

# 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

Convex Analysis

Tao Wu  
Thomas Möllenhoff  
Emanuel Laude



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



# Convex Set



## 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 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

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



## Convex Set

Convex Function

Existence of Minimizer

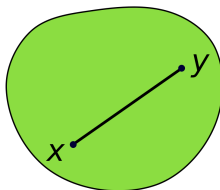
Subdifferential

Convex Conjugate

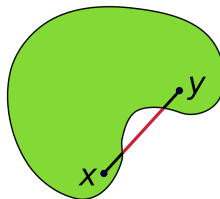
Duality Theory

Proximal Operator

convex



non-convex



## 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 convex set  $C \subset \mathbb{E}$  is

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



## Convex Set

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

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.





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.

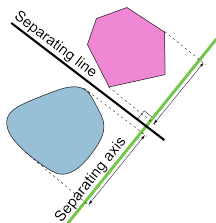
### Convex cone

$C$  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.

[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.





## 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 real Hilbert space.
- 2 Corollary: In a real Hilbert space, any (strongly) closed convex subset  $C$  is weakly closed.

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

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





- $J_{data}(u) = \|u - f\|_q^q$  where  $q \geq 1$  and  $\|\cdot\|_q$  is the  $\ell^q$ -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:

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



(As exercises)

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

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

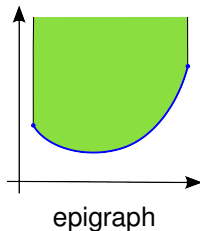
whenever  $\{u^i\}_{i=1}^n \subset \mathbb{E}$ ,  $\{\alpha_i\}_{i=1}^n \subset [0, 1]$ ,  $\sum_{i=1}^n \alpha_i = 1$ .

# Epigraph

## Definition

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

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



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

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

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



## Definition

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

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

## Theorem

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

Proof: on board.

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

## Global vs. Local minimizers

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



## Global vs. Local minimizers

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

### 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.





# Existence of Minimizer

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

# Does a minimizer always exist?

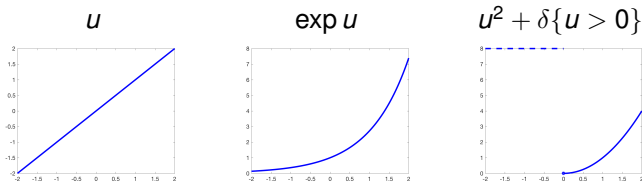


- Consider

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

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

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



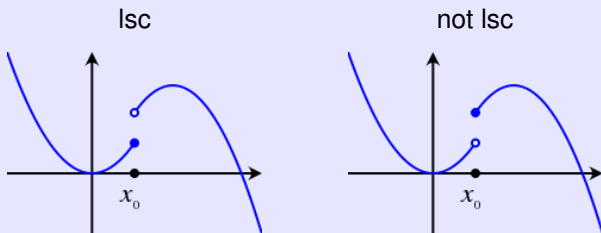
- We shall formalize our observations and derive sufficient conditions for existence.

# 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-semicontinuous** (lsc) if

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



- Proposition:  $J$  is lsc iff  $\text{epi } J$  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  $J(u) \rightarrow \infty$  whenever  $\|u\| \rightarrow \infty$ .
- 3  $J$  is **lower-semicontinuous** (lsc) if

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

## Theorem

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

Proof: on board.





## 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  $\|u\| \rightarrow \infty$ .
- 3  $J$  is **lower-semicontinuous** (lsc) if

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



### Theorem

Any proper function  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ , which is bounded from below, coercive, and lsc, 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  $u, v \in \text{dom } J$ ,  $u \neq v$ ,  $\alpha \in (0, 1)$ .

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

## 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

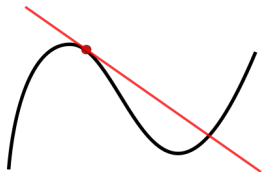
Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



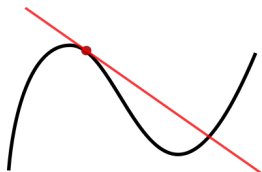
## Definition

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

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

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

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

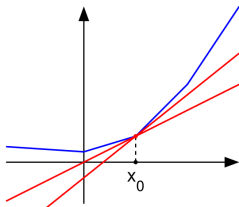
$J$  is said **continuously differentiable** at  $u \in \text{int dom } J$  if  $\nabla J(\cdot)$  is continuous on  $\text{dom } J \cap B_\epsilon(u)$  for some  $\epsilon > 0$ .

## 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*.

## Subdifferential

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



### Definition

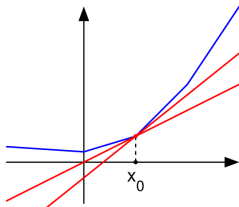
The **subdifferential** of a convex function  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$  at  $u \in \text{dom } J$  is defined by

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



## Subdifferential

Now we generalize differentiability from differentiable functions to nonsmooth (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

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

### Geometric interpretation

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



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 (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  $u$ .

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

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

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

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



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

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

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

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

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

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

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

## Properties of subdifferential map

### Theorem

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

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

Proof: on board.



## Properties of subdifferential map

### Theorem

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

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

Proof: on board.

### Theorem

Let  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$  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.



## Properties of subdifferential map

### Theorem

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

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

Proof: on board.

### Theorem

Let  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$  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.  $\xi^* \in \partial J(u^*)$  whenever

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

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.



## 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

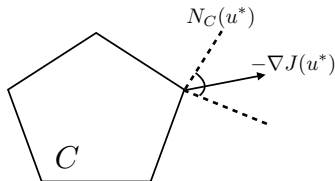
$$0 \in \partial J(u^*).$$

Proof: on board.

### Constrained optimization as special case

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

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





## 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  $\tilde{J}(u) = J(u) + \delta_C(u)$  with convex function  $J : \mathbb{E} \rightarrow \mathbb{R}$  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 relies on the **algebraic representation of  $N_C(u^*)$** , e.g. the **Karush-Kuhn-Tucker (KKT) conditions** under certain **constraint qualifications**.





# Convex Conjugate

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

# Legendre-Fenchel transform

## Definition

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

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

Convex Analysis

Tao Wu  
Thomas Möllenhoff  
Emanuel Laude



Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

# Legendre-Fenchel transform

## Definition

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

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

## Examples (as exercises)

- 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  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ , the **convex conjugate** of  $J$  is defined by

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

## Examples (as exercises)

- 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  $J(\cdot) = \|\cdot\| \Rightarrow J^*(\cdot) = \delta\{\|\cdot\|_* \leq 1\}. (\|\cdot\|_* \text{ is the dual norm of } \|\cdot\|, \text{ i.e. } \|v\|_* = \sup_{\|u\| \leq 1} \langle u, v \rangle)$
- 4  $J(\cdot) = \delta\{\|\cdot\|_* \leq 1\} \Rightarrow J^*(\cdot) = \|\cdot\|.$

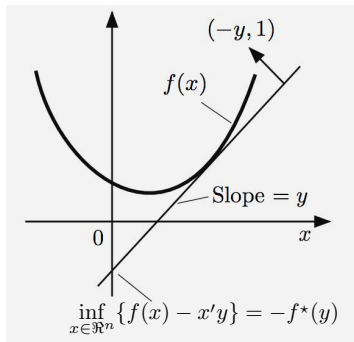
## Basic facts

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



## Geometric interpretation

Convex conjugation maps:  
(non-vertical) supporting hyperplane of the epigraph  
to:  
intersection with the vertical axis.



Courtesy of Bertsekas





## Theorem (Fenchel-Young inequality)

For all  $u \in \text{dom } J$ ,  $p \in \text{dom } J^*$ , 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:  $u = \arg \min_{v \in \mathbb{E}} J(v) - \langle v, p \rangle$ .

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator



## Theorem

Assume  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$  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

$J : \mathbb{E} \rightarrow \mathbb{R}$  is said to be  $\mu$ -strongly convex if  $\exists \mu > 0$  s.t.  
 $J(\cdot) - \frac{\mu}{2} \|\cdot\|^2$  is convex.

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

## Regularity of $J$ and $J^*$

### Definition

$J : \mathbb{E} \rightarrow \mathbb{R}$  is said to be  $\mu$ -strongly convex if  $\exists \mu > 0$  s.t.  $J(\cdot) - \frac{\mu}{2} \|\cdot\|^2$  is convex.

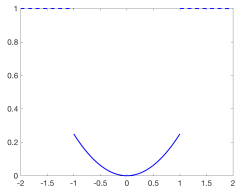
### Theorem

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

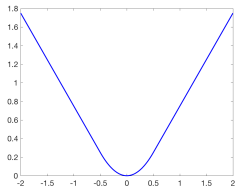
Proof: on board.



$J =$  truncated quadratic



$J^* =$  Huber function





# Duality Theory

Convex Set

Convex Function

Existence of Minimizer

Subdifferential

Convex Conjugate

Duality Theory

Proximal Operator

# Fenchel-Rockafellar duality

- Consider

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

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



# Fenchel-Rockafellar duality



- Consider

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

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

- The **weak duality** always holds:

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



- Consider

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

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

- The **weak duality** always holds:

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

- Define the **primal-dual 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.

# 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 solution 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



## Application: Total-variation image restoration

- Primal problem:

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

where  $\Omega \subset \mathbb{R}^d$ ,  $\|p\|_{1,q} = \sum_{j \in \Omega} |p_j|_q$  for each  $p \in (\mathbb{R}^\Omega)^d$ .



## Application: Total-variation image restoration

- Primal problem:

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

where  $\Omega \subset \mathbb{R}^d$ ,  $\|p\|_{1,q} = \sum_{j \in \Omega} |p_j|_q$  for each  $p \in (\mathbb{R}^\Omega)^d$ .

- Apply the Fenchel-Rockafellar duality with

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



## Application: Total-variation image restoration

- Primal problem:

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

where  $\Omega \subset \mathbb{R}^d$ ,  $\|p\|_{1,q} = \sum_{j \in \Omega} |p_j|_q$  for each  $p \in (\mathbb{R}^\Omega)^d$ .

- 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 ( $1/q + 1/q' = 1$ ):

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



## Application: Total-variation image restoration



- Primal problem:

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

where  $\Omega \subset \mathbb{R}^d$ ,  $\|p\|_{1,q} = \sum_{j \in \Omega} |p_j|_q$  for each  $p \in (\mathbb{R}^\Omega)^d$ .

- 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:

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