



Chapter 1

Convex Analysis

Convex Optimization for Machine Learning & Computer Vision
WS 2019/20

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Tao Wu
Zhenzhang Ye

Computer Vision Group
Department of Informatics
TU Munich



Convex Set

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Notations

- \mathbb{E} is a *Euclidean space* (i.e., finite dimensional inner product space), equipped with

① Inner product $\langle \cdot, \cdot \rangle$, e.g., $\langle u, v \rangle = u^\top v$ if $\mathbb{E} = \mathbb{R}^n$;

② Norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$ satisfying polarization identity:

$$2\|u\|^2 + 2\|v\|^2 = \|u + v\|^2 + \|u - v\|^2.$$

- C is a closed, convex subset of \mathbb{E} .
- J is a convex *objective* function.

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



Notations

- \mathbb{E} is a *Euclidean space* (i.e., finite dimensional inner product space), equipped with

① Inner product $\langle \cdot, \cdot \rangle$, e.g., $\langle u, v \rangle = u^\top v$ if $\mathbb{E} = \mathbb{R}^n$;

② Norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$ satisfying polarization identity:

$$2\|u\|^2 + 2\|v\|^2 = \|u + v\|^2 + \|u - v\|^2.$$

- C is a closed, convex subset of \mathbb{E} .
- J is a convex *objective* function.

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Convex optimization

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

First questions:

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

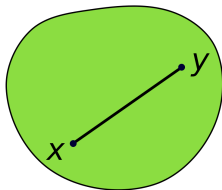
Convex set

Definition

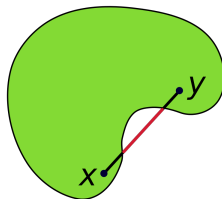
A set C is said to be **convex** if

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

convex



non-convex



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

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

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



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

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

with $\text{aff } C$ the **affine hull** of C . If C is a *convex* set, then

$$\text{rint } C = \{u \in C : \forall v \in C, \exists \alpha > 1 \text{ s.t. } v + \alpha(u - v) \in C\}.$$

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 and relative interior: $\text{int } C, \text{rint } C$.

In general, the union of convex sets is not convex.



Basic properties

The following operations preserve the convexity:

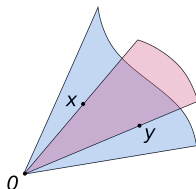
- Intersection: $C_1 \cap C_2$.
- Summation: $C_1 + C_2 := \{u^1 + u^2 : u^1 \in C_1, u^2 \in C_2\}$.
- Closure: $\text{cl } C$.
- Interior and relative interior: $\text{int } C, \text{rint } C$.

In general, the union of convex sets is not 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.



Convex cone.





Theorem (separation of convex sets)

Let C_1, C_2 be nonempty convex subsets of \mathbb{E} .

- ① Assume C_1 is closed and $C_2 = \{w\} \subset \mathbb{E} \setminus C_1$. Then $\exists v \in \mathbb{E}, v \neq 0, \alpha \in \mathbb{R}$ s.t.

$$\langle v, w \rangle > \alpha \geq \langle v, u \rangle, \quad \forall u \in C_1.$$

- ② Assume C_1 is open and $C_2 = \{w\} \subset \mathbb{E} \setminus C_1$. Then $\exists v \in \mathbb{E}, v \neq 0, \alpha \in \mathbb{R}$ s.t.

$$\langle v, w \rangle \geq \alpha \geq \langle v, u \rangle, \quad \forall u \in C_1.$$

- ③ Assume $C_1 \cap C_2 = \emptyset$ and C_1 is open. Then $\exists v \in \mathbb{E}, v \neq 0, \alpha \in \mathbb{R}$ s.t.

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

- ④ Assume $\emptyset \neq \text{int } C_1 \subset \mathbb{E} \setminus C_2$. Then $\exists v \in \mathbb{E}, v \neq 0, \alpha \in \mathbb{R}$ s.t.

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

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 Function



- 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 $u \neq v$.

Examples

- $J_{data}(u) = \|u - z\|_p^p$, where $p \geq 1$ and $\|\cdot\|_p$ is ℓ^p -norm.
- $J_{regu}(u) = \|Ku\|_q^q$, where K is linear transform and $q \geq 1$.
- $J(u) = J_{data}(u) + \alpha J_{regu}(u)$, where $\alpha > 0$.



Examples

- $J_{data}(u) = \|u - z\|_p^p$, where $p \geq 1$ and $\|\cdot\|_p$ is ℓ^p -norm.
- $J_{regu}(u) = \|Ku\|_q^q$, where K is linear transform and $q \geq 1$.
- $J(u) = J_{data}(u) + \alpha J_{regu}(u)$, where $\alpha > 0$.
- Negative binary entropy ($\epsilon > 0$):
 $J_\epsilon(u) = \epsilon(u \log(u) + (1 - u) \log(1 - u))$.
- Soft plus: $J_\epsilon(v) = \epsilon \log(1 + \exp(v/\epsilon))$.



Examples

- $J_{data}(u) = \|u - z\|_p^p$, where $p \geq 1$ and $\|\cdot\|_p$ is ℓ^p -norm.
- $J_{regu}(u) = \|Ku\|_q^q$, where K is linear transform and $q \geq 1$.
- $J(u) = J_{data}(u) + \alpha J_{regu}(u)$, where $\alpha > 0$.
- Negative binary entropy ($\epsilon > 0$):
 $J_\epsilon(u) = \epsilon(u \log(u) + (1 - u) \log(1 - u))$.
- Soft plus: $J_\epsilon(v) = \epsilon \log(1 + \exp(v/\epsilon))$.
- **Indicator function** of a convex set $C \subset \mathbb{E}$:

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

Formulate *constrained optimization* with indicator function:

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





(As exercises)

- Any norm (over a normed vector space) is a convex function.
- J is a convex function and A is an affine transform
 $\Rightarrow u \mapsto J(A(u))$ is a 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 $\{u^i\}_{i=1}^n \subset \mathbb{E}$, $\{\alpha_i\}_{i=1}^n \subset [0, 1]$, $\sum_{i=1}^n \alpha_i = 1$.

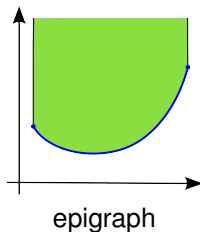
(Hence it is an equivalent definition of a convex function.)

Epigraph

Definition

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

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



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: found in script.

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

Existence of Minimizer

Global vs. Local minimizer

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 $J(u^*) < J(u)$.

Global vs. Local minimizer

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.

Does a minimizer always exist?

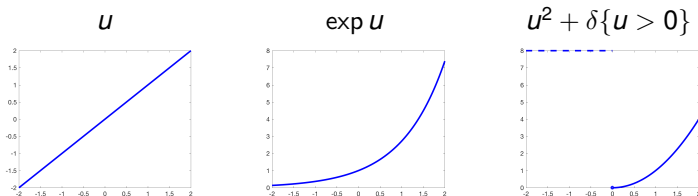


- Consider

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

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

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



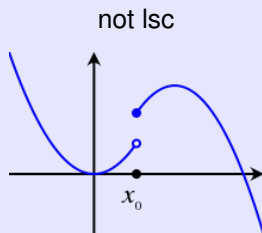
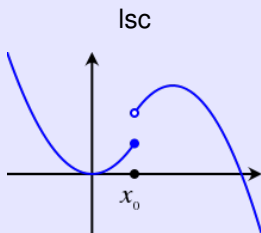
- Next we formalize our observations and derive sufficient conditions for existence.

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

$$J(u^*) \leq \liminf_{k \rightarrow \infty} J(u^k), \text{ whenever } u^k \rightarrow u^*.$$



- 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

$$J(u^*) \leq \liminf_{k \rightarrow \infty} J(u^k), \text{ whenever } u^k \rightarrow u^*.$$

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.

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



- 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.



Subdifferential

Convex Set

Convex Function

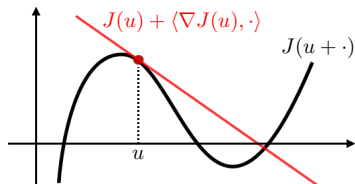
Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



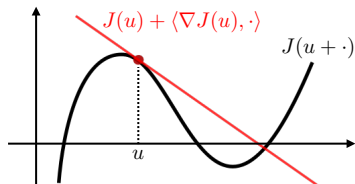
Definition

$J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ 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 $(\text{int dom } J) \cap B_\epsilon(u)$ for some $\epsilon > 0$.





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 $(\text{int dom } J) \cap B_\epsilon(u)$ for some $\epsilon > 0$.

Remark

For a convex function J , differentiability of J can be defined by considering $J : \text{aff dom } J \rightarrow \overline{\mathbb{R}}$ and $u \in \text{rint dom } J$.

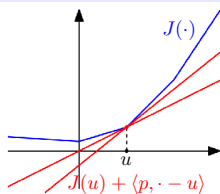


Subdifferential

Definition

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

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

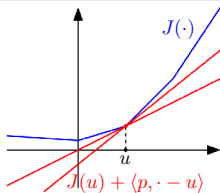


Subdifferential

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

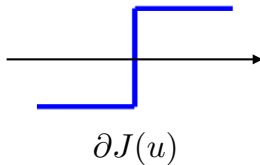
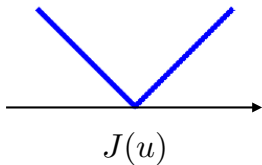
$p \in \partial J(u)$ iff $(p, -1)$ is a normal vector for the supporting hyperplane of $\text{epi } J$ at $(u, J(u))$, i.e.,

$$\left\langle \begin{bmatrix} p \\ -1 \end{bmatrix}, \begin{bmatrix} u \\ J(u) \end{bmatrix} \right\rangle \geq \left\langle \begin{bmatrix} p \\ -1 \end{bmatrix}, \begin{bmatrix} v \\ \alpha \end{bmatrix} \right\rangle, \quad \forall (v, \alpha) \in \text{epi } J.$$



Subdifferential: Examples

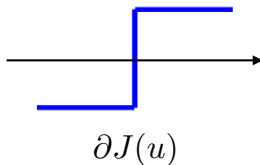
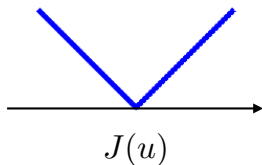
① $J(u) = |u|$.



Subdifferential: Examples



① $J(u) = |u|$.



② Given a 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

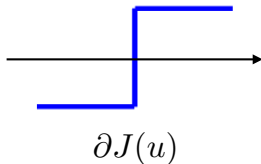
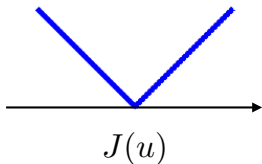
Duality Theory

Proximal Operator

Subdifferential: Examples



① $J(u) = |u|$.



② Given a 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 $J(u) = |||u||| \Rightarrow \partial J(0) = \{p : |||p|||_* \leq 1\}$. $||| \cdot |||_*$ is the dual norm of $||| \cdot |||$, i.e., $|||p|||_* = \sup\{\langle p, u \rangle : |||u||| \leq 1\}$.

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

Subdifferential calculus



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)

Let $\tilde{J}(\cdot) = J_1(\cdot) + J_2(\cdot)$, where J_1, J_2 are convex functions s.t.

$$\text{rint dom } J_1 \cap \text{rint dom } J_2 \neq \emptyset.$$

Then for any $u \in \text{dom } J_1 \cap \text{dom } J_2$, we have

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

Properties of subdifferential map

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 Analysis

Tao Wu
Zhenzhang Ye



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 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 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.



Properties of subdifferential map

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 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 proper, convex, lsc function. Then the set-valued map ∂J is **closed**, i.e. $p^* \in \partial J(u^*)$ whenever

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

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.



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

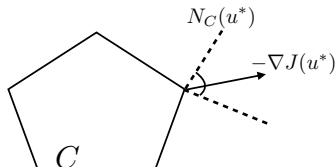
$$0 \in \partial J(u^*).$$

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 $0 \in \partial \tilde{J}(u^*) \Leftrightarrow$

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

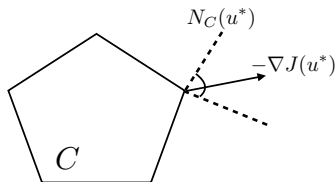


Optimality condition

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 relies on the *algebraic* representation of $N_C(u^*)$ (e.g., the **Karush-Kuhn-Tucker (KKT) conditions**) typically under certain *constraint qualifications*.



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 relies on the *algebraic* representation of $N_C(u^*)$ (e.g., the **Karush-Kuhn-Tucker (KKT) conditions**) typically under certain *constraint qualifications*.

Example: Normal cone of linear-inequality constraints

Let $C = \{u \in \mathbb{R}^n : Au \leq b\}$ with given $b \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n}$. Then

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





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

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



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 $J(u) = \| \|u\| \| \Rightarrow J^*(p) = \delta\{ \| \|p\| \|_* \leq 1 \}$. $\| \cdot \|_*$ is the dual norm of $\| \cdot \|$, i.e., $\| \|p\| \|_* = \sup\{ \langle p, u \rangle : \| \|u\| \| \leq 1 \}$.
- 2 $J(u) = \frac{1}{q} \| \|u\| \|_q^q \Rightarrow J^*(p) = \frac{1}{r} \| \|p\| \|_r^r$. ($1 < q < \infty$, $\frac{1}{q} + \frac{1}{r} = 1$)
- 3 $J(u) = \sum_{i=1}^n u_i \log u_i + \delta\{u \in \Delta^n\}$ (negative entropy)
 $\Rightarrow J^*(p) = \log(\sum_{i=1}^n \exp(p_i))$. (log-sum-exp)



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 $J(u) = \| \|u\| \| \Rightarrow J^*(p) = \delta\{ \| \|p\| \|_* \leq 1 \}$. $\| \cdot \|_*$ is the dual norm of $\| \cdot \|$, i.e., $\| \|p\| \|_* = \sup\{ \langle p, u \rangle : \| \|u\| \| \leq 1 \}$.
- 2 $J(u) = \frac{1}{q} \| \|u\| \|_q^q \Rightarrow J^*(p) = \frac{1}{r} \| \|p\| \|_r^r$. ($1 < q < \infty$, $\frac{1}{q} + \frac{1}{r} = 1$)
- 3 $J(u) = \sum_{i=1}^n u_i \log u_i + \delta\{ u \in \Delta^n \}$ (negative entropy)
 $\Rightarrow J^*(p) = \log(\sum_{i=1}^n \exp(p_i))$. (log-sum-exp)

Calculus rules

- 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

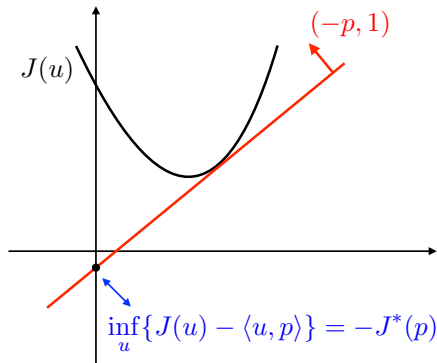
Geometric interpretation

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

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

If J is also convex, the equality holds iff $p \in \partial J(u)$ with $(u, p) \in \text{dom } J \times \text{dom } J^*$.

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: $\min_{u \in \mathbb{E}} J(u) - \langle u, p \rangle$.



Fenchel-Young inequality, order reversing property



Theorem (Fenchel-Young inequality)

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

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

If J is also convex, the equality holds iff $p \in \partial J(u)$ with $(u, p) \in \text{dom } J \times \text{dom } J^*$.

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 $J_1, J_2 : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, $J_1^*(\cdot) \leq J_2^*(\cdot)$ whenever $J_1(\cdot) \geq J_2(\cdot)$.

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 .



Regularity of J and J^*

Definition

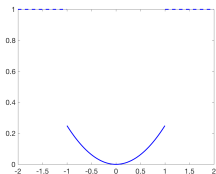
- 1 $J : \mathbb{E} \rightarrow \mathbb{R}$ is μ -strongly convex if $\exists \mu > 0$ s.t. $J(\cdot) - \frac{\mu}{2} \|\cdot\|^2$ is convex.
- 2 $J : \mathbb{E} \rightarrow \mathbb{R}$ is L -Lipschitz differentiable (a.k.a. L -smooth) if J is differentiable and ∇J is Lipschitz with modulus L .

Theorem

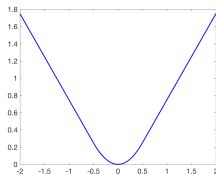
Assume that $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ is proper, convex, and lsc. Then J is μ -strongly convex iff J^* is $\frac{1}{\mu}$ -Lipschitz differentiable.

Proof: on board.

compactly supported quadratic



Huber function





Duality Theory

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Fenchel-Rockafellar duality

- Consider the model problem:

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

with $K \in \mathbb{R}^{m \times n}$, and $F : \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$, $G : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ proper, convex, and lsc functions.



Fenchel-Rockafellar duality

- Consider the model problem:

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

with $K \in \mathbb{R}^{m \times n}$, and $F: \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$, $G: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ proper, convex, and lsc functions.

- The **weak duality** ($\mathcal{P}^* \geq \mathcal{D}^*$) always holds:

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



Fenchel-Rockafellar duality

- Consider the model problem:

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

with $K \in \mathbb{R}^{m \times n}$, and $F: \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$, $G: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ proper, convex, and lsc functions.

- The **weak duality** ($\mathcal{P}^* \geq \mathcal{D}^*$) 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 (i) $\mathcal{G}(u, p) \geq 0 \quad \forall (u, p)$; (ii) $\mathcal{G}(u, p) = 0$ iff (u, p) is an optimal primal-dual pair (used as an optimality criterion).



Fenchel-Rockafellar duality



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

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



- $\mathcal{G}(u^*, p^*) = 0 \Leftrightarrow \mathcal{P}^* = \mathcal{D}^* \Leftrightarrow (u^*, p^*)$ 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 primal-dual pair iff

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

Proof: on board.

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Example: Total-variation image denoising

- 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 $\|p\|_{1,q} = \sum_{j \in \Omega} (\sum_{l=1}^d |p_{j,l}|^q)^{1/q}$ for each $p \in \mathbb{R}^{|\Omega| \times d}$.



Example: Total-variation image denoising

- 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 $\|p\|_{1,q} = \sum_{j \in \Omega} (\sum_{l=1}^d |p_{j,l}|^q)^{1/q}$ for each $p \in \mathbb{R}^{|\Omega| \times d}$.

- Apply the Fenchel-Rockafellar duality with

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



Example: Total-variation image denoising

- 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 $\|p\|_{1,q} = \sum_{j \in \Omega} (\sum_{l=1}^d |p_{j,l}|^q)^{1/q}$ 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 ($1/q + 1/q' = 1$):

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



Example: Total-variation image denoising

- 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 $\|p\|_{1,q} = \sum_{j \in \Omega} (\sum_{l=1}^d |p_{j,l}|^q)^{1/q}$ 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 ($1/q + 1/q' = 1$):

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

- 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

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

Observations

- 1 The minimization in prox always has a unique minimizer.
- 2 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 .



Definition

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

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

Observations

- 1 The minimization in prox always has a unique minimizer.
- 2 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 .

- 3 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

Examples of proximal operator

- 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 \Rightarrow projected gradient.



Examples of proximal operator

- 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 \Rightarrow projected gradient.

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

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



Examples of proximal operator

- 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 \Rightarrow 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.

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



Proximal gradient

Let us solve the convex optimization:

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

with G continuously differentiable and F non-differentiable.

The **proximal gradient** iteration appears as:

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

Derive proximal gradient as fixed-point iteration:

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



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Logistic regression (programming exercise)



- Classification of MNIST¹ handwritten digits.
- $C = 10$ classes; grayscale images of pixels $M = 28 \times 28$.
- $N = 60000$ training images $X \in \mathbb{R}^{N \times M}$ with ground-truth labels $Y \in \{1, \dots, C\}^N$; 10000 test samples for evaluation.
- Performance by error rate: Convolutional neural network $< 1\%$; **Logistic regression** $< 8\%$.



¹<http://yann.lecun.com/exdb/mnist/>

Logistic regression (programming exercise)



- Classification of MNIST¹ handwritten digits.
- $C = 10$ classes; grayscale images of pixels $M = 28 \times 28$.
- $N = 60000$ training images $X \in \mathbb{R}^{N \times M}$ with ground-truth labels $Y \in \{1, \dots, C\}^N$; 10000 test samples for evaluation.
- Performance by error rate: Convolutional neural network $< 1\%$; **Logistic regression** $< 8\%$.
- Logistic regression: Train a *linear classifier* $x \mapsto Wx + b$ with $W \in \mathbb{R}^{C \times M}$ and $b \in \mathbb{R}^C$, which parameterizes *likelihood* via *softmax function*:

$$\mathcal{P}(y|x; W, b) = \frac{\exp(\langle W_{y,\cdot}, x \rangle + b_y)}{\sum_{k=1}^C \exp(\langle W_{k,\cdot}, x \rangle + b_k)}.$$

¹<http://yann.lecun.com/exdb/mnist/>



Logistic regression (programming exercise)



- Classification of MNIST¹ handwritten digits.
- $C = 10$ classes; grayscale images of pixels $M = 28 \times 28$.
- $N = 60000$ training images $X \in \mathbb{R}^{N \times M}$ with ground-truth labels $Y \in \{1, \dots, C\}^N$; 10000 test samples for evaluation.
- Performance by error rate: Convolutional neural network $< 1\%$; **Logistic regression** $< 8\%$.
- Logistic regression: Train a *linear classifier* $x \mapsto Wx + b$ with $W \in \mathbb{R}^{C \times M}$ and $b \in \mathbb{R}^C$, which parameterizes *likelihood* via *softmax function*:

$$\mathcal{P}(y|x; W, b) = \frac{\exp(\langle W_{y,\cdot}, x \rangle + b_y)}{\sum_{k=1}^C \exp(\langle W_{k,\cdot}, x \rangle + b_k)}.$$

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

$$\min_{W, b} -\frac{1}{N} \sum_{n=1}^N \log \mathcal{P}(Y_n | X_{n,\cdot}; W, b) + \mathcal{R}_W(W) + \mathcal{R}_b(b).$$

¹<http://yann.lecun.com/exdb/mnist/>



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:

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



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:

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

Remark

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



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](#)

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.



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](#)

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



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

Properties of Moreau envelope

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

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



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.





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:

$$p = \nabla \text{env}_{\tau J}(u) \Leftrightarrow u \in \partial(\text{env}_{\tau J})^*(p) = \partial J^*(p) + \tau p$$

$$\Leftrightarrow u - \tau p \in \partial J^*(p) \Leftrightarrow \partial J(u - \tau p) \ni p$$

$$\Leftrightarrow \tau \partial J(u - \tau p) \ni \tau p \Leftrightarrow (I + \tau \partial J)(u - \tau p) \ni u$$

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

$$\Leftrightarrow \nabla \text{env}_{\tau J}(u) = p = \frac{1}{\tau}(u - \text{prox}_{\tau J}(u)).$$

Hence,

$$\text{prox}_{\tau J}(u) = u - \tau \nabla \text{env}_{\tau J}(u).$$