Subdifferentials

Pontus Giselsson

Today's lecture

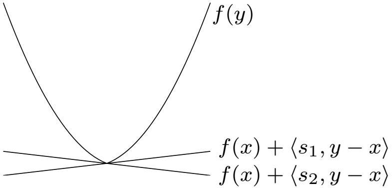
- subdifferentials and subgradients
- existence of subgradients
- relation between directional derivative and subdifferential
- Fermat's rule
- subdifferential calculus rules

Subdifferentials

- let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ (not necessarily convex)
- the subdifferential of f at x is the set of vectors s satisfying

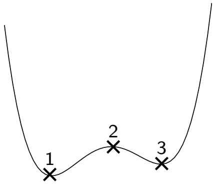
$$f(y) \geq f(x) + \langle s, y - x \rangle \quad \text{for all } y \in \mathbb{R}^n \tag{1}$$

- notation:
  - subdifferential: ∂f
  - subdifferential at x: $\partial f(x) = \{s \mid (1) \text{ holds}\}$
  - any element $s \in \partial f(x)$ is called *subgradient* of f at x
- subgradients define affine minorizers that coincide with f at x



Subdifferential example

- consider the following nonconvex function:



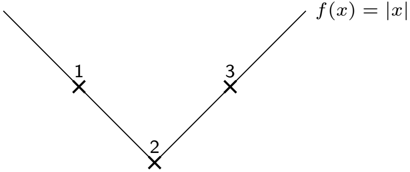
- what is the subdifferential at 1? 0
- what is the subdifferential at 2? $\emptyset$
- what is the subdifferential at 3? $\emptyset$

conclusion:

- subdifferential for nonconvex functions may be empty for some x

Subdifferential example

- consider the following convex function:



- what is the subdifferential at 1? -1
- what is the subdifferential at 2? $[-1, 1]$
- what is the subdifferential at 3? 1

fact:

- for finite-valued convex functions, a subgradient exists for every x

Extended-valued functions

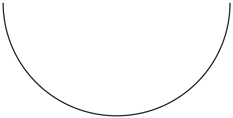
- let $f \ : \ \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex
- existence of subgradient if $x \notin \text{dom } f$?:
- subgradient definition:

$$f(y) \geq f(x) + \langle s, y - x \rangle \quad \text{for all } y \in \mathbb{R}^n$$

with $f(x) = \infty$, since l.h.s. finite for some y, $\partial f(x) = \emptyset$

Extended-valued functions

- let $f\ :\ \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex
- existence of subgradient for all $x \in \mathrm{dom} f$?
- counter-example: half-circle



$f(x) = -\sqrt{1-x^2}$
dom $f = [-1\ 1]$

- "vertical slope" at $x = 1$ (and $x = -1$)
- no affine function h with $h(1) = 0$ minorizes f

fact:

- for convex f subgradient exists for all $x \in$ ri dom$f$

Converse?

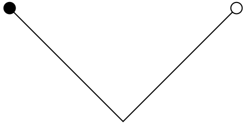
- know that subgradient exists for all $x \in$ ri dom f if f convex
- is f convex if subgradient exists for all $x \in$ ri dom f?might not be
- however: if we restrict ourselves to closed functions we have

a closed function is convex if and only if dom f is convex and dom $\partial f \supseteq$ ri dom f

- (construct counter-examples in exercises if conditions violated)

Proof sketch

- consider the function $f(x) = |x|$ with domain $[-1, 1)$



- construct a function g with domain $[-1, 1]$ that satisfies

$$g(x) = \begin{cases} f(x) & \text{if } x \in [-1, 1) \\ c & \text{else} \end{cases}$$

- if g must be convex, what does c have to satisfy?$c \geq 1$
- if g must be closed, what does c have to satisfy?$c \leq 1$
- if g closed and convex $\Rightarrow c = 1$
- (behavior on boundary controlled by behavior on ri dom$f$)

Subdifferentials and epigraphs

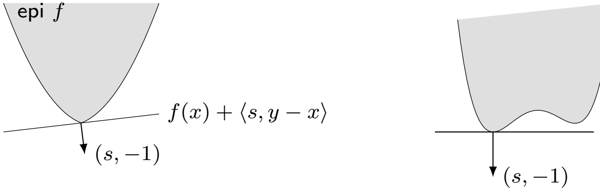
- it holds that:

$$s \in \partial f(x) \quad \text{if and only if} \quad (s, -1) \in N_{\text{epi}f}(x, f(x))$$

  or equivalently

$$N_{\text{epi}f}(x, f(x)) = \{(\lambda s, -\lambda) | \text{ for all } s \in \partial f(x), \lambda \geq 0\}$$

- subdifferentials define non-vertical supporting hyperplanes to $\text{epi}f$



- holds also for nonconvex and extended-valued f

Proof

- recall definition of normal cone operator to C: $s \in N_C(x)$ iff

$$\langle s, y - x\rangle \leq 0 \quad \text{for all } y \in C$$

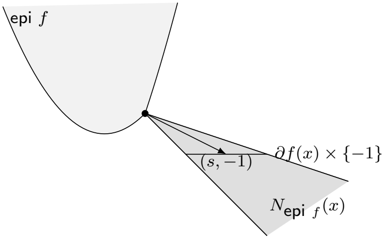
- apply to epi f: $(s, -1) \in N_{\text{epi } f}(x, f(x))$ iff

$$\begin{aligned}
& \langle (s, -1), (y, r) - (x, f(x))\rangle \leq 0 \quad \text{for all } y \in \mathbb{R}^n, r \geq f(y) \\
\iff \quad & r \geq f(x) + \langle s, y - x\rangle \quad \text{for all } y \in \mathbb{R}^n \text{ and } r \geq f(y) \\
\iff \quad & f(y) \geq f(x) + \langle s, y - x\rangle \quad \text{for all } y \in \mathbb{R}^n
\end{aligned}$$

which is the subgradient definition

Example

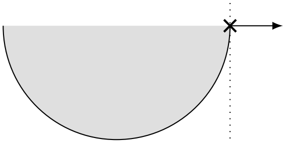
- consider the function $f(x) = \frac{1}{2}x^2 + |x - 2|$



- the normal vector $(s, -1)$ is in normal cone
- $(N_{\text{epi } f}(x, f(x)) = \mathbb{R}_+(\partial f(x) \times \{-1\}))$

Counter-example?

- counter-example?: half-circle



$f(x) = -\sqrt{1-x^2}$
dom $f = [-1\ 1]$

- normal cone at $(1,0)$ is

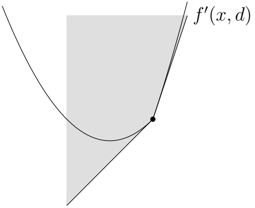
$$N_{\mathrm{epi}f}(1,0) = \{s \mid s = (s_1, s_2) \text{ with } s_1 \geq 0, s_2 = 0\}$$

- already now that $\partial f(1) = \emptyset$, is this a counter-example?
- No: no element of $N_{\mathrm{epi}f}(1,0)$ cannot be written as $(s,-1)$
  $(s,-1)$ models *non-vertical* supporting hyperplanes!

Tangent cone to epigraph

- here we assume that f is finite-valued and convex!
- tangent cone of epigraph is epigraph of directional derivative

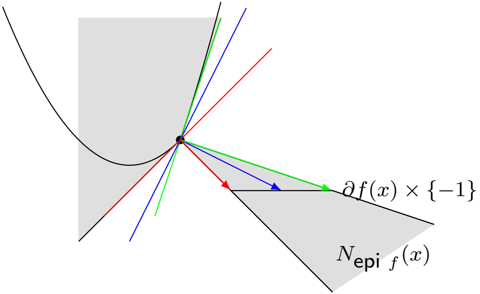
$$T_{\text{epi } f}(x, f(x)) = \text{epi } f'(x, d)$$



Alternative representation of tangent cone

- tangent cone is intersection of halfspaces defined by subgradients

$$T_{\text{epi } f}(x, f(x)) = \{(d, r) \mid \langle (s, -1), (d, r) \rangle \leq 0 \text{ for all } s \in \partial f(x)\}$$



Relation: Directional derivative and subdifferential

- from previous slide:

$$\begin{aligned} T_{\text{epi } f}(x, f(x)) &= \{(d, r) \mid \langle (s, -1), (d, r) \rangle \leq 0 \text{ for all } s \in \partial f(x)\} \\ &= \{(d, r) \mid \langle s, d \rangle \leq r \text{ for all } s \in \partial f(x)\} \\ &= \{(d, r) \mid \sup_{s \in \partial f(x)} \langle s, d \rangle \leq r\} = \text{epi } f'(x, d) \end{aligned}$$

- therefore the directional derivative satisfies

$$f'(x, d) = \sup_{s \in \partial f(x)} \langle s, d \rangle$$

  i.e., it is the support function of the subdifferential

- (for finite-valued convex f: subdifferential can be defined as set whose support function is the directional derivative)

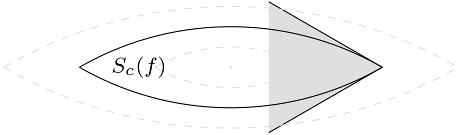
Tangent cone to levelsets

- let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and define levelset

$$S_c(f) = \{y \mid f(y) \le c\}$$

- assume that $\exists \bar{d}$ with $f'(x, \bar{d}) < 0$ and that $f(x) = c$, then

$$T_{S_c(f)}(x) = \{d \in \mathbb{R}^n \mid f'(x, d) \le 0\}$$



- tangent cone is directions with non-increasing function values
- (since $f(x) = c$ we look at elements on boundary of levelset)

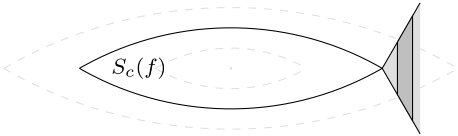
Normal cone to levelsets

- let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and define levelset

$$S_c(f) = \{y \mid f(y) \leq c\}$$

- assume that $\exists \bar{x}$ with $f(\bar{x}) < c$ and that $f(x) = c$, then

$$N_{S_c(f)}(x) = \mathbb{R}_+ \partial f(x)$$



- proven by showing that $(T_{S_c(f)}(x)) = (\mathbb{R}_+ \partial f(x))^\circ$

Are assumptions necessary?

- are the assumptions

$$\exists \bar{x} \text{ with } f(\bar{x}) < f(c) \quad \text{and} \quad \exists \bar{d} \text{ with } f'(x, \bar{d}) < 0$$

  necessary for the set equalities

$$N_{S_c(f)}(x) = \mathbb{R}_+ \partial f(x) \quad \text{and} \quad T_{S_c(f)}(x) = \{d \in \mathbb{R}^n \mid f'(x, d) \leq 0\}?$$

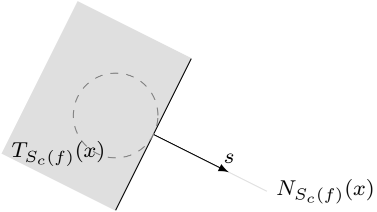
- consider $f = \frac{1}{2}\|\cdot\|^2$ and the levelset

$$S_0(f) = \{x \mid f(x) \leq 0\} = \{x \mid \tfrac{1}{2}\|x\|^2 \leq 0\} = \{0\}$$

- what is normal cone of $S_0(f)$ at $x = 0$?: $\mathbb{R}^n$
- what is the subdifferential at $x = 0$?: $\partial f(0) = \{0\}$
- what is the tangent cone of $S_0(f)$ at $x = 0$?: polar normal, i.e., $\{0\}$
- what is the set of d with nonpositive directional derivative at $x = 0$?: $\mathbb{R}^n$

Example

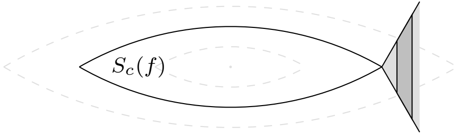
- f is finite, $\inf_x f(x) < f(x) = c$, $\partial f(x) = \{\nabla f(x)\}$
- compute the normal cone and tangent cone operator of $S_c(f)$
- normal cone:
  $N_{S_c(f)}(x) = \mathbb{R}_+\{\nabla f(x)\} = \{s \mid s = \lambda \nabla f(x), \lambda \geq 0\}$
- tangent cone: polar to $N_{S_c(f)}(x)$ is $T_{S_c(f)}(x) = \{d \mid \langle v, d\rangle \leq 0\}$



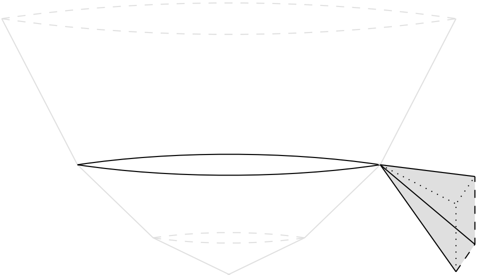
- (dashed curve is potential level curve)
- (gradient points "outwards" from level curve)

Relation between normal cones

- normal cone to level set of f, i.e., $S_c(f)$:



- normal cone to epi f, i.e., $N_{\text{epi } f}$:



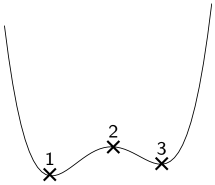
(note dim epi $f =$ dim $S_c(f) + 1$)

Relation to gradient

- if f differentiable at x and $\partial f(x) \neq \emptyset$ then $\partial f(x) = \{\nabla f(x)\}$

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \quad \text{for all } y \in \mathbb{R}^n$$

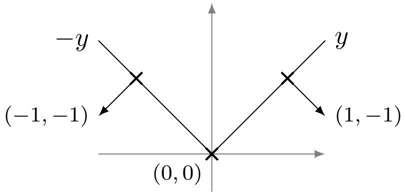
- if f differentiable and convex, then $\partial f(x) = \{\nabla f(x)\}$ for all x
- a function can be differentiable at x but $\partial f(x) = \emptyset$, e.g., "2","3":



- gradient is a local concept, subdifferential is a global property
- however, for convex functions gradient gives global under-estimator (since $\partial f(x) = \{\nabla f(x)\}$)

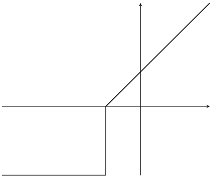
Example – Subdifferentials as lower bounds

- f convex, $f(-1) = 1$, $\partial f(-1) = \{-1\}$, $f(1) = 1$ and $\partial f(1) = \{1\}$
- compute a lower bound to the optimal value of f
- we know that optimal value of f is ≥ 0

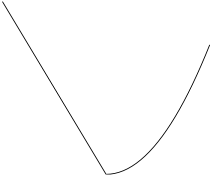


Construct function from subdifferential

- we have the following subdifferential



- draw the corresponding function and find the optimal point



(linear to the left and quadratic to the right)
(no axes since any constant can be added)

Fermat's rule

- Let f be proper, then x minimizes f if and only if

$$0 \in \partial f(x)$$

- proof: x minimizes f iff

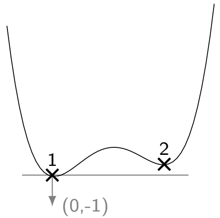
$$f(y) \geq f(x) + \langle 0, y - x \rangle \quad \text{for all } y \in \mathbb{R}^n$$

  which by definition of subdifferential is equivalent to $0 \in \partial f(x)$

- holds also for nonconvex functions

Example of Fermat's rule

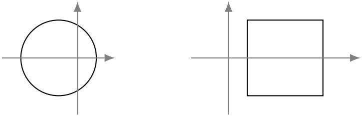
- Fermat's rule holds also for nonconvex functions:



- $\partial f(\text{"}1\text{"}) = 0$
- $\partial f(\text{"}2\text{"}) = \emptyset$

Examples of Fermat's rule

- (a): $\partial f(x)$, (b): $\partial g(y)$, does x resp. y optimize f resp. g?



(a) (b)

- if convex, can we conclude existence of optimal point in (b)? No!
- draw an example of a corresponding function

Subdifferential calculus rules

- how to compute $\partial(f_1 + f_2)(x)$?
- how to compute $\partial(g \circ L)(x)$?
- how to compute $\partial(Lg)(x)$?

Subdifferential sum

- if $x \in \textbf{dom}\partial f_1 \cap \textbf{dom}\partial f_2$, we have

$$\partial(f_1 + f_2)(x) \supseteq \partial f_1(x) + \partial f_2(x)$$

- proof:
  let $s_1 \in \partial f_1(x)$ and $s_2 \in \partial f_2(x)$, add subdifferential definition:

$$f_1(y) + f_2(y) \geq f_1(x) + f_2(x) + \langle s_1 + s_2, y - x \rangle$$

  i.e. $s_1 + s_2 \in \partial(f_1 + f_2)(x)$

- under additional assumptions, we also have reverse inclusion, i.e.,

$$\partial(f_1 + f_2) \subseteq \partial f_1 + \partial f_2$$

  (will be shown after conjugate functions introduced)

Composition

- let L be a linear operator
- if $Lx \in \text{dom} g$ we have

$$\partial(g \circ L)(x) \supseteq L^* \partial g(Lx)$$

- under additional assumptions also other inclusion holds (will be shown after conjugate functions)
- if f differentiable, we have chain rule

$$\nabla(g \circ L)(x) = L^* \nabla g(Lx)$$

Image function

- assume that $x \in \text{dom}(Lg) = L(\text{dom}g)$ and suppose that there exists $\bar{y}$ such that $L\bar{y} = x$ and $g(\bar{y}) = (Lg)(x)$, then

$$\partial(Lg)(x) = \{s \in \mathbb{R}^n \mid L^* s \in \partial g(\bar{y})\}$$

- will be shown after conjugate functions