Chapter 1
Convex Analysis

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

Tao Wu
Zhenzhang Ye

Computer Vision Group
Department of Informatics
TU Munich

Last updated: 28.10.2019

Convex Set

Convex Optimization

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 Optimization

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 optimization

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

First questions:

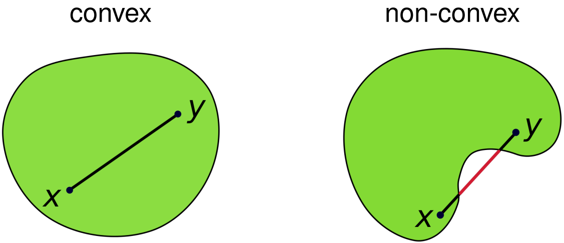
- What is a convex set?
- What is a convex function?

Convex set

Definition

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

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



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 $\mathbb{E} \backslash C$ is open.
- The **closure** of a set $C \subset \mathbb{E}$ is

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

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

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

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 $\mathbb{E} \backslash C$ is open.
- The **closure** of a set $C \subset \mathbb{E}$ is

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

- 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 aff C the **affine hull** of C. If C is a *convex* set, then

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

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: $\operatorname{cl} C$.
- Interior and relative interior: $\operatorname{int} C$, $\operatorname{rint} C$.

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

Basic properties

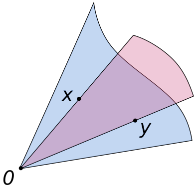
The following operations preserve the convexity:

- 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: $\operatorname{cl} C$.
- Interior and relative interior: $\operatorname{int} C$, $\operatorname{rint} C$.

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.

Separation of convex sets

Theorem (separation of convex sets)

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

1. Assume C_1 is closed and $C_2 = \{w\} \subset \mathbb{E} \backslash C_1$. 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 $C_2 = \{w\} \subset \mathbb{E} \backslash C_1$. 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.
$$\left\langle v, u^1 \right\rangle \geq \alpha \geq \left\langle v, u^2 \right\rangle , \quad \forall u^1 \in C_1,\ u^2 \in C_2 .$$
4. Assume $\emptyset \neq \operatorname{int} C_1 \subset \mathbb{E} \backslash C_2$. Then $\exists v \in \mathbb{E},\ v \neq 0,\ \alpha \in \mathbb{R}$ s.t.
$$\left\langle v, u^1 \right\rangle \geq \alpha \geq \left\langle v, u^2 \right\rangle , \quad \forall u^1 \in C_1,\ u^2 \in C_2 .$$

Convex Function

Convex functions

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

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

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

Definition

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

1. $\operatorname{dom} J$ is a convex set.
2. For all $u, v \in \operatorname{dom} J$ 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 ($\epsilon > 0$):
  $J_\epsilon(u) = \epsilon\big(u \log(u) + (1-u)\log(1-u)\big)$.
- 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 ($\epsilon > 0$):
  $J_\epsilon(u) = \epsilon\big(u \log(u) + (1-u)\log(1-u)\big)$.
- 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:

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

Basic facts

(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) $J : \mathbb{E} \to \overline{\mathbb{R}}$ 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 $\{u^i\}_{i=1}^n \subset \mathbb{E}$, $\{\alpha_i\}_{i=1}^n \subset [0,1]$, $\sum_{i=1}^n \alpha_i = 1$.

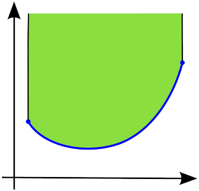
  (Hence it is an equivalent definition of a convex function.)

Epigraph

Definition

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

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



epigraph

Theorem

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

Proof: as exercise.

Lipschitz continuity

Definition

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

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

Lipschitz continuity

Definition

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

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

Theorem

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

Proof: found in script.

Existence of Minimizer

Global vs. Local minimizer

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

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

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 $J : \mathbb{E} \to \overline{\mathbb{R}}$:

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

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 $J : \mathbb{E} \to \overline{\mathbb{R}}$, if $u^* \in \operatorname{dom} J$ is a local minimizer of J, then it is also a global minimizer.

Proof: on board.

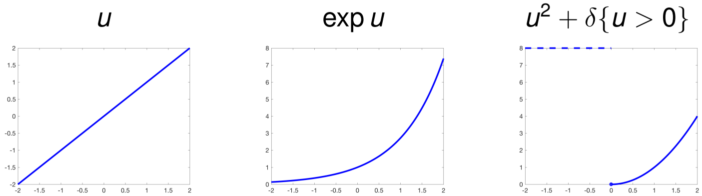
Does a minimizer always exist?

- Consider

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

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

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



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

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

Sufficient conditions for existence

Definition

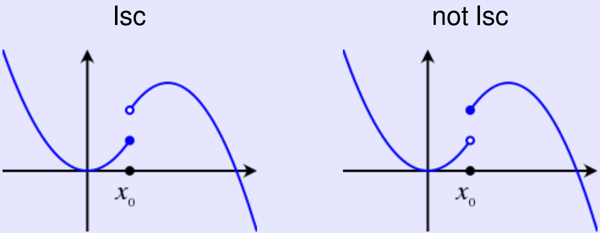
① J is **bounded from below** if $J(\cdot) \geq C$ for some $C \in \mathbb{R}$.

② J is **coercive** if $J(u) \to \infty$ whenever $\|u\| \to \infty$.

  - Proposition: J is coercive if $\operatorname{dom} J$ is bounded.

③ J is **lower semi-continuous** (lsc) at u^* if

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



lsc

not lsc

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

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 $J(u) \to \infty$ whenever $\|u\| \to \infty$.
3. J is **lower semi-continuous** (lsc) at u^* if

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

Theorem

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

Proof: on board.

Uniqueness

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

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

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

Theorem

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

Proof: on board.