Operators

Pontus Giselsson

Today's lecture

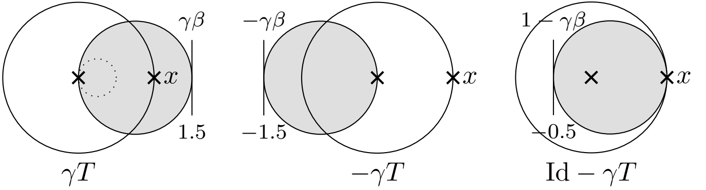
- forward step operators
- gradient step operators (subclass of forward step operators)
- resolvents
- proximal operators (subclass of resolvents)
- reflected resolvents
- reflected proximal operators (or proximal reflectors)

Forward step operator

- suppose that $T : \mathbb{R}^n \to \mathbb{R}^n$ is single-valued
- the forward step operator is $(\mathrm{Id} - \gamma T)$

Cocoercivity

- suppose that T is $\frac{1}{\beta}$-cocoercive with $\beta = \frac{1}{2}$
- then $\mathrm{Id} - \gamma T$ with $\gamma = 3$ is α-averaged, decide α:



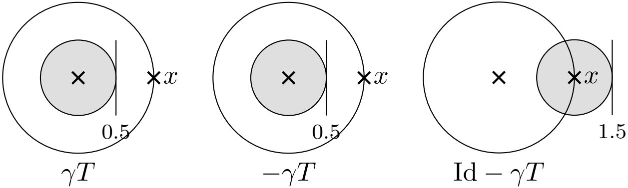
- T is 2-cocoercive $\Leftrightarrow$ γT is $\frac{2}{3}$-cocoercive
  $\Leftrightarrow$ $(\mathrm{Id} - \gamma T)$ is $\frac{3}{4}$-averaged
- generally: suppose $\gamma \in (0, \frac{2}{\beta})$
- then: $\frac{1}{\beta}$-cocoercivity of T $\Leftrightarrow$ $\frac{\gamma\beta}{2}$-averagedness of $(\mathrm{Id} - \gamma T)$

Iterating the forward step operator

- since $\frac{1}{\beta}$-cocoercivity of $T \Leftrightarrow \frac{\gamma\beta}{2}$-averagedness of $(\mathrm{Id} - \gamma T)$
- iterating $x^{k+1} = (\mathrm{Id} - \gamma T)x^k$ converges to fixed-point (if exists)
- (if $\gamma \in (0, \frac{2}{\beta})$)

Lipschitz continuity

- suppose that T is $\frac{1}{2}$-Lipschitz and $\gamma = 1$
- motivate that $\mathrm{Id} - \gamma T$ is not averaged



- cannot make $\mathrm{Id} - \gamma T$ nonexpansive independent of γ
- iterating forward step of Lipschitz T not guaranteed to converge

Gradient step operator

- suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable
- the forward step becomes the gradient step operator of f

$$I - \gamma \nabla f$$

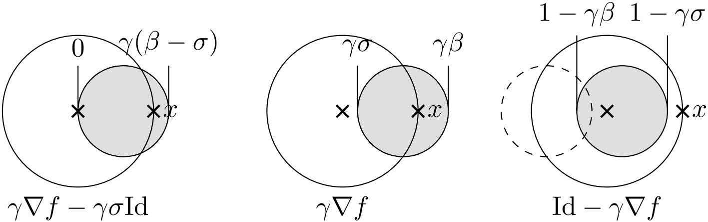
- if f (proper closed and) convex, then
  $\frac{1}{\beta}$-cocoercivity of $\nabla f \Leftrightarrow \beta$-Lipschitz continuity of ∇f
  $\Leftrightarrow \beta$-smoothness of f
- if f is β-smooth, the gradient method converges for $\gamma \in (0, \frac{2}{\beta})$

$$x^{k+1} = (\mathrm{Id} - \gamma \nabla f)x^k$$

  (since ∇f $\frac{1}{\beta}$-cocoercive)

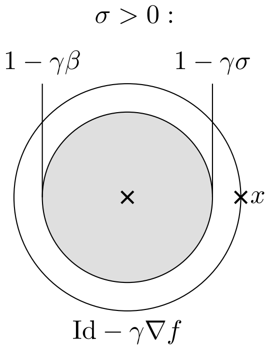
Stronger properties

- assume f is β-smooth and σ-strongly convex
- then γf is $\gamma\beta$-smooth and $\gamma\sigma$-strongly convex
- then $\gamma f - \frac{\gamma\sigma}{2}\|\cdot\|^2$ is $\gamma(\beta-\sigma)$-smooth
- or $\gamma\nabla f - \gamma\sigma\mathrm{Id}$ is $\frac{1}{\gamma(\beta-\sigma)}$-cocoercive
- $(\mathrm{Id} - \gamma\nabla f)$ is δ-Lipschitz, decide δ

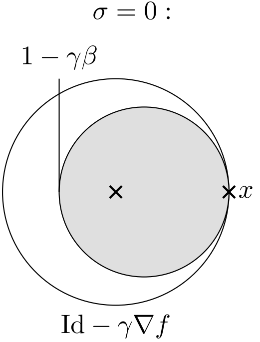


- $(\mathrm{Id} - \gamma T)$ is $\max(\gamma\beta - 1, 1 - \gamma\sigma)$-Lipschitz
- contractive if $1 - \gamma\sigma < 1$ and $\gamma\beta - 1 < 1$, i.e., $\gamma \in (0, \frac{2}{\beta})$
- gradient method $x^{k+1} = (\mathrm{Id} - \gamma T)x^k$ then converges linearly
- optimal γ (center circle) given by $\gamma = \frac{2}{\beta+\sigma} \Rightarrow \delta = \frac{\beta/\sigma - 1}{\beta/\sigma + 1}$

Summary: Gradient step operator



- $\gamma \in (0, \frac{2}{\beta}) \Rightarrow$ contractive
- optimal $\gamma = \frac{2}{\beta+\sigma} \Rightarrow$ factor $\frac{\beta/\sigma-1}{\beta/\sigma+1}(= \gamma\beta - 1 = 1 - \gamma\sigma)$



- $\gamma = 2\alpha/\beta, \alpha \in (0, 1)$
  $\Rightarrow 1 - \gamma\beta = 1 - 2\alpha$
  $\Rightarrow (\mathrm{Id} - \gamma\nabla f)$ α-averaged

Resolvent

- resolvent $J_A \ : \ \mathcal{D} \to \mathbb{R}^n$ to monotone operator is defined as

$$J_A = (\mathrm{Id} + A)^{-1}$$

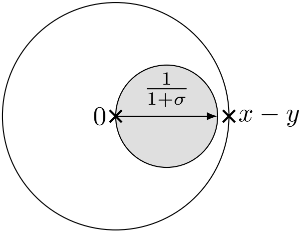
- due to Minty, if A maximally monotone, then $\mathcal{D} = \mathbb{R}^n$
(dom$(\mathrm{Id} + A^{-1})$ = ran$(\mathrm{Id} + A) = \mathbb{R}^n$ iff A maximally monotone)
- this is important for algorithms involving resolvent
- we will consider resolvents to maximally monotone operators

Properties of resolvent

- assume A is σ-strongly monotone ($\sigma = 0$ implies monotone)
- $\mathrm{Id} + A$ is $(1+\sigma)$-strongly monotone

$$\langle Ax - Ay + (x-y), x-y\rangle \geq \sigma\|x-y\|^2 + \|x-y\|^2 = (1+\sigma)\|x-y\|^2$$

- properties of $J_A = (\mathrm{Id} + A)^{-1}$?
- $J_A = (\mathrm{Id} + A)^{-1}$ is $(1+\sigma)$-cocoercive



- $\sigma = 0$: J_A is $\frac{1}{2}$-averaged (or 1-cocoercive or firmly nonexpansive)
- $\sigma > 0$: J_A is $\frac{1}{1+\sigma}$-contractive
- (iteration of the resolvent converges to a fixed-point, if exists)
- note: resolvent is single-valued

Lipschitz continuity

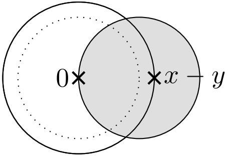
- assume A is β-Lipschitz continuous, then

$$2\langle J_A x - J_A y, x - y\rangle \geq \|x - y\|^2 + (1 - \beta^2)\|J_A x - J_A y\|^2$$

  (besides being 1-cocoercive)

proof sketch:

- $A + \beta\mathrm{Id}$ is $\frac{1}{2\beta}$-cocoercive

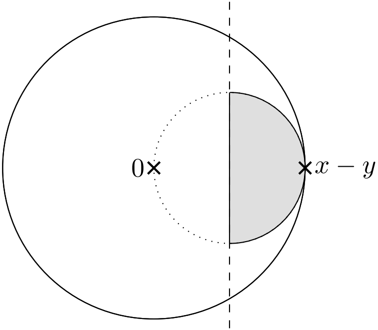


  - dotted: $Ax - Ay$
  - gray: $(\beta\mathrm{Id} + A)x - (\beta\mathrm{Id} + A)y$

- using $\beta\mathrm{Id} = \mathrm{Id} + (\beta - 1)\mathrm{Id}$, the definition of a cocoercive operator, and the definition of the inverse, gives the result

Graphical representation

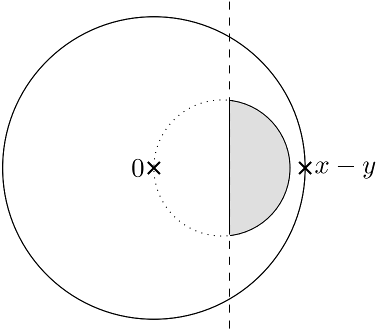
- assume A is 1-Lipschitz continuous
- then (besides being 1-cocoercive) J_A is $\frac{1}{2}$-strongly monotone



- 1-cocoercivity: $J_A x - J_A y$ in dotted region
- $\frac{1}{2}$-strong monotonicity: $J_A x - J_A y$ to the right of dashed line

Lipschitz continuity and strong monotonicity

- let A be 1-Lipschitz and σ-strongly monotone (with $0 \leq \sigma < 1$)
- σ-strong monotonicity of $A \Rightarrow (1+\sigma)$-cocoercivity of J_A
- 1-Lipschitz continuity of $A \Rightarrow \frac{1}{2}$-strong monotonicity of J_A
- intersect regions to find region when both properties are present



- $J_A x - J_A y$ ends up in gray region
- ($\sigma = \frac{1}{9}$ and $\beta = 1$ in figure)

Proximal operators

- assume f is proper closed and convex
- then ∂f maximally monotone
- let $A = \partial f$, then:

$$J_A(z) = \operatorname*{argmin}_x \left\{ f(x) + \tfrac{1}{2}\|x - z\|^2 \right\} =: \mathrm{prox}_f(z)$$

  where prox_f is called prox operator
- proof: $x = \mathrm{prox}_f(z)$ if and only if

$$\begin{aligned} & 0 \in \partial f(x) + x - z \\ \Leftrightarrow \quad & z \in \partial f(x) + x \\ \Leftrightarrow \quad & z \in (\mathrm{Id} + \partial f)x \\ \Leftrightarrow \quad & x = (\mathrm{Id} + \partial f)^{-1} z \end{aligned}$$

Proximal operator characterization

- the proximal operator satisfies

$$\mathrm{prox}_f = \nabla h^*$$

  where $h = f + \frac{1}{2}\|\cdot\|^2$

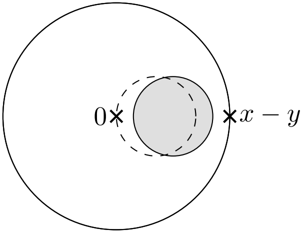
- why?
  - h is proper closed and convex, and $\partial h = \partial f + \mathrm{Id}$
  - therefore $\nabla h^* = (\partial h)^{-1} = (\partial f + \mathrm{Id})^{-1} = J_{\partial f}$
- can this be used to derive tighter properties of $J_{\partial f}$?

Proximal operator properties

- we have $\text{prox}_f(z) = \nabla h^*(z)$ where $h = f + \frac{1}{2}\|\cdot\|^2$
- recall equivalent dual properties
  - (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
- assume f is σ-strongly convex $\Rightarrow h$ is $(1+\sigma)$-strongly convex
  $\Rightarrow \nabla h^* = \text{prox}_f$ is $(1+\sigma)$-cocoercive
  (same as in general case)

Lipschitz continuity

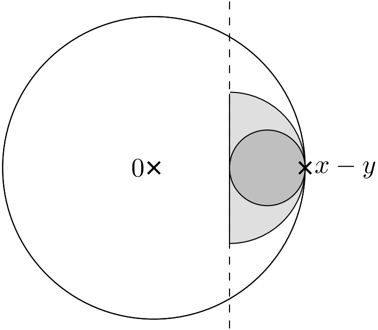
- let $h = f + \frac{1}{2}\|\cdot\|^2$, i.e. $\text{prox}_f = \nabla h^*$
- assume that f is β-smooth and σ-strongly convex $0 \leq \sigma \leq \beta$
- then h is $(\beta + 1)$-smooth and $(\sigma + 1)$-strongly convex
- therefore h^* is $\frac{1}{1+\beta}$-strongly convex and $\frac{1}{1+\sigma}$-smooth
- and $h^* - \frac{1}{2(1+\beta)}\|\cdot\|^2$ is $(\frac{1}{1+\sigma} - \frac{1}{1+\beta})$-smooth
- finally $\nabla h^* - \frac{1}{1+\beta}\mathrm{Id}$ is $\frac{1}{\frac{1}{1+\sigma} - \frac{1}{1+\beta}}$-cocoercive (if $\beta > \sigma$)



- $\nabla h^* - \frac{1}{1+\beta}\mathrm{Id} = \mathrm{prox}_f - \frac{1}{1+\beta}\mathrm{Id}$ inside dashed circle
- $\nabla h^* = \mathrm{prox}_f$ in gray area (shift by $\frac{1}{1+\beta}\mathrm{Id}$)
- (figure has $\beta = \frac{17}{3}$ and $\sigma = \frac{3}{17}$)

Comparison

- assume A is a maximal monotone operator and that f is PCC
- assume that A and ∂f are 1-Lipschitz
- J_A and prox_f end up in darker and lighter gray area respectively



- **conclusion**: under Lipschitz assumptions, the resolvent of subdifferentials are confined to smaller regions

Proximal operator for separable functions

- consider a separable function $g(x) = \sum_{i=1}^{n} g_i(x_i)$
- the prox is also separable:

$$
\begin{aligned}
\mathrm{prox}_g(z) &= \operatorname*{argmin}_x \{g(x) + \tfrac{1}{2}\|x - z\|^2\} \\
&= \operatorname*{argmin}_x \{\sum_{i=1}^{n} g_i(x_i) + \tfrac{1}{2}\sum_{i=1}^{n}(x_i - z_i)^2\} \\
&= \begin{bmatrix} \operatorname{argmin}_{x_1}\{g_1(x_1) + \tfrac{1}{2}(x_1 - z_1)^2\} \\ \ddots \\ \operatorname{argmin}_{x_n}\{g_n(x_n) + \tfrac{1}{2}(x_n - z_n)^2\} \end{bmatrix}
\end{aligned}
$$

- cheap evaluation $\Rightarrow$ good to have in algorithms

Separability and compositions

- assume that g is separable, i.e., $g(x) = \sum_{i=1}^{n} g_i(x_i)$
- let $h = g \circ L$ where L is arbitrary linear operator
- the prox becomes

$$\begin{aligned} \mathsf{prox}_h(z) &= \operatorname*{argmin}_x \{h(x) + \tfrac{1}{2}\|x - z\|^2\} \\ &= \operatorname*{argmin}_x \{g(Lx) + \tfrac{1}{2}\|x - z\|^2\} \end{aligned}$$

- separability is lost in general

Moreau's identity

- the following relation holds between the prox of f and f^*

$$\mathsf{prox}_f + \mathsf{prox}_{f^*} = \mathrm{Id}$$

- when f scaled by γ, we have

$$\mathsf{prox}_{\gamma f} + \mathsf{prox}_{(\gamma f)^*} = \mathsf{prox}_{\gamma f} + \gamma \mathsf{prox}_{\gamma^{-1} f^*} \circ \gamma^{-1}\mathrm{Id} = \mathrm{Id}$$

- when f composed with L, we have

$$\mathsf{prox}_{\gamma(f\circ L)}(z) = z - \gamma L^* \mu^*$$

  where

$$\mu^* \in \operatorname*{Argmin}_{\mu}\{f^*(\mu) + \tfrac{\gamma}{2}\|L^*\mu - \gamma^{-1}z\|^2\}$$

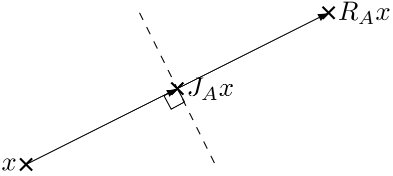
  (assuming that the Argmin is nonempty)
- these identities are very useful!

Reflected resolvent

- the reflected resolvent R_A to a monotone operator A is defined as

$$R_A := 2J_A - \mathrm{Id}$$

- it gives the reflection point (therefore its name)



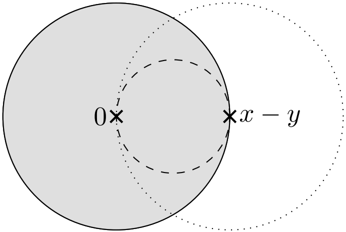
- if $A = \partial f$ then *reflected proximal operator* given by

$$R_{\partial f} = 2\mathrm{prox}_f - \mathrm{Id}$$

(sometimes denoted rprox_f or R_f)

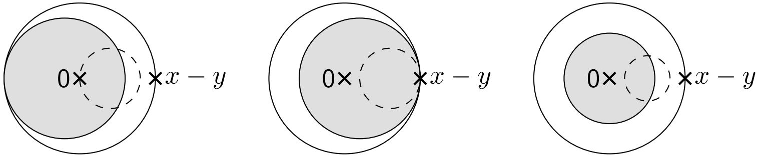
Properties of reflected resolvent

- in the general case, A monotone
- reflected resolvent R_A is β-Lipschitz, what is β?
- $\beta = 1$, i.e., R_A is nonexpansive
  "proof":
  1. $J_A x - J_A y$ within dashed region (since J_A 1-cocoercive in general case)
  2. $2J_A x - J_A y$ within dotted region (multiply by 2)
  3. $(2J_A - \mathrm{Id})x - (2J_A - \mathrm{Id})y = (2J_A x - 2J_A y) - (x - y)$ in gray area (shift by $-\mathrm{Id}$)



Further properties of reflected resolvent

- properties of R_A obtained by multiplying resolvent (J_A) area by 2 (radially) and shifting with $-\mathrm{Id}$
- examples: $A = \partial f$ is β-smooth and σ-strongly monotone

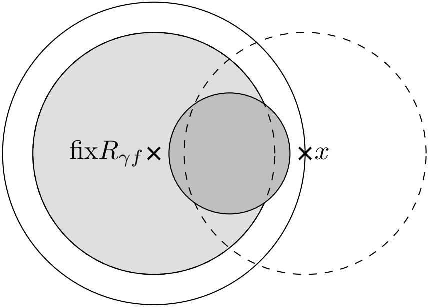


$\beta = \infty, \sigma = \frac{1}{4}$ $\quad$ $\beta = 4, \sigma = 0$ $\quad$ $\beta = 4, \sigma = \frac{1}{4}$

- left: negatively averaged, middle: averaged, right: contractive
- (fairly easy to visualize, can be harder to prove)

More properties of reflected proximal operator

- assume ∇f is σ-strongly monotone and β-Lipschitz
- then $\text{prox}_{\gamma f} - \frac{1}{1+\gamma\beta}\text{Id}$ is $\frac{1}{\frac{1}{1+\gamma\sigma} - \frac{1}{1+\gamma\beta}}$-cocoercive (if $\beta > \sigma$)
- it can be shown that $R_{\gamma f}$ is $\max\left(\frac{1-\gamma\sigma}{1+\gamma\sigma}, \frac{\gamma\beta-1}{1+\gamma\beta}\right)$-contractive



- contraction factor optimized for $\gamma = \frac{1}{\sqrt{\sigma\beta}}$

  (gives a contraction factor of $\frac{\sqrt{\beta/\sigma}-1}{\sqrt{\beta/\sigma}+1}$)

A reflected resolvent identity

- assume that A is a maximally monotone operator and $\gamma \in (0, \infty)$
- then

$$R_{\gamma A}(\mathrm{Id} + \gamma A) = \mathrm{Id} - \gamma A$$

- proof

$$\begin{aligned} R_{\gamma A}(\mathrm{Id} + \gamma A) &= 2(\mathrm{Id} + \gamma A)^{-1}(\mathrm{Id} + \gamma A) - (\mathrm{Id} + \gamma A) \\ &= 2\mathrm{Id} - (\mathrm{Id} + \gamma A) \\ &= (\mathrm{Id} - \gamma A) \end{aligned}$$

  where second step holds since $(\mathrm{Id} + \gamma A)^{-1} = J_{\gamma A}$ is single-valued