Subdifferential Properties

Pontus Giselsson

Today's lecture

- last lecture: properties of monotone operators
- today: properties of subdifferential operators

Monotonicity

- we know that the subdifferential operator is monotone, i.e.,

$$\langle s_1 - s_2, x_1 - x_2 \rangle \geq 0$$

  for all $(s_1, x_1), (s_2, x_2) \in$ gph ∂f

- proof: add subdifferential definitions

$$\begin{aligned} f(x_2) &\geq f(x_1) + \langle s_1, x_2 - x_1 \rangle \\ f(x_1) &\geq f(x_2) + \langle s_2, x_1 - x_2 \rangle \end{aligned}$$

- (note: does not require convexity of f)

Maximal monotonicity

- ∂f is maximally monotone if no $(\bar{x}, \bar{s}) \notin \text{gph } \partial f$ exists such that

$$\langle \bar{s} - s, \bar{x} - x \rangle \geq 0$$

  for all $(x, s) \in \text{gph } \partial f$
- ∂f of a (proper) closed convex function f is maximally monotone

Strong monotonicity

- recall: the subdifferential operator is σ-strongly monotone iff

$$\langle s_1 - s_2, x_1 - x_2 \rangle \geq \sigma \|x_1 - x_2\|^2$$

holds for all $(s_1, x_1), (s_2, x_2) \in \text{gph } \partial f$

Strong monotonicity characterization

$$\partial f \text{ is } \sigma\text{-strongly monotone} \iff \partial f - \sigma \mathrm{Id} \text{ is monotone}$$

- proof: σ-strong monotonicity:

$$\begin{aligned} & \langle u - v, x - y \rangle \geq \sigma \|x - y\|^2 \\ \Leftrightarrow \quad & \langle (u - \sigma x) - (v - \sigma y), x - y \rangle \geq 0 \end{aligned}$$

  the result holds since $u - \sigma x \in \partial f(x) - \sigma x$

Strong monotonicity and strong convexity

- assume $f\ :\ \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper closed and convex:

$$\boxed{\begin{array}{c} f \text{ is } \sigma\text{-strongly convex} \\ \Longrightarrow \\ \partial f \text{ is } \sigma\text{-strongly monotone} \end{array}}$$

- first statement is equivalent to that $g = f - \frac{\sigma}{2}\|\cdot\|^2$ is convex
- further $f = g + \frac{\sigma}{2}\|\cdot\|^2$ and $\partial f = \partial(g + \frac{\sigma}{2}\|\cdot\|^2) = \partial g + \sigma\mathrm{Id}$
- therefore $\partial g = \partial f - \sigma\mathrm{Id}$
- definition of ∂g with $s_x \in \partial f(x)$

$$\begin{aligned} & f(y) - \tfrac{\sigma}{2}\|y\|^2 \geq f(x) - \tfrac{\sigma}{2}\|x\|^2 + \langle s_x - \sigma x, y - x\rangle \\ \Leftrightarrow \quad & f(y) \geq f(x) + \langle s_x, y - x\rangle + \tfrac{\sigma}{2}\|x - y\|^2 \end{aligned}$$

- add definitions with x and y swapped $\Rightarrow$ σ-strong monotonicity
- (actually $\Leftarrow$ also holds, will show this later)

More properties

- assume again that f is proper closed and convex, then
- what we actually showed on previous slide was

$$
\boxed{\begin{array}{c}
f \text{ is } \sigma\text{-strongly convex} \\
\Longrightarrow \\
f(y) \geq f(x) + \langle s, y - x\rangle + \frac{\sigma}{2}\|x - y\|^2 \\
\text{for all } x \in \text{dom}\partial f, y \in \mathbb{R}^n, s \in \partial f(x) \\
\Longrightarrow \\
\partial f \text{ is } \sigma\text{-strongly monotone}
\end{array}}
$$

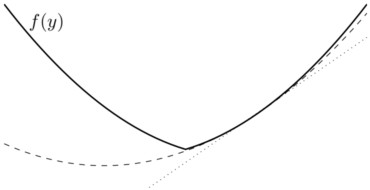
- (since third $\Rightarrow$ first (shown later) we have equivalence)

Graphical interpretation - strong convexity

- strong convexity:

$$f(y) \geq f(x) + \langle s, y - x \rangle + \frac{\sigma}{2}\|x - y\|^2$$

  i.e., has affine plus norm squared minorizer



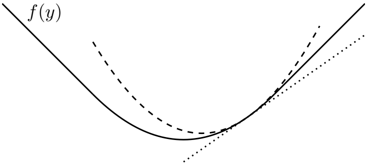
- compare to standard convex function with affine minorizer

Norm squared majorizers

- what happens if we instead have a norm-square majorizer

$$f(y) \leq f(x) + \langle s, y - x \rangle + \frac{\beta}{2}\|x - y\|^2$$

  to a convex function?



- f is squeezed between the affine minorizer and quadratic majorizer

Smoothness

- we define this property as smoothness
- assume that $f : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable
- then f is β-smooth if

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\beta}{2} \|x - y\|^2$$

  holds for all $x, y \in \mathbb{R}^n$
- unlike strong convexity, smoothness $\not\Rightarrow$ convexity $\Rightarrow$ assume!

Alternative definition of smoothness

- assume $f : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable
- then f is β-smooth if $g := \frac{\beta}{2}\|\cdot\|^2 - f$ is convex

equivalent to other definition?

- since g is differentiable, convexity of g is equivalent to that

  $$g(y) \geq g(x) + \langle \nabla g(x), y - x \rangle$$

  holds for all $x, y \in \mathbb{R}^n$
- insert $g = \frac{\beta}{2}\|\cdot\|^2 - f$ and $\nabla g = \beta \mathrm{Id} - \nabla f$, to get

  $$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \tfrac{\beta}{2}(-\|x\|^2 + \|y\|^2 - 2\langle x, y - x \rangle)$$

  which is equivalent to β-smoothness since

  $$-\|x\|^2 + \|y\|^2 - 2\langle x, y - x \rangle = \|x\|^2 + \|y\|^2 - 2\langle x, y \rangle = \|x - y\|^2$$

Cocoercivity

- gradients of β-smooth convex functions are $\frac{1}{\beta}$-cocoercive
- let $\phi(y) = f(y) - \langle \nabla f(x_0), y \rangle$ with $\nabla \phi(y) = \nabla f(y) - \nabla f(x_0)$
- ϕ is convex and is optimized at $y^\star = x_0$ since Fermat's rule says:

$$0 = \nabla f(y^\star) - \nabla f(x_0)$$

- further, ϕ is β-smooth: let $h = \frac{\beta}{2}\|\cdot\|^2 - f$ (h is convex)

$$\frac{\beta}{2}\|y\|^2 - \phi(y) = \frac{\beta}{2}\|y\|^2 - f(y) + \langle \nabla f(x_0), y \rangle = h + \langle \nabla f(x_0), y \rangle$$

which is convex, i.e., ϕ is β-smooth

Cocercivity cont'd

- next, use that
  - x_0 optimizes ϕ
  - ϕ is β-smooth
  - $\nabla\phi(y) = \nabla f(y) - \nabla f(x)$
- let $s_y = \nabla f(y)$ and $s_x = \nabla f(x_0)$, then

$$\begin{aligned}\phi(x_0) &\leq \phi(y - \tfrac{1}{\beta}(s_y - s_x)) \\ &\leq \phi(y) + \langle s_y - s_x, y - \tfrac{1}{\beta}(s_y - s_x) - y\rangle + \tfrac{\beta}{2}\|y - \tfrac{1}{\beta}(s_y - s_x) - y\|^2 \\ &= \phi(y) - \tfrac{1}{2\beta}\|s_y - s_x\|^2\end{aligned}$$

- or, using $\phi(y) = f(y) - \langle\nabla f(x_0), y\rangle$ and letting $x = x_0$:

$$f(x) \leq f(y) + \langle s_x, x - y\rangle - \tfrac{1}{2\beta}\|s_x - s_y\|^2$$

- add with arguments interchanged $\Rightarrow$ $\frac{1}{\beta}$-cocoercivity:

$$\langle s_x - s_y, x - y\rangle \geq \tfrac{1}{\beta}\|s_x - s_y\|^2$$

Differentiability

- we assumed differentiability of f before smoothness definition
- convexity and smoothness imply cocoercivity without differentiability assumption
- since cocoercivity $\Rightarrow$ single-valuedness of subdifferential $\Rightarrow$ differentiability, the differentiability assumption is not needed

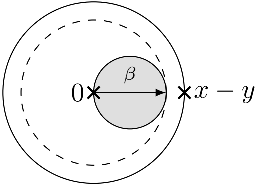
Lipschitz continuity

Assume $f : \mathbb{R}^n \to \mathbb{R}$ is convex

- we know that β-smooth functions have $\frac{1}{\beta}$-cocoercive gradients
- also $\frac{1}{\beta}$-cocoercivity implies β-Lipschitz continuity
- proof: by Cauchy-Schwarz, we have

$$\begin{aligned}\|\nabla f(x) - \nabla f(y)\|\|x - y\| &\geq \langle \nabla f(x) - \nabla f(y), x - y\rangle \\ &\geq \tfrac{1}{\beta}\|\nabla f(x) - \nabla f(y)\|^2\end{aligned}$$

  divide by $\|\nabla f(x) - \nabla f(y)\|$ and multiply by β to get result
- graphical representation:



Lipschitz implies cocoercivity

Assume $f : \mathbb{R}^n \to \mathbb{R}$ is convex

- actually, the reverse also holds!
- i.e., β-Lipschitz continuity of ∇f implies $\frac{1}{\beta}$-cocoercivity
- to show this, let $h(\tau) = f(x + \tau(y - x))$, then

$$\nabla h(\tau) = \langle y - x, \nabla f(x + \tau(y - x))\rangle$$

- since $f(y) = h(1)$ and $f(x) = h(0)$, we get

$$\begin{aligned} f(y) - f(x) = h(1) - h(0) &= \int_0^1 \nabla h(\tau) d\tau \\ &= \int_0^1 \langle y - x, \nabla f(x + \tau(y - x))\rangle d\tau \end{aligned}$$

Lipschitz implies cocoercivity cont'd

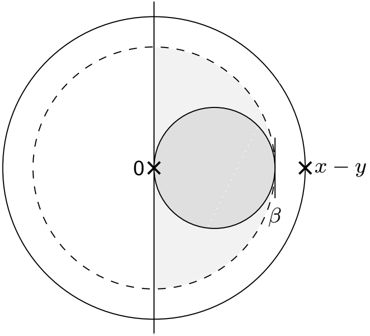
- using the previous characterization of $f(y) - f(x)$, we get

$$
\begin{aligned}
f(y) - f(x) - \langle \nabla f(x), y - x \rangle \\
&= \int_0^1 \langle \nabla f(x + \tau(y-x)), y - x \rangle d\tau - \langle \nabla f(x), y - x \rangle \\
&= \int_0^1 \langle \nabla f(x + \tau(y-x)) - \nabla f(x), y - x \rangle d\tau \\
&\leq \int_0^1 \|\nabla f(x + \tau(y-x)) - \nabla f(x)\| \|x - y\| d\tau \\
&\leq \beta \|x - y\| \int_0^1 \|x + \tau(y-x) - x\| d\tau \\
&= \beta \|x - y\| \int_0^1 \tau \|y - x\| d\tau = \tfrac{\beta}{2} \|x - y\|^2
\end{aligned}
$$

- i.e., f is β-smooth which implies $\frac{1}{\beta}$-cocoercivity of ∇f

Graphical interpretation

Assume f convex $\Rightarrow \nabla f$ monotone



- result: ∇f is $\frac{1}{\beta}$-cocoercive if and only if ∇f is β-Lipschitz
- Lipschitz ∇f cannot end up in lighter gray area!
- (does not hold for general Lipschitz operators T)

Summary: Smooth functions

Let f be convex, then the following are equivalent

- f is β-smooth
- ∇f is $\frac{1}{\beta}$-cocoercive
- ∇f is β-Lipschitz continuous

Duality correspondence

- we know that for proper closed convex f

$$s \in \partial f(x) \quad \Leftrightarrow \quad x \in \partial f^*(s)$$

  (since $f = f^{**}$ in that case)
- that is $\partial f^* = (\partial f)^{-1}$
- therefore, for proper closed convex f:

σ-cocoercivity of ∇f
$\Longleftrightarrow$
σ-strong monotonicity of $(\partial f)^{-1} = \partial f^*$

- (recall that cocoercivity is also called inverse strong monotonicity)

Duality correspondence

- summarizing results from lecture:

> Let f be proper closed and convex, and consider:
>
> (i) f^* is σ-strongly convex
> (ii) ∂f^* is σ-strongly monotone
> (iii) ∇f is σ-cocoercive
> (iv) ∇f is $\frac{1}{\sigma}$-Lipschitz continuous
> (v) f is $\frac{1}{\sigma}$-smooth

- all these conditions are equivalent
- proof: we have shown (i)$\Rightarrow$(ii)$\Leftrightarrow$(iii)$\Leftrightarrow$(iv)$\Leftrightarrow$(v)
- we know from before that (v)$\Leftrightarrow$(i), i.e., equivalences everywhere

Duality correspondence 2

- for proper closed and convex f, we have $f = f^{**}$
- therefore duality correspondence holds with f and f^* swapped:

> Let f be proper closed and convex, and consider:
>
> (i) f is σ-strongly convex
> (ii) ∂f is σ-strongly monotone
> (iii) ∇f^* is σ-cocoercive
> (iv) ∇f^* is $\frac{1}{\sigma}$-Lipschitz continuous
> (v) f^* is $\frac{1}{\sigma}$-smooth

Composition

- consider a problem of the form

$$\text{minimize } f(x) + g(Lx)$$

- the dual is then

$$\text{minimize } f^*(-L^*\mu) + g^*(\mu)$$

- properties of $d = f^* \circ -L^*$?

Composition cont'd

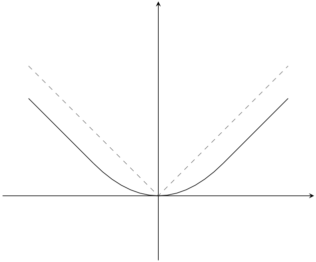
- assume that f is σ-strongly convex
- then f^* is $\frac{1}{\sigma}$-smooth and $(f \circ -L^*)$ is $\frac{\|L^*\|^2}{\sigma}$-smooth
- assume that f is β-smooth and $\|L^*\mu\| \geq \theta\|\mu\|$ for all μ ($\mathcal{N}(L^*) = 0$ or $\mathcal{R}(L) = \mathbb{R}^m$)
- then f^* is $\frac{1}{\beta}$-strongly convex and $(f \circ -L^*)$ is $\frac{\theta^2}{\beta}$-strongly convex

Second order differentiability

- assume that f is σ-strongly convex
- then f^* is $\frac{1}{\sigma}$-smooth
- and ∇f^* is Lipschitz continuous (even cocoercive)
- what about $\nabla^2 f^*$?
- ∇f Lipschitz $\Rightarrow$ $\nabla^2 f$ differentiable almost everywhere
- that is, a unique Hessian $\nabla^2 f^*$ exists almost everywhere

Example

- consider $f(y) = \inf_x \{\|x\|_1 + \|x-y\|_2^2\}$
- epi$f$ is set sum of $\|\cdot\|_1$ and $\|\cdot\|_2^2$



- that is

$$f(x) = \begin{cases} -x - 0.25 & \text{if } x \leq -0.5 \\ x^2 & \text{if } -0.5 \leq x \leq 0.5 \\ x - 0.25 & \text{if } x \geq 0.5 \end{cases}$$

Example cont'd

- we have

$$f(x) = \begin{cases} -x - 0.25 & \text{if } x \le -0.5 \\ x^2 & \text{if } -0.5 \le x \le 0.5 \\ x - 0.25 & \text{if } x \ge 0.5 \end{cases}$$

- therefore

$$\nabla f(x) = \begin{cases} -1 & \text{if } x \le -0.5 \\ 2x & \text{if } -0.5 \le x \le 0.5 \\ 1 & \text{if } x \ge 0.5 \end{cases}$$

- and

$$\nabla^2 f(x) = \begin{cases} 0 & \text{if } x \le -0.5 \\ 2 & \text{if } -0.5 \le x \le 0.5 \\ 0 & \text{if } x \ge 0.5 \end{cases}$$

(unique almost everywhere)