Tao Wu
Thomas Möllenhoff
Emanuel Laude

Chapter 1
Convex Analysis

Convex Optimization for Machine Learning & Computer Vision
SS 2017

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Computer Vision Group
Department of Informatics
TU München

updated 10.05.2017

Convex Set

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Convex Optimization

Assume

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

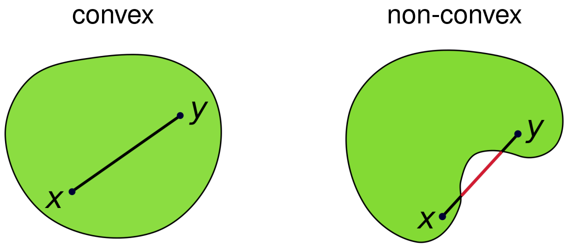
Tao Wu
Thomas Möllenhoff
Emanuel Laude

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

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

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

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

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

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

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

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

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

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

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: $\operatorname{int} C$

– The union of convex sets is not convex in general.

Tao Wu
Thomas Möllenhoff
Emanuel Laude

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: $\operatorname{int} C$

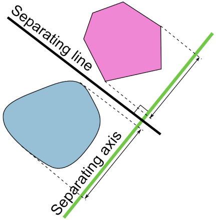
– The union of convex sets is not convex in general.

– *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.

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.

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

<u>Proof</u>: on board.

Tao Wu
Thomas Möllenhoff
Emanuel Laude

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.

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

Proof: on board.

Remarks

1. The proof works in any real Hilbert space.
2. Corollary: In a real Hilbert space, any (strongly) closed convex subset C is weakly closed.

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Convex Function

Tao Wu
Thomas Möllenhoff
Emanuel Laude

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

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

Examples

- $J_{data}(u) = \|u - f\|_q^q$ where $q \geq 1$ and $\|\cdot\|_q$ is the ℓ^q-norm.
- $J_{regu}(u) = \|Ku\|_q^q$ where K is linear transform.
- $J(u) = J_{data}(u) + \alpha J_{regu}(u)$ where $\alpha > 0$.
- **Indicator function**:

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

  where C is a convex subset of $\mathbb{E}$.

- Alternative formulation of constrained optimization:

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

Basic facts

(As exercises)

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

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

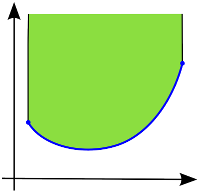
whenever $\{u^i\}_{i=1}^n \subset \mathbb{E}$, $\{\alpha_i\}_{i=1}^n \subset [0, 1]$, $\sum_{i=1}^n \alpha_i = 1$.

Epigraph

Definition

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

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



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 $\operatorname{ri}\operatorname{dom} J \neq \emptyset$. We say J is **locally Lipschitz** at $u \in \operatorname{ri}\operatorname{dom} J$ with modulus $L_u > 0$ if there exists $\epsilon > 0$ s.t.

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

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Lipschitz continuity

Definition

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

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

Theorem

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

Proof: on board.

Global vs. Local minimizers

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

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Global vs. Local minimizers

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

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.

Existence of Minimizer

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Does a minimizer always exist?

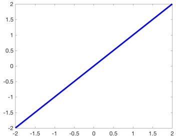
- Consider

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

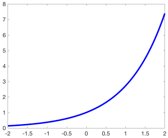
where $J : \mathbb{E} \to \overline{\mathbb{R}}$ is a proper, convex function.

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

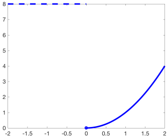
u



$\exp u$



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



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

Tao Wu
Thomas Möllenhoff
Emanuel Laude

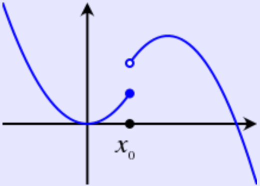
Sufficient conditions for existence

Definition

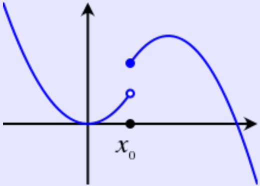
1. J is **bounded from below** if $J(\cdot) \geq C$ for some $C \in \mathbb{R}$.
2. J is **coercive** if $J(u) \to \infty$ whenever $\|u\| \to \infty$.
   - Proposition: J is coercive if dom J is bounded.
3. J is **lower-semicontinuous** (l.s.c.) if

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

l.s.c.



not l.s.c.



- Proposition: J is l.s.c. iff 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-semicontinuous** (l.s.c.) if

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

Theorem

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

Proof: on board.

Tao Wu
Thomas Möllenhoff
Emanuel Laude

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-semicontinuous** (l.s.c.) if

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

Theorem

Any proper function $J : \mathbb{E} \to \overline{\mathbb{R}}$, which is bounded from below, coercive, and l.s.c., 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 l.s.c. $\Rightarrow$ J is weakly l.s.c.

Tao Wu
Thomas Möllenhoff
Emanuel Laude

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.

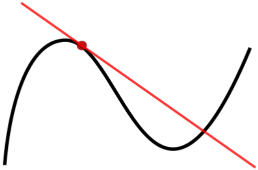
Proof: on board.

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Subdifferential

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Differential



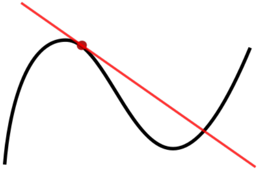
Definition

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

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

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

Differential



Definition

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

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

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

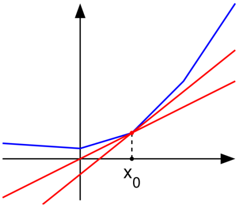
Remark

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

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Subdifferential

Now we generalize differentiability from differentiable functions to nonsmooth (convex) functions.



Definition

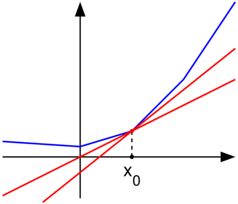
The **subdifferential** of a convex function $J : \mathbb{E} \to \overline{\mathbb{R}}$ at $u \in \operatorname{dom} J$ is defined by

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

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Subdifferential

Now we generalize differentiability from differentiable functions to nonsmooth (convex) functions.



Definition

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

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

Geometric interpretation

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

Examples and basic facts

Basic facts

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

Tao Wu
Thomas Möllenhoff
Emanuel Laude

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 (as exercises)

1. $u \in \mathbb{R}^n \mapsto \|u\|_1$.
2. $u \in \mathbb{R}^n \mapsto \|u\|_\infty$.
3. $X \in \mathbb{R}^{n \times n} \mapsto \|X\|_{1,2} = \sum_i \left(\sum_j |X_{i,j}|^2 \right)^{1/2}$.
4. $X \in \mathbb{R}^{n \times n} \mapsto \|X\|_{nuc} = \sum_i \sigma_i(X)$ (sum of singular values).
5. Given a closed convex subset $C \subset \mathbb{E}$ and $u \in C$,

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

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

Subdifferential calculus

Theorem (chain rule under linear transform)

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

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

for any $u \in \operatorname{ri}\operatorname{dom} J$.

Example: $J(u) = \|Ku\|_1 \Rightarrow \partial J(u) = K^\top \partial\|\cdot\|_1(Ku)$.

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Subdifferential calculus

Theorem (chain rule under linear transform)

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

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

for any $u \in \operatorname{ri} \operatorname{dom} J$.

Example: $J(u) = \|Ku\|_1 \Rightarrow \partial J(u) = K^\top \partial \|\cdot\|_1(Ku)$.

Theorem (summation rule)

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

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

for any $u \in \operatorname{ri} \operatorname{dom} J_1 \cap \operatorname{ri} \operatorname{dom} J_2$.

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

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Properties of subdifferential map

Theorem

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

$$\left\langle \xi^1 - \xi^2, u^1 - u^2 \right\rangle \geq 0.$$

Proof: on board.

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Properties of subdifferential map

Theorem

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

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

Proof: on board.

Theorem

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

Proof: on board.

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Properties of subdifferential map

Theorem

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

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

Proof: on board.

Theorem

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

Proof: on board.

Theorem

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

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

Proof: on board.

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Optimality condition

Theorem

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

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

Proof: on board.

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Optimality condition

Theorem

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

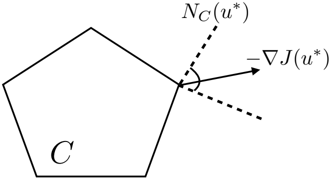
$$0 \in \partial J(u^*).$$

Proof: on board.

Constrained optimization as special case

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

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



Tao Wu
Thomas Möllenhoff
Emanuel Laude

Optimality condition

Theorem

Given any proper convex function $J : \mathbb{E} \to \overline{\mathbb{R}}$, 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 $\widetilde{J}(u) = J(u) + \delta_C(u)$ with convex function $J : \mathbb{E} \to \mathbb{R}$ and closed convex subset $C \in \mathbb{E}$, then $0 \in \partial J(u^*) \ \Leftrightarrow$

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

Remark

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

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Convex Conjugate

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Legendre-Fenchel transform

Definition

Given a convex function $J : \mathbb{E} \to \overline{\mathbb{R}}$, the **convex conjugate** of J is defined by

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

Tao Wu
Thomas Möllenhoff
Emanuel Laude

Legendre-Fenchel transform

Definition

Given a convex function $J : \mathbb{E} \to \overline{\mathbb{R}}$, the **convex conjugate** of J is defined by

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

Examples

1. $J(\cdot) = \langle w, \cdot \rangle \;\Rightarrow\; J^*(\cdot) = \delta\{\cdot = w\}$. $(w \in \mathbb{E})$
2. $J(\cdot) = \frac{1}{q}\| \cdot \|_q^q \;\Rightarrow\; J^*(\cdot) = \frac{1}{q'}\| \cdot \|_{q'}^{q'}$. $(0 < q < 1, \frac{1}{q} + \frac{1}{q'} = 1)$
3. $J(\cdot) = \| \cdot \| \;\Rightarrow\; J^*(\cdot) = \delta\{\| \cdot \|_* \leq 1\}$. ($\| \cdot \|_*$ is the *dual norm* of $\| \cdot \|$, i.e. $\|v\|_* = \sup_{\|u\| \leq 1} \langle u, v \rangle$)
4. $J(\cdot) = \delta\{\| \cdot \|_* \leq 1\} \;\Rightarrow\; J^*(\cdot) = \| \cdot \|$.

Legendre-Fenchel transform

Definition

Given a convex function $J : \mathbb{E} \to \overline{\mathbb{R}}$, the **convex conjugate** of J is defined by

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

Examples

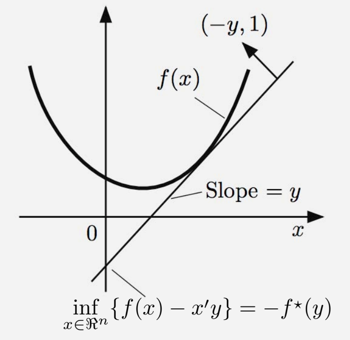
1. $J(\cdot) = \langle w, \cdot \rangle \;\Rightarrow\; J^*(\cdot) = \delta\{\cdot = w\}$. $(w \in \mathbb{E})$
2. $J(\cdot) = \frac{1}{q}\|\cdot\|_q^q \;\Rightarrow\; J^*(\cdot) = \frac{1}{q'}\|\cdot\|_{q'}^{q'}$. $(0 < q < 1, \frac{1}{q} + \frac{1}{q'} = 1)$
3. $J(\cdot) = \|\cdot\| \;\Rightarrow\; J^*(\cdot) = \delta\{\|\cdot\|_* \le 1\}$. ($\|\cdot\|_*$ is the *dual norm* of $\|\cdot\|$, i.e. $\|v\|_* = \sup_{\|u\| \le 1} \langle u, v \rangle$)
4. $J(\cdot) = \delta\{\|\cdot\|_* \le 1\} \;\Rightarrow\; J^*(\cdot) = \|\cdot\|$.

Basic facts

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

Geometric interpretation

Convex conjugation maps from:
  non-vertical hyperplanes supporting the epigraph
to:
  crossing points of vertical axis.



Courtesy of Bertsekas

Fenchel-Young inequality

Theorem (Fenchel-Young inequality)

For all $u \in \operatorname{dom} J,\ p \in \operatorname{dom} J^*$, we have

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

The equality holds iff $p \in \partial J(u)$.

Proof: $J(u) + J^*(p) \geq \langle u, p \rangle$ follows directly from the definition of convex conjugate; $p \in \partial J(u)$ is the sufficient and necessary condition for: $u = \arg\min_{v \in \mathbb{E}} J(v) - \langle v, p \rangle$.

Biconjugate

Theorem

Assume $J : \mathbb{E} \to \overline{\mathbb{R}}$ and $J^{**} = (J^*)^*$ is the **biconjugate** of J. In general:

1. $J^{**}(\cdot) \leq J(\cdot)$.
2. J^* is convex and l.s.c.

If J is proper, convex, and l.s.c., then:

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

Proof: on board.

Tao Wu
Thomas Möllenhoff
Emanuel Laude