

known as the **normal cone** of C at u .

- ③ (Exercise) $X \in \mathbb{R}^{m \times n} \mapsto \|X\|_{1,2} = \sum_{i=1}^m \left(\sum_{j=1}^n |X_{i,j}|^2 \right)^{1/2}$.
- ④ (Exercise) $X \in \mathbb{R}^{m \times n} \mapsto \|X\|_{nuclear} = \sum_i \sigma_i(X)$, i.e., sum of singular values.

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Theorem (chain rule under linear transform)

Let $\tilde{J}(\cdot) = J(K \cdot)$ with some convex function J and linear transform K . Then

$$\partial \tilde{J}(u) = K^\top \partial J(Ku)$$

whenever $Ku \in \text{rint dom } J$.

Example: $J(u) = \|Ku\| \Rightarrow \partial J(u) = K^\top \partial \|\cdot\|(Ku).$



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Theorem (chain rule under linear transform)

Let $\tilde{J}(\cdot) = J(K\cdot)$ with some convex function J and linear transform K . Then

$$\partial \tilde{J}(u) = K^\top \partial J(Ku)$$

whenever $Ku \in \text{rint dom } J$.

Example: $J(u) = \|Ku\| \Rightarrow \partial J(u) = K^\top \partial \|\cdot\|(Ku)$.

Theorem (summation rule)

If $\tilde{J}(\cdot) = J_1(\cdot) + J_2(\cdot)$ for some convex functions J_1 and J_2 , then

$$\partial \tilde{J}(u) = \partial J_1(u) + \partial J_2(u)$$

for any $u \in \text{rint dom } J_1 \cap \text{rint dom } J_2$.

Warning: not true if J_1 or J_2 is non-convex, e.g. $0 = |\cdot| + (-|\cdot|)$.

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Properties of subdifferential map

Theorem

Let $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ be a convex function. Then ∂J is a **monotone operator**, i.e. $\forall u^1, u^2 \in \text{dom } J, p^1 \in \partial J(u^1), p^2 \in \partial J(u^2) :$

$$\langle p^1 - p^2, u^1 - u^2 \rangle \geq 0.$$

Proof: on board.



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Properties of subdifferential map

Theorem

Let $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ be a convex function. Then ∂J is a **monotone operator**, i.e. $\forall u^1, u^2 \in \text{dom } J, p^1 \in \partial J(u^1), p^2 \in \partial J(u^2)$:

$$\langle p^1 - p^2, u^1 - u^2 \rangle \geq 0.$$

Proof: on board.

Theorem

Let $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ be a convex function. Then for any $u \in \text{int dom } J$, $\partial J(u)$ is a nonempty, compact, and convex subset.

Proof: on board.



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Properties of subdifferential map

Theorem

Let $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ be a convex function. Then ∂J is a **monotone operator**, i.e. $\forall u^1, u^2 \in \text{dom } J, p^1 \in \partial J(u^1), p^2 \in \partial J(u^2)$:

$$\langle p^1 - p^2, u^1 - u^2 \rangle \geq 0.$$

Proof: on board.

Theorem

Let $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ be a convex function. Then for any $u \in \text{int dom } J$, $\partial J(u)$ is a nonempty, compact, and convex subset.

Proof: on board.

Theorem

Let $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ be a proper, convex, lsc function. Then the set-valued map $\partial J(\cdot)$ is **closed**, i.e. $p^* \in \partial J(u^*)$ whenever

$$\exists (u^k, p^k) \rightarrow (u^*, p^*) \in (\text{dom } J) \times \mathbb{E} \text{ s.t. } p^k \in \partial J(u^k) \ \forall k.$$

Proof: on board.



Optimality condition

Theorem

Given any proper convex function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, the sufficient and necessary condition for u^* being a (global) minimizer for J is

$$0 \in \partial J(u^*).$$

Proof: on board.

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Optimality condition

Theorem

Given any proper convex function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, the sufficient and necessary condition for u^* being a (global) minimizer for J is

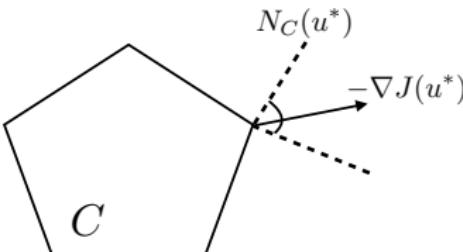
$$0 \in \partial J(u^*).$$

Proof: on board.

Constrained optimization as special case

If u^* minimizes $\tilde{J} = J + \delta_C$ with convex function $J : \mathbb{E} \rightarrow \mathbb{R}$ and closed convex subset $C \subset \mathbb{E}$, then $0 \in \partial \tilde{J}(u^*) \Leftrightarrow$

$$0 \in \partial J(u^*) + N_C(u^*).$$



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Convex Set

Convex Function

Existence of Minimizer

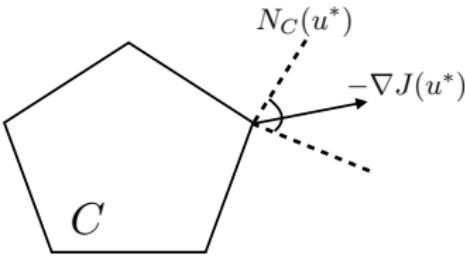
Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

$$0 \in \partial J(u^*) + N_C(u^*).$$



Remark

The optimality condition $0 \in \partial J(u^*) + N_C(u^*)$ is *geometric*. More explicit characterization replies on the *algebraic* representation of $N_C(u^*)$, e.g., the **Karush-Kuhn-Tucker (KKT) conditions**, typically under certain *constraint qualifications*.



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Optimality condition

Constrained optimization as special case

If u^* minimizes $\tilde{J} = J + \delta_C$ with convex function $J : \mathbb{E} \rightarrow \mathbb{R}$ and closed convex subset $C \subset \mathbb{E}$, then $0 \in \partial \tilde{J}(u^*) \Leftrightarrow$

$$0 \in \partial J(u^*) + N_C(u^*).$$

Remark

The optimality condition $0 \in \partial J(u^*) + N_C(u^*)$ is *geometric*. More explicit characterization replies on the *algebraic* representation of $N_C(u^*)$, e.g., the **Karush-Kuhn-Tucker (KKT) conditions**, typically under certain *constraint qualifications*.

Example: Linear-inequality constraints

Let $C = \{u \in \mathbb{R}^n : Au \leq b\}$ where $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$ has linearly independent rows. Then

$$N_C(u) = \{A^\top \lambda : \lambda \geq 0, \lambda_i = 0 \text{ if } (Au - b)_i < 0\}.$$



Convex Conjugate

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Convex conjugate

Definition

Given a function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, the **convex conjugate** (a.k.a. Legendre-Fenchel transform) of J is defined by

$$J^*(p) = \sup_{u \in \mathbb{E}} \{ \langle u, p \rangle - J(u) \} \quad \forall p \in \mathbb{E}.$$

[Convex Set](#)[Convex Function](#)[Existence of Minimizer](#)[Subdifferential](#)[Convex Conjugate](#)[Duality Theory](#)[Proximal Operator](#)



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Convex conjugate

Definition

Given a function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, the **convex conjugate** (a.k.a. Legendre-Fenchel transform) of J is defined by

$$J^*(p) = \sup_{u \in \mathbb{E}} \{ \langle u, p \rangle - J(u) \} \quad \forall p \in \mathbb{E}.$$

Examples (as exercise)

- ① $J(u) = \langle w, u \rangle \Rightarrow J^*(p) = \delta\{p = w\}. (w \in \mathbb{E} \text{ given})$
- ② $J(u) = \|u\| \Rightarrow J^*(p) = \delta\{\|p\|_* \leq 1\}. (\|\cdot\|_* \text{ is the dual norm of } \|\cdot\|, \text{ i.e. } \|p\|_* = \sup\{\langle p, u \rangle : \|u\| \leq 1\})$
- ③ $J(u) = \frac{1}{q} \|u\|_q^q \Rightarrow J^*(p) = \frac{1}{q'} \|p\|_{q'}^{q'}. (q \in [1, \infty], \frac{1}{q} + \frac{1}{q'} = 1)$
- ④ $J(u) = \sum_i u_i \log u_i + \delta_{\Delta^{n-1}}(u) \Rightarrow J^*(p) = \log(\sum_i \exp(p_i)).$



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Convex conjugate

Definition

Given a function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, the **convex conjugate** (a.k.a. Legendre-Fenchel transform) of J is defined by

$$J^*(p) = \sup_{u \in \mathbb{E}} \{ \langle u, p \rangle - J(u) \} \quad \forall p \in \mathbb{E}.$$

Examples (as exercise)

- ① $J(u) = \langle w, u \rangle \Rightarrow J^*(p) = \delta\{p = w\}. (w \in \mathbb{E} \text{ given})$
- ② $J(u) = \|u\| \Rightarrow J^*(p) = \delta\{\|p\|_* \leq 1\}. (\|\cdot\|_* \text{ is the dual norm of } \|\cdot\|, \text{ i.e. } \|p\|_* = \sup\{\langle p, u \rangle : \|u\| \leq 1\})$
- ③ $J(u) = \frac{1}{q} \|u\|_q^q \Rightarrow J^*(p) = \frac{1}{q'} \|p\|_{q'}^{q'}. (q \in [1, \infty], \frac{1}{q} + \frac{1}{q'} = 1)$
- ④ $J(u) = \sum_i u_i \log u_i + \delta_{\Delta^{n-1}}(u) \Rightarrow J^*(p) = \log(\sum_i \exp(p_i)).$

Basic facts (as exercise)

- Scalar multiplication: $\tilde{J}(\cdot) = \alpha J(\cdot) \Rightarrow \tilde{J}^*(\cdot) = \alpha J^*(\cdot/\alpha).$
- Translation: $\tilde{J}(\cdot) = J(\cdot - z) \Rightarrow \tilde{J}^*(\cdot) = J^*(\cdot) + \langle \cdot, z \rangle.$



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

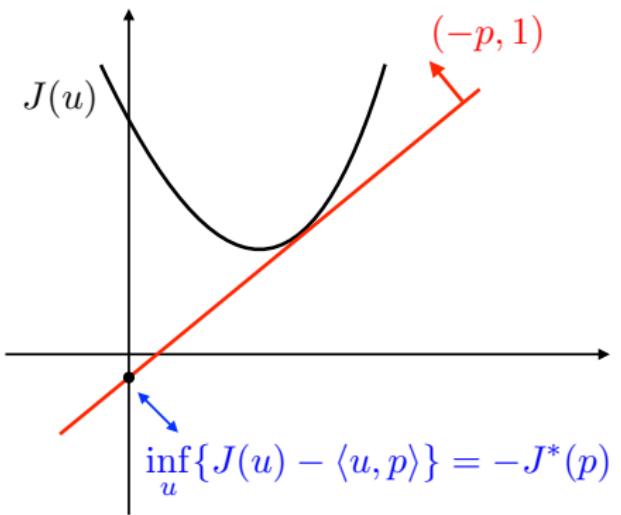
Geometric interpretation

Geometrically, convex conjugation maps

the normal vector of a supporting hyperplane to the epigraph

to

the intersection with the vertical axis.





Fenchel-Young inequality, order reversing property

Theorem (Fenchel-Young inequality)

For any $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ and $(u, p) \in \mathbb{E} \times \mathbb{E}$, we have

$$J(u) + J^*(p) \geq \langle u, p \rangle.$$

The equality holds iff $p \in \partial J(u)$ for $(u, p) \in \text{dom } J \times \text{dom } J^*$.

Proof: (i) $J(u) + J^*(p) \geq \langle u, p \rangle$ follows directly from the definition of convex conjugate.

(ii) The equality holds only if $(u, p) \in \text{dom } J \times \text{dom } J^*$.

Moreover, $p \in \partial J(u)$ is the sufficient and necessary condition for: $\min_{u \in \mathbb{E}} J(u) - \langle u, p \rangle$.

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Fenchel-Young inequality, order reversing property

Theorem (Fenchel-Young inequality)

For any $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ and $(u, p) \in \mathbb{E} \times \mathbb{E}$, we have

$$J(u) + J^*(p) \geq \langle u, p \rangle.$$

The equality holds iff $p \in \partial J(u)$ for $(u, p) \in \text{dom } J \times \text{dom } J^*$.

Proof: (i) $J(u) + J^*(p) \geq \langle u, p \rangle$ follows directly from the definition of convex conjugate.

(ii) The equality holds only if $(u, p) \in \text{dom } J \times \text{dom } J^*$.

Moreover, $p \in \partial J(u)$ is the sufficient and necessary condition for: $\min_{u \in \mathbb{E}} J(u) - \langle u, p \rangle$.

Theorem (order reversing)

For any $J_1, J_2 : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, we have $J_1^* \leq J_2^*$ whenever $J_1 \geq J_2$.

Proof: For any (u, p) , we have $\langle u, p \rangle - J_1(u) \leq \langle u, p \rangle - J_2(u)$. Taking supremum over u on both sides yields $J_1^*(p) \leq J_2^*(p)$.



Theorem

Let $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, and $J^{**} = (J^*)^*$ is the **biconjugate** of J .

In general:

- ① $J^{**}(\cdot) \leq J(\cdot)$.
- ② J^* is convex and lsc.

If J is proper, convex, and lsc, then:

- ③ $J^{**}(\cdot) = J(\cdot)$.
- ④ $p \in \partial J(u)$ iff $u \in \partial J^*(p)$.

Proof: on board.

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Definition

- ① $J : \mathbb{E} \rightarrow \mathbb{R}$ is **μ -strongly convex** if $\exists \mu > 0$ s.t. $J(\cdot) - \frac{\mu}{2} \|\cdot\|^2$ is convex.
- ② $J : \mathbb{E} \rightarrow \mathbb{R}$ is **L -Lipschitz differentiable** (a.k.a. L -smooth) if J is differentiable and ∇J is Lipschitz with modulus L .



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



- [Convex Set](#)
- [Convex Function](#)
- [Existence of Minimizer](#)
- [Subdifferential](#)
- [Convex Conjugate](#)
- [Duality Theory](#)
- [Proximal Operator](#)

Regularity of J and J^*

Definition

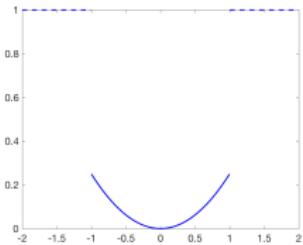
- ① $J : \mathbb{E} \rightarrow \mathbb{R}$ is **μ -strongly convex** if $\exists \mu > 0$ s.t. $J(\cdot) - \frac{\mu}{2} \|\cdot\|^2$ is convex.
- ② $J : \mathbb{E} \rightarrow \mathbb{R}$ is **L -Lipschitz differentiable** (a.k.a. L -smooth) if J is differentiable and ∇J is Lipschitz with modulus L .

Theorem

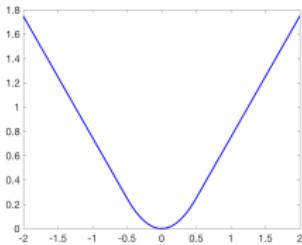
Assume that $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ is proper, convex, and lsc. Then J is μ -strongly convex iff J^* is $\frac{1}{\mu}$ -Lipschitz differentiable.

Proof: on board.

truncated quadratic



Huber function





Duality Theory

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



- Consider

$$\inf_{u \in \mathbb{R}^n} \{F(Ku) + G(u)\},$$

where $K \in \mathbb{R}^{m \times n}$, and F, G are proper, convex, and lsc.

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



[Convex Set](#)
[Convex Function](#)
[Existence of Minimizer](#)
[Subdifferential](#)
[Convex Conjugate](#)
Duality Theory
[Proximal Operator](#)

Fenchel-Rockafellar duality

- Consider

$$\inf_{u \in \mathbb{R}^n} \{F(Ku) + G(u)\},$$

where $K \in \mathbb{R}^{m \times n}$, and F, G are proper, convex, and lsc.

- The **weak duality** always holds:

$$\begin{aligned}
 \mathcal{P}^* &:= \inf_u \{F(Ku) + G(u)\} \\
 &= \inf_u \sup_p \{\langle p, Ku \rangle - F^*(p) + G(u)\} \\
 &\geq \sup_p \inf_u \{\langle K^\top p, u \rangle + G(u) - F^*(p)\} \\
 &= \sup_p \{-G^*(-K^\top p) - F^*(p)\} =: \mathcal{D}^*.
 \end{aligned}$$



[Convex Set](#)
[Convex Function](#)
[Existence of Minimizer](#)
[Subdifferential](#)
[Convex Conjugate](#)
Duality Theory
[Proximal Operator](#)

Fenchel-Rockafellar duality

- Consider

$$\inf_{u \in \mathbb{R}^n} \{F(Ku) + G(u)\},$$

where $K \in \mathbb{R}^{m \times n}$, and F, G are proper, convex, and lsc.

- The **weak duality** always holds:

$$\begin{aligned}
 \mathcal{P}^* &:= \inf_u \{F(Ku) + G(u)\} \\
 &= \inf_u \sup_p \{\langle p, Ku \rangle - F^*(p) + G(u)\} \\
 &\geq \sup_p \inf_u \{\langle K^\top p, u \rangle + G(u) - F^*(p)\} \\
 &= \sup_p \{-G^*(-K^\top p) - F^*(p)\} =: \mathcal{D}^*.
 \end{aligned}$$

- Define the **duality gap**:

$$\mathcal{G}(u, p) = F(Ku) + G(u) + G^*(-K^\top p) + F^*(p).$$

Note that $\mathcal{G}(u, p) = 0$ is an optimality criterion.



- $\mathcal{G}(u^*, p^*) = 0 \Leftrightarrow \mathcal{P}^* = \mathcal{D}^* \Leftrightarrow (u^*, p^*)$ solves the **saddle point problem** with $\mathcal{L}(u, p) := \langle p, Ku \rangle - F^*(p) + G(u)$:

$$\mathcal{L}(u^*, p) \leq \mathcal{L}(u^*, p^*) \leq \mathcal{L}(u, p^*) \quad \forall (u, p).$$

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Fenchel-Rockafellar duality

- $\mathcal{G}(u^*, p^*) = 0 \Leftrightarrow \mathcal{P}^* = \mathcal{D}^* \Leftrightarrow (u^*, p^*)$ solves the **saddle point problem** with $\mathcal{L}(u, p) := \langle p, Ku \rangle - F^*(p) + G(u)$:

$$\mathcal{L}(u^*, p) \leq \mathcal{L}(u^*, p^*) \leq \mathcal{L}(u, p^*) \quad \forall (u, p).$$

Theorem (Fenchel-Rockafellar duality)

Assume $\exists \bar{u} \in \text{dom } G$ s.t. F is continuous at $K\bar{u}$. Then the **strong duality** holds: $\mathcal{P}^* = \mathcal{D}^*$. Moreover, (u^*, p^*) is the optimal solution pair iff

$$\begin{cases} Ku^* \in \partial F^*(p^*), \\ -K^\top p^* \in \partial G(u^*). \end{cases}$$

Proof: on board.



Example: Total-variation image restoration

- Primal problem (given $\Omega \subset \mathbb{R}^d$, $f \in \mathbb{R}^\Omega$, $\alpha > 0$, $q \in [1, \infty]$):

$$\min_{u \in \mathbb{R}^\Omega} \alpha \|\nabla u\|_{1,q} + \frac{1}{2} \|u - f\|^2.$$

Here $\|p\|_{1,q} = \sum_{j \in \Omega} |p_j|^{\ell q}$ for each $p \in \mathbb{R}^{|\Omega| \times d}$.

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Example: Total-variation image restoration

- Primal problem (given $\Omega \subset \mathbb{R}^d$, $f \in \mathbb{R}^\Omega$, $\alpha > 0$, $q \in [1, \infty]$):

$$\min_{u \in \mathbb{R}^\Omega} \alpha \|\nabla u\|_{1,q} + \frac{1}{2} \|u - f\|^2.$$

Here $\|p\|_{1,q} = \sum_{j \in \Omega} |p_j|^{\ell q}$ for each $p \in \mathbb{R}^{|\Omega| \times d}$.

- Apply the Fenchel-Rockafellar duality with

$$F(\cdot) = \alpha \|\cdot\|_{1,q}, \quad K = \nabla, \quad G(\cdot) = \frac{1}{2} \|\cdot - f\|^2.$$

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Example: Total-variation image restoration

- Primal problem (given $\Omega \subset \mathbb{R}^d$, $f \in \mathbb{R}^\Omega$, $\alpha > 0$, $q \in [1, \infty]$):

$$\min_{u \in \mathbb{R}^\Omega} \alpha \|\nabla u\|_{1,q} + \frac{1}{2} \|u - f\|^2.$$

Here $\|p\|_{1,q} = \sum_{j \in \Omega} |p_j|^{1/q}$ for each $p \in \mathbb{R}^{|\Omega| \times d}$.

- Apply the Fenchel-Rockafellar duality with

$$F(\cdot) = \alpha \|\cdot\|_{1,q}, \quad K = \nabla, \quad G(\cdot) = \frac{1}{2} \|\cdot - f\|^2.$$

- Saddle point problem ($1/q + 1/q' = 1$):

$$\max_p \min_u \langle p, \nabla u \rangle - \delta\{\|p\|_{\infty, q'} \leq \alpha\} + \frac{1}{2} \|u - f\|^2.$$



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Example: Total-variation image restoration

- Primal problem (given $\Omega \subset \mathbb{R}^d$, $f \in \mathbb{R}^\Omega$, $\alpha > 0$, $q \in [1, \infty]$):

$$\min_{u \in \mathbb{R}^\Omega} \alpha \|\nabla u\|_{1,q} + \frac{1}{2} \|u - f\|^2.$$

Here $\|p\|_{1,q} = \sum_{j \in \Omega} |p_j|^{\ell q}$ for each $p \in \mathbb{R}^{|\Omega| \times d}$.

- Apply the Fenchel-Rockafellar duality with

$$F(\cdot) = \alpha \|\cdot\|_{1,q}, \quad K = \nabla, \quad G(\cdot) = \frac{1}{2} \|\cdot - f\|^2.$$

- Saddle point problem ($1/q + 1/q' = 1$):

$$\max_p \min_u \langle p, \nabla u \rangle - \delta\{\|p\|_{\infty, q'} \leq \alpha\} + \frac{1}{2} \|u - f\|^2.$$

- Dual problem:

$$\min_p \frac{1}{2} \|\nabla^\top p\|^2 + \langle \nabla^\top p, f \rangle + \delta\{\|p\|_{\infty, q'} \leq \alpha\}.$$



Proximal Operator

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Proximal operator

Definition

Given a proper, convex, lsc function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, we define the **proximal operator** of J by

$$\text{prox}_{\tau J}(v) = \arg \min_u J(u) + \frac{1}{2\tau} \|u - v\|^2.$$

Note: The above minimization always has a unique minimizer.

[Convex Set](#)[Convex Function](#)[Existence of Minimizer](#)[Subdifferential](#)[Convex Conjugate](#)[Duality Theory](#)[Proximal Operator](#)



Proximal operator

Definition

Given a proper, convex, lsc function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, we define the **proximal operator** of J by

$$\text{prox}_{\tau J}(v) = \arg \min_u J(u) + \frac{1}{2\tau} \|u - v\|^2.$$

Note: The above minimization always has a unique minimizer.

Proximal operator in an optimization algorithm

Let us solve the convex optimization:

$$\min_u F(u) + G(u),$$

where G is cont'lly differentiable but F is non-differentiable.

The **proximal gradient** iteration appears as:

$$u^{k+1} = \text{prox}_{\tau F}(u^k - \tau \nabla G(u^k)).$$

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Observations

- ① By checking the optimality condition,

$$u = \text{prox}_{\tau J}(v) \Leftrightarrow 0 \in \tau \partial J(u) + u - v \Leftrightarrow u = (I + \tau \partial J)^{-1}(v).$$

Thus, $\text{prox}_{\tau J} = (I + \tau \partial J)^{-1}$, a.k.a. the **resolvent** of ∂J .



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Observations

- ① By checking the optimality condition,

$$u = \text{prox}_{\tau J}(v) \Leftrightarrow 0 \in \tau \partial J(u) + u - v \Leftrightarrow u = (I + \tau \partial J)^{-1}(v).$$

Thus, $\text{prox}_{\tau J} = (I + \tau \partial J)^{-1}$, a.k.a. the **resolvent** of ∂J .

- ② u^* is a **fixed point** of $\text{prox}_{\tau J}$, i.e. $u^* = \text{prox}_{\tau J}(u^*)$,
 $\Leftrightarrow 0 \in \partial J(u^*)$, i.e. u^* is a minimizer of J .



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Observations

- ① By checking the optimality condition,

$$u = \text{prox}_{\tau J}(v) \Leftrightarrow 0 \in \tau \partial J(u) + u - v \Leftrightarrow u = (I + \tau \partial J)^{-1}(v).$$

Thus, $\text{prox}_{\tau J} = (I + \tau \partial J)^{-1}$, a.k.a. the **resolvent** of ∂J .

- ② u^* is a **fixed point** of $\text{prox}_{\tau J}$, i.e. $u^* = \text{prox}_{\tau J}(u^*)$,
 $\Leftrightarrow 0 \in \partial J(u^*)$, i.e. u^* is a minimizer of J .

- ③ Derivation of proximal gradient algorithm:

$$\begin{aligned} u^* &\in \arg \min_u F(u) + G(u) \\ &\Leftrightarrow 0 \in \partial F(u^*) + \nabla G(u^*) \\ &\Leftrightarrow u^* + \tau \partial F(u^*) \ni u^* - \tau \nabla G(u^*) \\ &\Leftrightarrow u^* = (I + \tau \partial F)^{-1}(u^* - \tau \nabla G(u^*)) \\ &\rightsquigarrow u^{k+1} = \text{prox}_{\tau F}(u^k - \tau \nabla G(u^k)). \end{aligned}$$



Examples

- ① Indicator function. Let C be nonempty, closed, convex \Rightarrow

$$\text{prox}_{\tau \delta_C}(\cdot) = \text{proj}_C(\cdot).$$

In this case: proximal gradient \Leftrightarrow projected gradient.

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Examples

- ① Indicator function. Let C be nonempty, closed, convex \Rightarrow

$$\text{prox}_{\tau \delta_C}(\cdot) = \text{proj}_C(\cdot).$$

In this case: proximal gradient \Leftrightarrow projected gradient.

- ② Linear approximation. $\tilde{J}(\cdot) := J(\bar{u}) + \langle \nabla J(\bar{u}), \cdot - \bar{u} \rangle \Rightarrow$

$$\begin{aligned}\text{prox}_{\tau \tilde{J}}(\bar{u}) &= \arg \min_u \left\{ \frac{1}{2\tau} \|u - \bar{u}\|^2 + \langle \nabla J(\bar{u}), u - \bar{u} \rangle \right\} \\ &= \bar{u} - \tau \nabla J(\bar{u}). \quad (\text{gradient descent step})\end{aligned}$$

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Examples

- ① Indicator function. Let C be nonempty, closed, convex \Rightarrow

$$\text{prox}_{\tau \delta_C}(\cdot) = \text{proj}_C(\cdot).$$

In this case: proximal gradient \Leftrightarrow projected gradient.

- ② Linear approximation. $\tilde{J}(\cdot) := J(\bar{u}) + \langle \nabla J(\bar{u}), \cdot - \bar{u} \rangle \Rightarrow$

$$\begin{aligned}\text{prox}_{\tau \tilde{J}}(\bar{u}) &= \arg \min_u \left\{ \frac{1}{2\tau} \|u - \bar{u}\|^2 + \langle \nabla J(\bar{u}), u - \bar{u} \rangle \right\} \\ &= \bar{u} - \tau \nabla J(\bar{u}). \quad (\text{gradient descent step})\end{aligned}$$

- ③ Quadratic approximation.

$$\tilde{J}(\cdot) := J(\bar{u}) + \langle \nabla J(\bar{u}), \cdot - \bar{u} \rangle + \frac{1}{2} \langle \nabla^2 J(\bar{u})(\cdot - \bar{u}), \cdot - \bar{u} \rangle \Rightarrow$$

$$\begin{aligned}\text{prox}_{\tau \tilde{J}}(\bar{u}) &= \arg \min_u \left\{ \frac{1}{2\tau} \|u - \bar{u}\|^2 + \langle \nabla J(\bar{u}), u - \bar{u} \rangle \right. \\ &\quad \left. + \frac{1}{2} \langle \nabla^2 J(\bar{u})(u - \bar{u}), u - \bar{u} \rangle \right\} \\ &= \bar{u} - \left(\nabla^2 J(\bar{u}) + \frac{1}{\tau} I \right)^{-1} \nabla J(\bar{u}). \quad (\text{damped Newton step})\end{aligned}$$

Logistic regression (programming exercise)



- MNIST¹ dataset - handwritten digit recognition.
- Train classifier on training set; Evaluate on test set.
- Conv Neural Network: 0.23%; **Logistic Regression:** 10%.

Convex Analysis

Tao Wu
Emmanuel Laude
Zhenzhang Ye



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

¹<http://yann.lecun.com/exdb/mnist/>

Logistic regression (programming exercise)

Convex Analysis

Tao Wu
Emmanuel Laude
Zhenzhang Ye



- MNIST¹ dataset - handwritten digit recognition.
- Train classifier on training set; Evaluate on test set.
- Conv Neural Network: 0.23%; **Logistic Regression:** 10%.
- $K = 10$ classes; grayscale images of pixels $M = 28 \times 28$.
- $N = 60000$ training images $X \in \mathbb{R}^{N \times M}$, with ground-truth labels $Y \in \{1, \dots, K\}^N$.



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

¹<http://yann.lecun.com/exdb/mnist/>

Logistic regression (programming exercise)

Convex Analysis

Tao Wu
Emmanuel Laude
Zhenzhang Ye



- MNIST¹ dataset - handwritten digit recognition.
- Train classifier on training set; Evaluate on test set.
- Conv Neural Network: 0.23%; **Logistic Regression:** 10%.
- $K = 10$ classes; grayscale images of pixels $M = 28 \times 28$.
- $N = 60000$ training images $X \in \mathbb{R}^{N \times M}$, with ground-truth labels $Y \in \{1, \dots, K\}^N$.
- Task: Train a *linear classifier* with $W \in \mathbb{R}^{K \times M}$ and $b \in \mathbb{R}^K$,
- ..., which parameterizes *likelihood* via *softmax*:

$$\mathcal{P}(Y_n = k | X_{n,\cdot}, W) = \frac{\exp(\langle W_{k,\cdot}, X_{n,\cdot} \rangle + b_k)}{\sum_{k'=1}^K \exp(\langle W_{k',\cdot}, X_{n,\cdot} \rangle + b_{k'})}.$$



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

¹<http://yann.lecun.com/exdb/mnist/>

Logistic regression (programming exercise)

Convex Analysis

Tao Wu
Emmanuel Laude
Zhenzhang Ye



- MNIST¹ dataset - handwritten digit recognition.
- Train classifier on training set; Evaluate on test set.
- Conv Neural Network: 0.23%; **Logistic Regression:** 10%.
- $K = 10$ classes; grayscale images of pixels $M = 28 \times 28$.
- $N = 60000$ training images $X \in \mathbb{R}^{N \times M}$, with ground-truth labels $Y \in \{1, \dots, K\}^N$.
- Task: Train a *linear classifier* with $W \in \mathbb{R}^{K \times M}$ and $b \in \mathbb{R}^K$,
- ..., which parameterizes *likelihood* via *softmax*:

$$\mathcal{P}(Y_n = k | X_{n,\cdot}, W) = \frac{\exp(\langle W_{k,\cdot}, X_{n,\cdot} \rangle + b_k)}{\sum_{k'=1}^K \exp(\langle W_{k',\cdot}, X_{n,\cdot} \rangle + b_{k'})}.$$

- Minimize: negative log-likelihood \mathcal{P} + regularizer \mathcal{R} , i.e.

$$\min_{W,b} \mathcal{R}(W, b) - \frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K \mathbf{1}_{Y_n=k} \log \mathcal{P}(Y_n = k | X_{n,\cdot}, W).$$

¹<http://yann.lecun.com/exdb/mnist/>

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

[Convex Set](#)[Convex Function](#)[Existence of Minimizer](#)[Subdifferential](#)[Convex Conjugate](#)[Duality Theory](#)[Proximal Operator](#)

Moreau identity

Theorem (Moreau identity)

Let $\tau > 0$ and $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ be proper, convex, and lsc. Then the following identity holds:

$$\text{id}(\cdot) = \text{prox}_{\tau J}(\cdot) + \tau \text{prox}_{\frac{1}{\tau}J^*}(\cdot/\tau).$$

In particular, $\tau = 1 \Rightarrow \text{id}(\cdot) = \text{prox}_J(\cdot) + \text{prox}_{J^*}(\cdot)$.

Proof: $v = \tau \text{prox}_{\frac{1}{\tau}J^*}(u/\tau)$

$$\Leftrightarrow \left(I + \frac{1}{\tau} \partial J^* \right)^{-1} (u/\tau) = v/\tau$$

$$\Leftrightarrow \partial J^*(v/\tau) \ni u - v$$

$$\Leftrightarrow v/\tau \in \partial J(u - v)$$

$$\Leftrightarrow u - v = (I + \tau \partial J)^{-1}(u) = \text{prox}_{\tau J}(u).$$



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Moreau identity

Theorem (Moreau identity)

Let $\tau > 0$ and $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ be proper, convex, and lsc. Then the following identity holds:

$$\text{id}(\cdot) = \text{prox}_{\tau J}(\cdot) + \tau \text{prox}_{\frac{1}{\tau}J^*}(\cdot/\tau).$$

In particular, $\tau = 1 \Rightarrow \text{id}(\cdot) = \text{prox}_J(\cdot) + \text{prox}_{J^*}(\cdot)$.

Proof: $v = \tau \text{prox}_{\frac{1}{\tau}J^*}(u/\tau)$

$$\Leftrightarrow \left(I + \frac{1}{\tau} \partial J^* \right)^{-1} (u/\tau) = v/\tau$$

$$\Leftrightarrow \partial J^*(v/\tau) \ni u - v$$

$$\Leftrightarrow v/\tau \in \partial J(u - v)$$

$$\Leftrightarrow u - v = (I + \tau \partial J)^{-1}(u) = \text{prox}_{\tau J}(u).$$

Remark

The Moreau identity suggests that if one of $\text{prox}_J(\cdot)$ and $\text{prox}_{J^*}(\cdot)$ is computable, so is the other.



Infimal convolution

Definition

Let $F, G : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ be proper, convex, and lsc. The **infimal convolution** (or inf convolution) of F and G is defined by

$$(F \square G)(u) = \inf_{v \in \mathbb{E}} \{F(u - v) + G(v)\},$$

with $\text{dom}(F \square G) = \text{dom } F + \text{dom } G$.

[Convex Set](#)[Convex Function](#)[Existence of Minimizer](#)[Subdifferential](#)[Convex Conjugate](#)[Duality Theory](#)[Proximal Operator](#)



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Infimal convolution

Definition

Let $F, G : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ be proper, convex, and lsc. The **infimal convolution** (or inf convolution) of F and G is defined by

$$(F \square G)(u) = \inf_{v \in \mathbb{E}} \{F(u - v) + G(v)\},$$

with $\text{dom}(F \square G) = \text{dom } F + \text{dom } G$.

Theorem

Let $F, G : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ be proper, convex, and lsc. Then

$$(F \square G)^* = F^* + G^*.$$

Proof: $(F \square G)^*(p) = \sup_{u,v} \{\langle p, u \rangle - F(v) - G(u - v)\} = \sup_{u,v} \{\langle p, v \rangle - F(v) + \langle p, u - v \rangle - G(u - v)\} = F^*(p) + G^*(p)$.

Analogy to integral convolution

By convolution theorem, $\widehat{F * G} = \widehat{F} \cdot \widehat{G}$ where $\widehat{\cdot}$ denotes the Fourier transform and $*$ the integral convolution.



Moreau envelope

Definition

The **Moreau envelope** of a proper, convex, lsc function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ is defined for each $u \in \mathbb{E}$ by

$$\begin{aligned}\text{env}_{\tau J}(u) &:= \left(J \square \frac{1}{2\tau} \|\cdot\|^2 \right)(u) \\ &= \inf_{v \in \mathbb{E}} \left\{ J(v) + \frac{1}{2\tau} \|v - u\|^2 \right\} \\ &= J(\text{prox}_{\tau J}(u)) + \frac{1}{2\tau} \|\text{prox}_{\tau J}(u) - u\|^2.\end{aligned}$$

[Convex Set](#)
[Convex Function](#)
[Existence of Minimizer](#)
[Subdifferential](#)
[Convex Conjugate](#)
[Duality Theory](#)
[Proximal Operator](#)



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Moreau envelope

Definition

The **Moreau envelope** of a proper, convex, lsc function $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ is defined for each $u \in \mathbb{E}$ by

$$\begin{aligned}\text{env}_{\tau J}(u) &:= \left(J \square \frac{1}{2\tau} \|\cdot\|^2 \right)(u) \\ &= \inf_{v \in \mathbb{E}} \left\{ J(v) + \frac{1}{2\tau} \|v - u\|^2 \right\} \\ &= J(\text{prox}_{\tau J}(u)) + \frac{1}{2\tau} \|\text{prox}_{\tau J}(u) - u\|^2.\end{aligned}$$

Example

$J : u \mapsto \|u\| \Rightarrow \text{env}_{\tau J}$ is the Huber function:

$$\text{env}_{\tau J}(u) = \begin{cases} \frac{1}{2\tau} \|u\|^2 & \text{if } \|u\| \leq \tau, \\ \|u\| - \frac{\tau}{2} & \text{if } \|u\| > \tau. \end{cases}$$

Observation: $\text{env}_{\tau J}$ performs smoothing on J .



- Recall the theorem: $(F \square G)^* = F^* + G^* \Rightarrow$

$$(\text{env}_{\tau J})^* = J^* + \left(\frac{1}{2\tau} \|\cdot\|^2 \right)^* = J^* + \frac{\tau}{2} \|\cdot\|^2.$$

[Convex Set](#)[Convex Function](#)[Existence of Minimizer](#)[Subdifferential](#)[Convex Conjugate](#)[Duality Theory](#)[Proximal Operator](#)



Properties of Moreau envelope

- Recall the theorem: $(F \square G)^* = F^* + G^* \Rightarrow$

$$(\text{env}_{\tau J})^* = J^* + \left(\frac{1}{2\tau} \|\cdot\|^2 \right)^* = J^* + \frac{\tau}{2} \|\cdot\|^2.$$

- Recall the theorem: J is μ -strongly convex iff J^* is $\frac{1}{\mu}$ -Lipschitz differentiable.
 $\Rightarrow \text{env}_{\tau J}$ is $\frac{1}{\tau}$ -Lipschitz differentiable.

[Convex Set](#)

[Convex Function](#)

[Existence of Minimizer](#)

[Subdifferential](#)

[Convex Conjugate](#)

[Duality Theory](#)

[Proximal Operator](#)



Properties of Moreau envelope

- Recall the theorem: $(F \square G)^* = F^* + G^* \Rightarrow$

$$(\text{env}_{\tau J})^* = J^* + \left(\frac{1}{2\tau} \|\cdot\|^2 \right)^* = J^* + \frac{\tau}{2} \|\cdot\|^2.$$

- Recall the theorem: J is μ -strongly convex iff J^* is $\frac{1}{\mu}$ -Lipschitz differentiable.

$\Rightarrow \text{env}_{\tau J}$ is $\frac{1}{\tau}$ -Lipschitz differentiable.

- $\nabla \text{env}_{\tau J}$ can be calculated as:

$$\begin{aligned} p = \nabla \text{env}_{\tau J}(u) &\Leftrightarrow u \in \partial(\text{env}_{\tau J})^*(p) = \partial J^*(p) + \tau p \\ &\Leftrightarrow u - \tau p \in \partial J^*(p) \Leftrightarrow \partial J(u - \tau p) \ni p \\ &\Leftrightarrow \tau \partial J(u - \tau p) \ni \tau p \Leftrightarrow (I + \tau \partial J)(u - \tau p) \ni u \\ &\Leftrightarrow u - \tau p = (I + \tau \partial J)^{-1}(u) = \text{prox}_{\tau J}(u) \\ &\Leftrightarrow \nabla \text{env}_{\tau J}(u) = p = \frac{1}{\tau}(u - \text{prox}_{\tau J}(u)). \end{aligned}$$

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator