Monotone Operators and Fixed-Point Iterations

Pontus Giselsson

Today's lecture

- operators and their properties
    - monotone operators
    - Lipschitz continuous operators
    - averaged operators
    - cocoercive operators
- relation between properties
- monotone inclusion problems
    - special case: composite convex optimization
- resolvents and reflected resolvents
- Douglas-Rachford splitting
    - convergence

Power set

- the *power set* of the set $\mathcal{X}$ is the set of all subsets of $\mathcal{X}$.
- notation: $2^{\mathcal{X}}$ (since if number of elements in $\mathcal{X}$ is finite (n), then number of elements in the power set is 2^n).

Operators

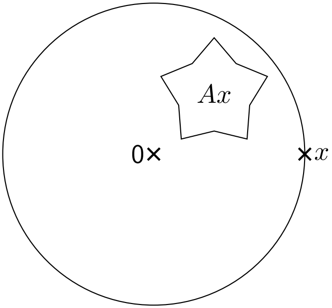
- an operator $A : \mathcal{H} \to 2^{\mathcal{H}}$ maps each point in $\mathcal{H}$ to a set in $\mathcal{H}$
- called *set-valued operator*
- Ax (or $A(x)$) means A operates on x (and gives a set back)
- if Ax is a singleton for all $x \in \mathcal{H}$, then A *single-valued*
  - can construct operator $B : \mathcal{H} \to \mathcal{H}$ with $\{Bx\} = Ax$ for all $x \in \mathcal{H}$
  - with slight abuse of notation, we treat these to be the same
- example:
  - the subdifferential operator ∂f is a set-valued operator
  - the gradient operator ∇f is a single-valued operator
- the graph of an operator $A : \mathcal{H} \to 2^{\mathcal{H}}$ is defined as

$$\mathrm{gph} A = \{(x, y) \mid y \in Ax\}$$

($\mathrm{gph} A$ is a subset of $\mathcal{H} \times \mathcal{H}$)

Graphical representation

- a set-valued operator $A : \mathcal{H} \to 2^{\mathcal{H}}$



- depending on where the set Ax is, A has different properties

Special operators

- the identity operator is denoted Id and is defined as

$$x = \mathrm{Id}(x)$$

- inverse of an operator

$$\mathrm{gph} A^{-1} = \{(y, x) \mid (x, y) \in \mathrm{gph} A\}$$

(therefore $y \in Ax$ if and only if $x \in A^{-1}y$)

Fixed points

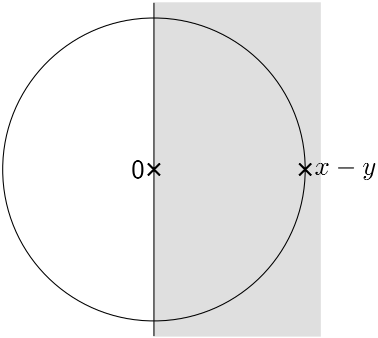
- a fixed-point y to the operator $A : \mathcal{H} \to \mathcal{H}$ satisfies $y = Ay$
- the set of fixed-points to $A : \mathcal{H} \to \mathcal{H}$ is denoted $\mathrm{fix} A$

Monotone operators

- an operator $A\ :\ \mathcal{H} \to 2^{\mathcal{H}}$ is monotone if

$$\langle x - y, u - v \rangle \geq 0$$

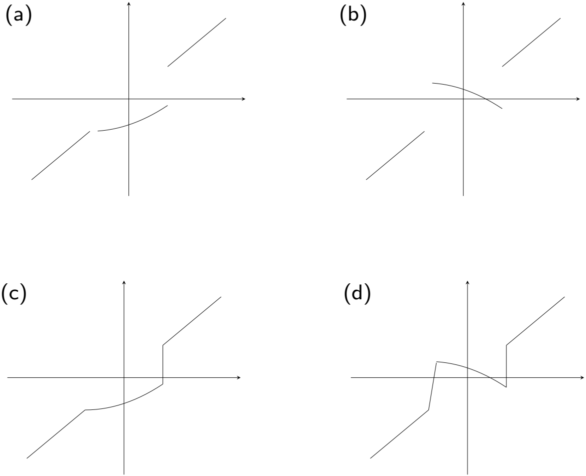
  for all $(x, u) \in \mathrm{gph} A$ and $(y, v) \in \mathrm{gph} A$
- graphical representation



then $u - v$ in gray area (since scalar product positive)
(or set $Ax \ominus Ay$ in gray area)

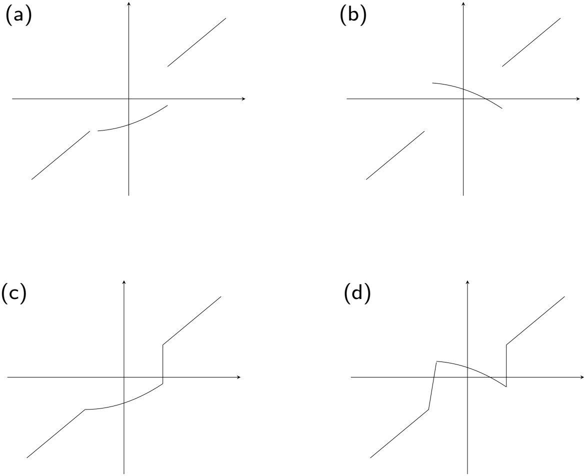
Monotonicity 1D

- which of the following operators $A : \mathbf{R} \to 2^{\mathbf{R}}$ are monotone?



Monotonicity 1D

- which of the following operators $A \;:\; \mathbf{R} \rightarrow 2^{\mathbf{R}}$ are monotone?



monotone: (a) and (c)
$(y - x > 0$ implies $v - u \geq 0$ where $(x,u),(y,v) \in \mathrm{gph(A)})$

Examples of monotone mappings

- the subdifferential ∂f of a proper, closed, convex function f
- proof: by convexity we have

$$f(x) \geq f(y) + \langle v, x - y \rangle$$
$$f(y) \geq f(x) + \langle u, y - x \rangle$$

for any $v \in \partial f(y)$ and $u \in \partial f(x)$, add these to get

$$\langle u - v, x - y \rangle \geq 0$$

Example of monotone mappings

- the subdifferential of the conjugate to a proper, closed, and convex function f, i.e., ∂f^* where

$$f^*(y) \triangleq \sup_x \{\langle y, x\rangle - f(x)\}$$

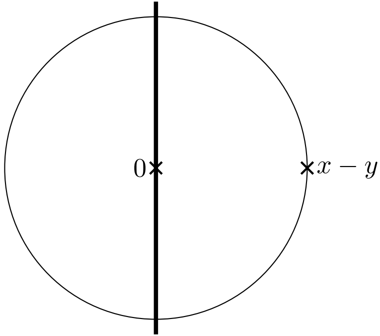
- we have $(\partial f)^{-1} = \partial f^*$

Examples of monotone mappings

- a (linear) skew-symmetric mapping (i.e., $A = -A^*$)
- proof:

$$\begin{aligned}\langle Ax - Ay, x - y\rangle &= \langle x - y, A^*(x - y)\rangle = -\langle x - y, A(x - y)\rangle \\ &= -\langle A(x - y), x - y\rangle = 0\end{aligned}$$

- graphical representation



then $Ax - Ay$ on thick black line

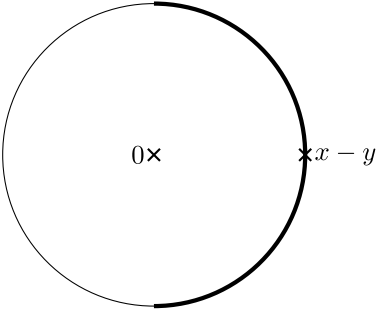
Examples of monotone mappings

- rotation $R_\theta \;:\; \mathbf{R}^2 \to \mathbf{R}^2$ with $|\theta| \le \frac{\pi}{2}$
- proof: let $v = x - y$

$$\langle R_\theta x - R_\theta y, x - y \rangle = \langle R_\theta v, v \rangle = \left\langle \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} v, v \right\rangle$$

$$= \left\langle \begin{bmatrix} v_1 \cos\theta - v_2 \sin\theta \\ v_1 \sin\theta + v_2 \cos\theta \end{bmatrix}, v \right\rangle = v_1^2 \cos\theta + v_2^2 \cos\theta \ge 0$$

- graphical representation



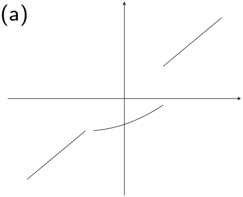
then $R_\theta(x - y)$ on thick semi-circle

Maximal monotonicity

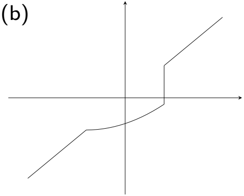
- a monotone operator $A : \mathcal{H} \to 2^{\mathcal{H}}$ is maximal monotone if no monotone operator $B : \mathcal{H} \to 2^{\mathcal{H}}$ exists such that $\mathrm{gph} A \subset \mathrm{gph} B$

- which of the following operators are maximal monotone

(a)



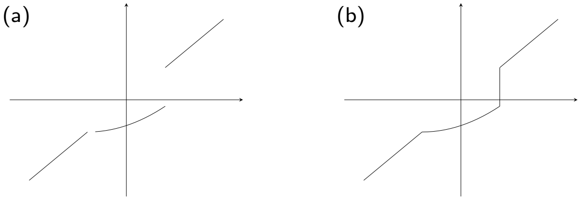
(b)



Maximal monotonicity

- a monotone operator $A : \mathcal{H} \to 2^{\mathcal{H}}$ is maximal monotone if no monotone operator $B : \mathcal{H} \to 2^{\mathcal{H}}$ exists such that $\mathrm{gph} A \subset \mathrm{gph} B$

- which of the following operators are maximal monotone

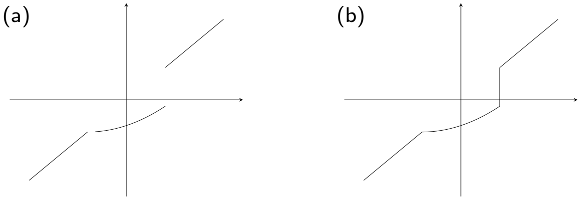


- maximally monotone: (b)

Maximal monotonicity

- a monotone operator $A : \mathcal{H} \to 2^{\mathcal{H}}$ is maximal monotone if no monotone operator $B : \mathcal{H} \to 2^{\mathcal{H}}$ exists such that $\mathrm{gph} A \subset \mathrm{gph} B$

- which of the following operators are maximal monotone



- maximally monotone: (b)
- subdifferentials of proper, closed, and convex functions are maximally monotone (not shown here)

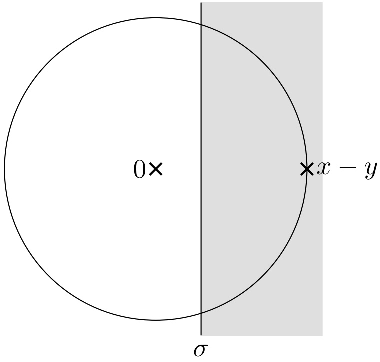
Strongly monotone operators

- an operator A is σ-strongly monotone if

$$\langle x - y, u - v\rangle \geq \sigma\|x - y\|^2$$

for all $(x, u) \in \mathrm{gph}A$ and $(y, v) \in \mathrm{gph}A$

- graphical representation



then $u - v$ in gray area (or set $Ax \ominus Ay$)

Strong convexity and strong monotonicity

- the subdifferential of a σ-strongly convex function is σ-strongly monotone
- proof:
  - by σ-strong convexity we have

$$f(x) \geq f(y) + \langle v, x - y \rangle + \tfrac{\sigma}{2}\|x - y\|^2$$
$$f(y) \geq f(x) + \langle u, y - x \rangle + \tfrac{\sigma}{2}\|x - y\|^2$$

  for any $v \in \partial f(y)$ and $u \in \partial f(x)$, add to get

$$\langle u - v, x - y \rangle \geq \sigma \|x - y\|^2$$

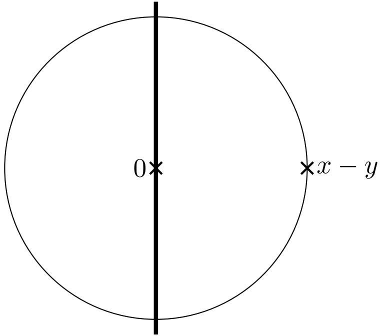
  - ($\sigma = 0$ shows that convexity of f implies monotonicity of ∂f)

Skew symmetric operator

- skew symmetric operator $A = -A^*$ (from before)

$$\langle Ax - Ay, x - y \rangle = 0$$

- not strongly monotone
- graphical representation

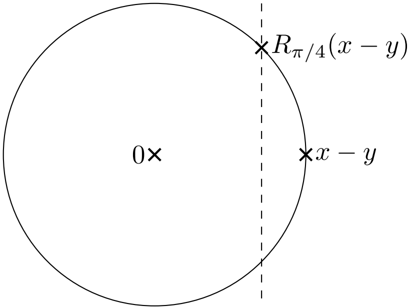


Rotation operator

- rotation operator R_θ with $|\theta| < \frac{\pi}{2}$ (from before)

$$\langle R_\theta x - R_\theta y, x - y \rangle \geq \cos\theta \|x - y\|^2$$

- R_θ is $\cos\theta$-strongly monotone
- graphical representation ($\theta = \frac{\pi}{4}$)

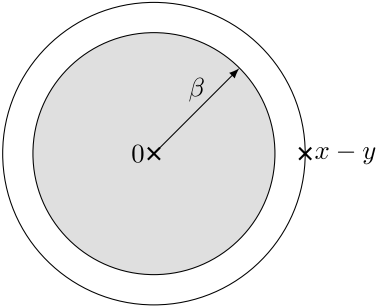


Lipschitz continuous operator

- an operator A is β-Lipschitz continuous if

$$\|Ax - Ay\| \leq \beta\|x - y\|$$

- A is single-valued (show by letting $y = x$ and use contradiction)
- graphical representation

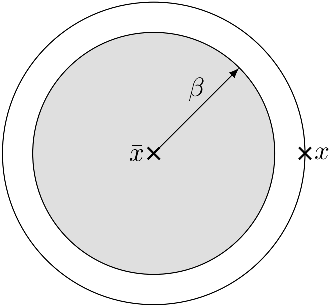


then $Ax - Ay$ is in gray area

Alternative graphical representation

- assume A has a fixed point $\bar{x} = A\bar{x}$ then

$$\|Ax - \bar{x}\| = \|Ax - A\bar{x}\| \leq \beta\|x - \bar{x}\|$$



then Ax in gray area

- interpretation: β relates to distance to fixed-point
- $\beta < 1$: contractive
- $\beta = 1$: nonexpansive

Examples

- a rotation is 1-Lipschitz continuous (nonexpansive)
- a linear mapping Mx is $\|M\|$-Lipschitz continuous

Convergence of contractive operator

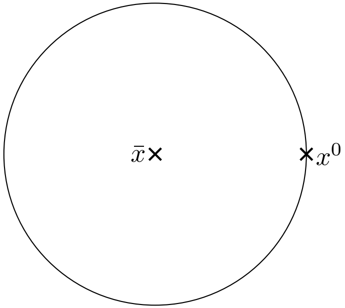
- a contractive ($\beta < 1$) operator A has a unique fixed-point $\bar{x}$ (Banach-Picard fixed-point theorem)
- the iteration $x^{k+1} = Ax^k$ converges linearly to the fixed-point ($\bar{x}$) if A is β-contractive:

$$\|x^{k+1} - \bar{x}\| = \|Ax^k - A\bar{x}\| \leq \beta\|x^k - \bar{x}\| \leq \beta^{k+1}\|x^0 - \bar{x}\|$$

Convergence of contractive operator

- a contractive ($\beta < 1$) operator A has a unique fixed-point $\bar{x}$ (Banach-Picard fixed-point theorem)
- the iteration $x^{k+1} = Ax^k$ converges linearly to the fixed-point ($\bar{x}$) if A is β-contractive:

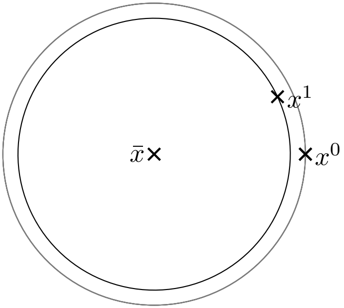
$$\|x^{k+1} - \bar{x}\| = \|Ax^k - A\bar{x}\| \leq \beta\|x^k - \bar{x}\| \leq \beta^{k+1}\|x^0 - \bar{x}\|$$



Convergence of contractive operator

- a contractive ($\beta < 1$) operator A has a unique fixed-point $\bar{x}$ (Banach-Picard fixed-point theorem)
- the iteration $x^{k+1} = Ax^k$ converges linearly to the fixed-point ($\bar{x}$) if A is β-contractive:

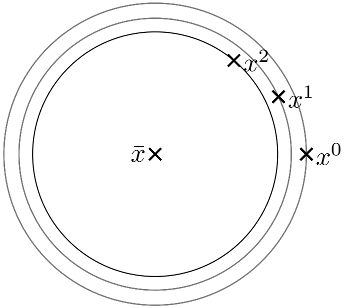
$$\|x^{k+1} - \bar{x}\| = \|Ax^k - A\bar{x}\| \leq \beta\|x^k - \bar{x}\| \leq \beta^{k+1}\|x^0 - \bar{x}\|$$



Convergence of contractive operator

- a contractive ($\beta < 1$) operator A has a unique fixed-point $\bar{x}$ (Banach-Picard fixed-point theorem)
- the iteration $x^{k+1} = Ax^k$ converges linearly to the fixed-point ($\bar{x}$) if A is β-contractive:

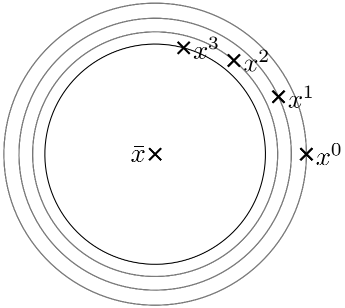
$$\|x^{k+1} - \bar{x}\| = \|Ax^k - A\bar{x}\| \leq \beta\|x^k - \bar{x}\| \leq \beta^{k+1}\|x^0 - \bar{x}\|$$



Convergence of contractive operator

- a contractive ($\beta < 1$) operator A has a unique fixed-point $\bar{x}$ (Banach-Picard fixed-point theorem)
- the iteration $x^{k+1} = Ax^k$ converges linearly to the fixed-point ($\bar{x}$) if A is β-contractive:

$$\|x^{k+1} - \bar{x}\| = \|Ax^k - A\bar{x}\| \leq \beta\|x^k - \bar{x}\| \leq \beta^{k+1}\|x^0 - \bar{x}\|$$



Fixed-points of nonexpansive operator

- nonexpansive operator need not have a fixed-point
- example:

Fixed-points of nonexpansive operator

- nonexpansive operator need not have a fixed-point
- example: $Ax = x + 2$

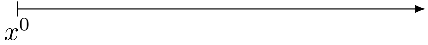
$$Ax = x + 2 \neq x$$

  for all $x \in \mathbf{R}$

- it is nonexpansive (1-Lipschitz continuous)

$$\|Ax - Ay\| = \|x + 2 - y - 2\| = \|x - y\|$$

- iteration $x^{k+1} = Ax^k$:



Fixed-points of nonexpansive operator

- nonexpansive operator need not have a fixed-point
- example: $Ax = x + 2$

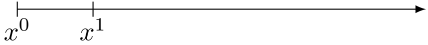
$$Ax = x + 2 \neq x$$

  for all $x \in \mathbf{R}$

- it is nonexpansive (1-Lipschitz continuous)

$$\|Ax - Ay\| = \|x + 2 - y - 2\| = \|x - y\|$$

- iteration $x^{k+1} = Ax^k$:



Fixed-points of nonexpansive operator

- nonexpansive operator need not have a fixed-point
- example: $Ax = x + 2$

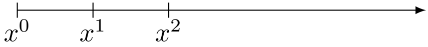
$$Ax = x + 2 \neq x$$

  for all $x \in \mathbf{R}$

- it is nonexpansive (1-Lipschitz continuous)

$$\|Ax - Ay\| = \|x + 2 - y - 2\| = \|x - y\|$$

- iteration $x^{k+1} = Ax^k$:



Fixed-points of nonexpansive operator

- nonexpansive operator need not have a fixed-point
- example: $Ax = x + 2$

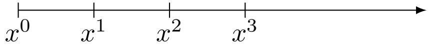
$$Ax = x + 2 \neq x$$

  for all $x \in \mathbf{R}$

- it is nonexpansive (1-Lipschitz continuous)

$$\|Ax - Ay\| = \|x + 2 - y - 2\| = \|x - y\|$$

- iteration $x^{k+1} = Ax^k$:



Fixed-points of nonexpansive operator

- nonexpansive operator need not have a fixed-point
- example: $Ax = x + 2$

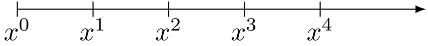
$$Ax = x + 2 \neq x$$

  for all $x \in \mathbf{R}$

- it is nonexpansive (1-Lipschitz continuous)

$$\|Ax - Ay\| = \|x + 2 - y - 2\| = \|x - y\|$$

- iteration $x^{k+1} = Ax^k$:



Fixed-points of nonexpansive operator

- nonexpansive operator need not have a fixed-point
- example: $Ax = x + 2$

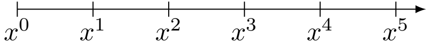
$$Ax = x + 2 \neq x$$

  for all $x \in \mathbf{R}$

- it is nonexpansive (1-Lipschitz continuous)

$$\|Ax - Ay\| = \|x + 2 - y - 2\| = \|x - y\|$$

- iteration $x^{k+1} = Ax^k$:



Fixed-points of nonexpansive operator

- nonexpansive operator need not have a fixed-point
- example: $Ax = x + 2$

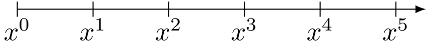
$$Ax = x + 2 \neq x$$

  for all $x \in \mathbf{R}$

- it is nonexpansive (1-Lipschitz continuous)

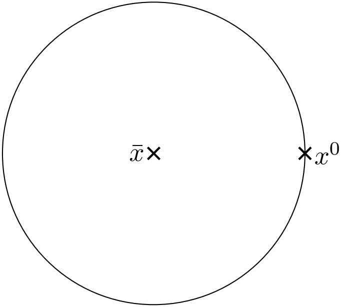
$$\|Ax - Ay\| = \|x + 2 - y - 2\| = \|x - y\|$$

- iteration $x^{k+1} = Ax^k$:



Convergence of nonexpansive operator

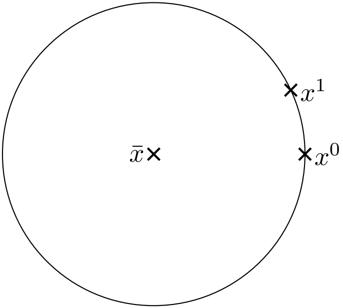
- if fixed-point $\bar{x}$ exists, iteration $x^{k+1} = Ax^k$ must not converge
- example: rotation by 25°



- however, the iterates are bounded

Convergence of nonexpansive operator

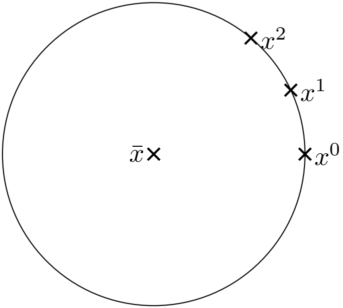
- if fixed-point $\bar{x}$ exists, iteration $x^{k+1} = Ax^k$ must not converge
- example: rotation by 25°



- however, the iterates are bounded

Convergence of nonexpansive operator

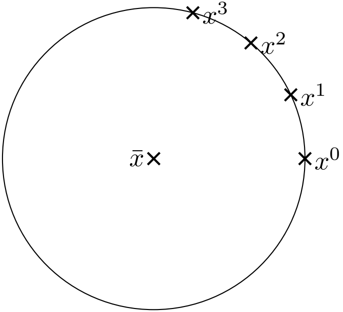
- if fixed-point $\bar{x}$ exists, iteration $x^{k+1} = Ax^k$ must not converge
- example: rotation by 25°



- however, the iterates are bounded

Convergence of nonexpansive operator

- if fixed-point $\bar{x}$ exists, iteration $x^{k+1} = Ax^k$ must not converge
- example: rotation by 25°



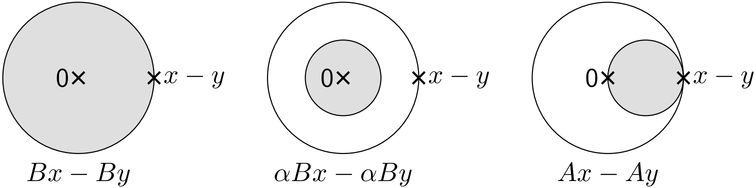
- however, the iterates are bounded

Averaged operators

- an operator A is α-averaged if and only if for some nonexpansive B and $\alpha \in (0,1)$:

$$A = (1-\alpha)\mathrm{Id} + \alpha B$$

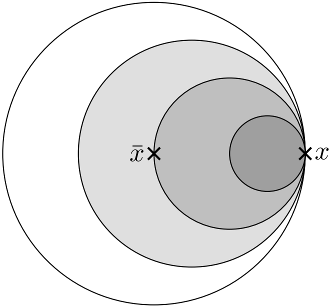
- graphical representation for $\alpha = 0.5$:



- for $\alpha = \frac{1}{2}$ we get $B = 2A - \mathrm{Id}$: A 0.5-averaged if and only if $2A - \mathrm{Id}$ nonexpansive
- $\frac{1}{2}$-averaged is called firmly nonexpansive

Additional graphical representation

- assume that $\bar{x}$ is a fixed-point to A_α which is α-averaged, then A_α can be represented as:



– 0.75-averaged – 0.5-averaged – 0.25-averaged

  where $A_\alpha x$ in respective gray areas
- why?
  - let $\bar{x} = y$
  - shift by $\bar{x}$: $(0 \to \bar{x}$, $x - \bar{x} \to x$, $Ax - A\bar{x} \to Ax - A\bar{x} + \bar{x} = Ax)$
- distance to fixed-point strictly decreased (except for if already at fixed-point)

Fixed-points

- the fixed-points of $A = (1-\alpha)\mathrm{Id} + \alpha B$ and B coincide (if they exist)
- proof
  - a fixed point $\bar{x}$ to B is a fixed-point to A:

$$A\bar{x} = (1-\alpha)\bar{x} + \alpha B\bar{x} = (1-\alpha+\alpha)\bar{x} = \bar{x}$$

  - a fixed-point $\bar{x}$ to A is a fixed-point to B:

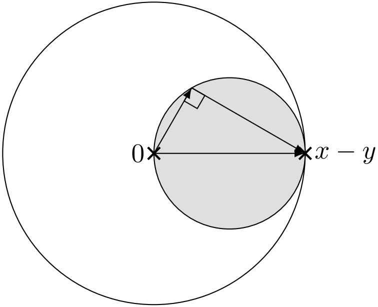
$$B\bar{x} = \tfrac{1}{\alpha}(A + (\alpha-1)\mathrm{Id})\bar{x} = \tfrac{1}{\alpha}(1+\alpha-1)\bar{x} = \bar{x}$$

Averaged operator formula

- α-averaged operator satisfies

$$\tfrac{1-\alpha}{\alpha}\|(I-A)x-(I-A)y\|^2+\|Ax-Ay\|^2 \leq \|x-y\|^2$$

- graphical representation for $\alpha = \frac{1}{2}$:



- can be used to show sub-linear convergence

Convergence

- the iterates for $x^{k+1} = Ax^k$ converge
- proof:

$$\tfrac{1-\alpha}{\alpha}\|x^k - x^{k+1}\|^2 = \tfrac{1-\alpha}{\alpha}\|(I-A)x^k - (I-A)x^\star\|^2$$
$$\leq \|x^k - x^\star\|^2 - \|x^{k+1} - x^\star\|^2$$

- summing over k:

$$\begin{aligned}(n+1)\|x^{n+1} - x^n\|^2 &\leq \sum_{k=0}^{n} \|x^{k+1} - x^k\|^2 \\ &\leq \frac{\alpha \sum_{k=1}^{n} \left(\|x^k - x^\star\|^2 - \|x^{k+1} - x^\star\|^2\right)}{1-\alpha} \\ &= \frac{\alpha\|x^0 - x^\star\|^2}{1-\alpha}\end{aligned}$$

- that is

$$\|x^{n+1} - x^n\|^2 \leq \frac{\alpha\|x^0 - x^\star\|^2}{(n+1)(1-\alpha)}$$

- optimize w.r.t. α gives $\alpha \to 0$ (not very informative since consecutive iterates close if short steps)

Convergence

- convergence towards fixed-point:
- proof:

$$\begin{aligned}\tfrac{1-\alpha}{\alpha}\|x^{n+1}-x^n\|^2 &= \tfrac{1-\alpha}{\alpha}\|(1-\alpha)x^n+\alpha B(x^n)-x^n\|^2\\ &= \alpha(1-\alpha)\|B(x^n)-x^n\|^2\end{aligned}$$

- therefore

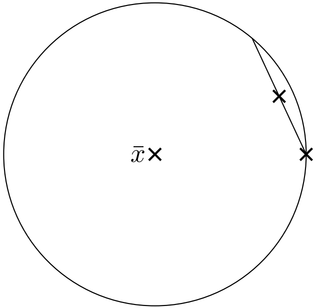
$$\|B(x^n)-x^n\|^2 = \tfrac{1}{\alpha^2}\|x^{n+1}-x^n\|^2 \leq \frac{\|x^0-x^\star\|^2}{(n+1)(1-\alpha)\alpha}$$

- optimize constant by letting $\alpha=\frac{1}{2}$:

$$\|B(x^n)-x^n\|^2 \leq \frac{4\|x^0-x^\star\|^2}{(n+1)}$$

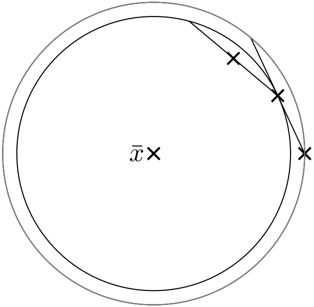
Convergence example - $\alpha = 0.5$

- rotation operator R_θ with $\theta = 50^\circ$
- fixed-point $\bar{x}$ at origin
- iterate 0.5-averaged operator



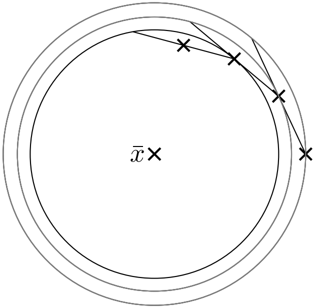
Convergence example - $\alpha = 0.5$

- rotation operator R_θ with $\theta = 50^\circ$
- fixed-point $\bar{x}$ at origin
- iterate 0.5-averaged operator



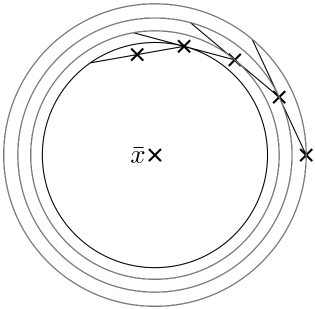
Convergence example - $\alpha = 0.5$

- rotation operator R_θ with $\theta = 50^\circ$
- fixed-point $\bar{x}$ at origin
- iterate 0.5-averaged operator



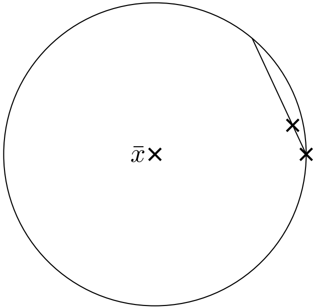
Convergence example - $\alpha = 0.5$

- rotation operator R_θ with $\theta = 50^\circ$
- fixed-point $\bar{x}$ at origin
- iterate 0.5-averaged operator



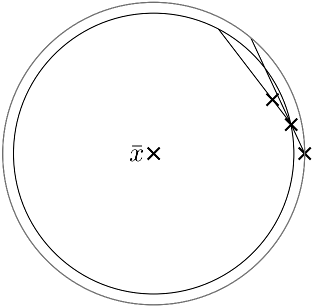
Example - $\alpha = 0.25$

- rotation operator R_θ with $\theta = 50^\circ$
- fixed-point $\bar{x}$ at origin
- iterate 0.25-averaged operator



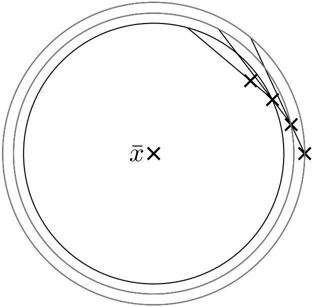
Example - $\alpha = 0.25$

- rotation operator R_θ with $\theta = 50^\circ$
- fixed-point $\bar{x}$ at origin
- iterate 0.25-averaged operator



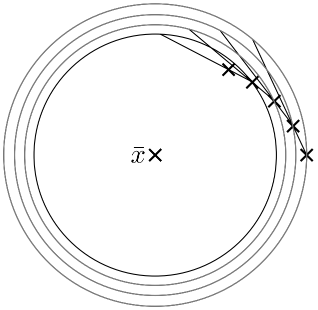
Example - $\alpha = 0.25$

- rotation operator R_θ with $\theta = 50^\circ$
- fixed-point $\bar{x}$ at origin
- iterate 0.25-averaged operator



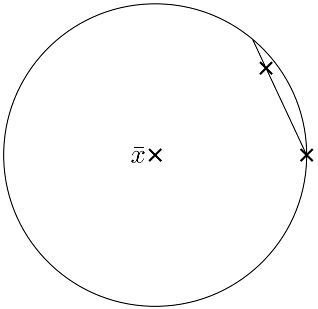
Example - $\alpha = 0.25$

- rotation operator R_θ with $\theta = 50^\circ$
- fixed-point $\bar{x}$ at origin
- iterate 0.25-averaged operator



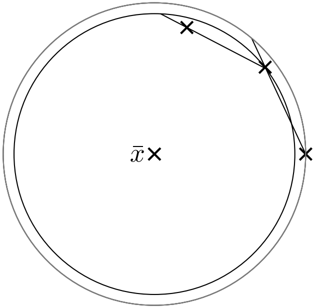
Example - $\alpha = 0.75$

- rotation operator R_θ with $\theta = 50^\circ$
- fixed-point $\bar{x}$ at origin
- iterate 0.75-averaged operator



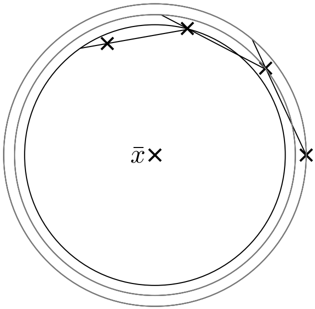
Example - $\alpha = 0.75$

- rotation operator R_θ with $\theta = 50^\circ$
- fixed-point $\bar{x}$ at origin
- iterate 0.75-averaged operator



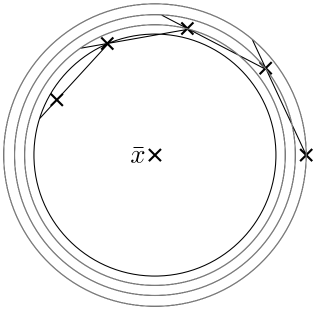
Example - $\alpha = 0.75$

- rotation operator R_θ with $\theta = 50^\circ$
- fixed-point $\bar{x}$ at origin
- iterate 0.75-averaged operator



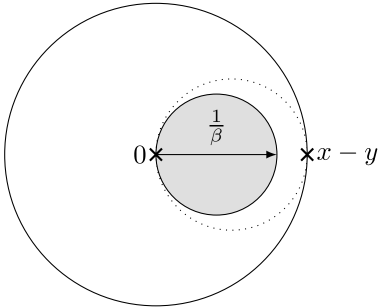
Example - $\alpha = 0.75$

- rotation operator R_θ with $\theta = 50^\circ$
- fixed-point $\bar{x}$ at origin
- iterate 0.75-averaged operator



Cocoercive operators

- an operator A is β-cocoercive if βA is $\frac{1}{2}$-averaged



- $Ax - Ay$ in gray area (dotted area shows that $\beta Ax - \beta Ax$ is $\frac{1}{2}$-averaged)

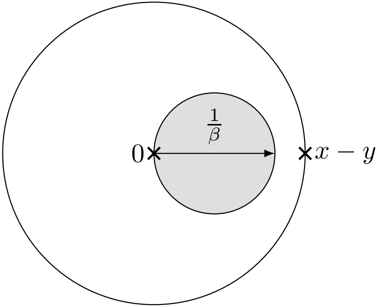
Cocoercive operator properties

- an operator A is β-cocoercive if βA is $\frac{1}{2}$-averaged, i.e.

$$\|(I - \beta A)x - (I - \beta A)y\|^2 + \|\beta Ax - \beta Ay\|^2 \leq \|x - y\|^2$$

- equivalently (by expanding the first square and div. by β)

$$\langle Ax - Ay, x - y\rangle \geq \beta\|Ax - Ay\|^2$$



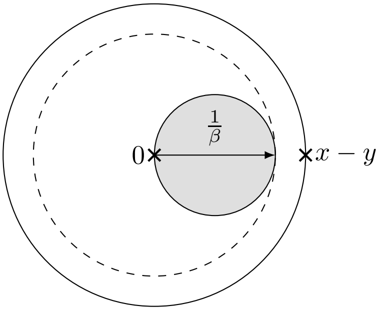
Properties

- β-cocoercivity implies γ-Lipschitz continuity:
- estimate γ?

Properties

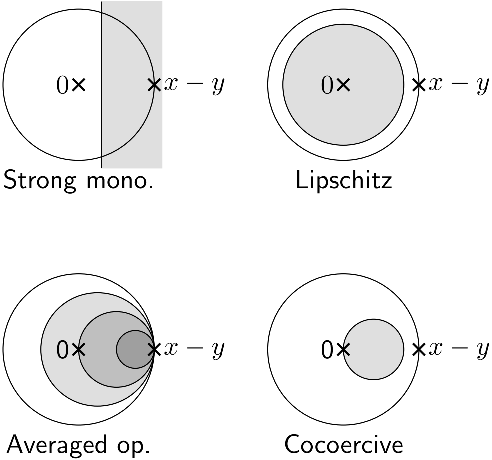
- β-cocoercivity implies γ-Lipschitz continuity:
- estimate γ?
- $\gamma = \frac{1}{\beta}$

$$\beta\|Ax - Ay\|^2 \leq \langle Ax - Ay, x - y\rangle \leq \|x - y\|\|Ax - Ay\|$$



Summary properties

- we have discussed operators A with the following properties



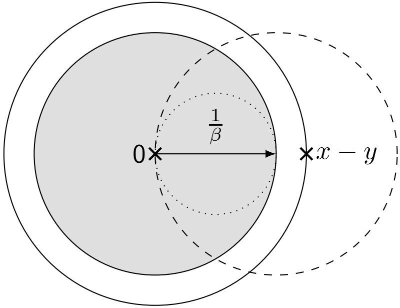
- the set (or point) $Ax \ominus Ay$ is in the respective gray areas

Exercise I

- assume that A is β-cocoercive
- estimate a small Lipschitz constant to $2A - \frac{1}{\beta}\mathrm{Id}$

Exercise I

- assume that A is β-cocoercive
- estimate a small Lipschitz constant to $2A - \frac{1}{\beta}\mathrm{Id}$
- a Lipschitz constant is $\frac{1}{\beta}$
  "proof":
  1. due to cocoercivity of A we have $Ax - Ay$ in dotted circle
  2. multiply by 2 ($2Ax - 2Ay$ in dashed)
  3. shift by $-\frac{1}{\beta}\mathrm{Id}$ ($(2A - \frac{1}{\beta}\mathrm{Id})x - (2A - \frac{1}{\beta}\mathrm{Id})y$ in gray)

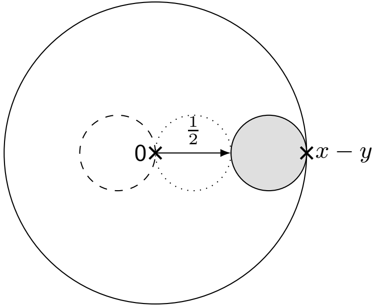


Exercise II

- assume that A is 2-cocoercive
- $\mathrm{Id} - A$ is α-averaged, compute α

Exercise II

- assume that A is 2-cocoercive
- $\mathrm{Id} - A$ is α-averaged, compute α
- $\mathrm{Id} - A$ is 0.25-averaged
  "proof":
  1. due to 2-cocoercivity of A, we have $Ax - Ay$ in dotted circle
  2. multiply by -1 ($-Ax + Ay$ in dashed)
  3. shift by Id ($(\mathrm{Id} - A)x - (\mathrm{Id} - A)y$ in gray)



Relation to (strong) monotonicity?

- can relate Lipschitz continuity, cocoercivity, and averagedness by scaling and shifting (they are all circles)
- cannot relate to (strong) monotonicity

Dual properties I

- consider the following list of properties
  (i) A is β-strongly monotone
  (ii) A^{-1} is β-cocoercive
  (iii) A^{-1} is $\frac{1}{\beta}$-Lipschitz continuous

  we have (i)$\Leftrightarrow$(ii) and (ii)$\Rightarrow$(iii)
- the result also holds with A and A^{-1} interchanged

Dual properties II

- for proper, closed, and convex f, the following are equivalent:
  (i) f is β-strongly convex

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

  for all $u \in \partial f(y)$
  (ii) ∂f is β-strongly monotone
  (iii) ∂f^* is β-cocoercive
  (iv) ∂f^* is $\frac{1}{\beta}$-Lipschitz continuous
  (v) f^* is $\frac{1}{\beta}$-smooth

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

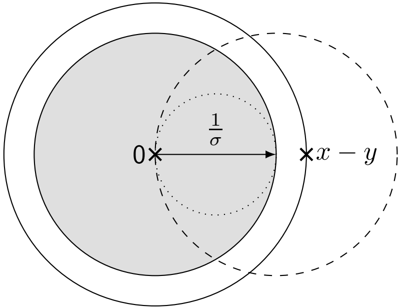
- the result also holds with f and f^* interchanged
- we have implication (iv)$\Rightarrow$(iii) as opposed to general case
- (recall $\partial f^* = (\partial f)^{-1}$)

Exercise I revisited

- A^{-1} is σ-strongly monotone
- estimate a small Lipschitz constant to $2A - \frac{1}{\sigma}\mathrm{Id}$

Exercise I revisited

- A^{-1} is σ-strongly monotone
- estimate a small Lipschitz constant to $2A - \frac{1}{\sigma}\mathrm{Id}$
- a Lipschitz constant is $\frac{1}{\sigma}$
  "proof":
  1. (i)$\Rightarrow$(ii) implies A is σ-cocoercive ($Ax - Ay$ in dotted)
  2. multiply by 2 ($2Ax - 2Ay$ in dashed)
  3. shift by $-\frac{1}{\sigma}\mathrm{Id}$ ($(2A - \frac{1}{\sigma}\mathrm{Id})x - (2A - \frac{1}{\sigma}\mathrm{Id})y$ in gray)

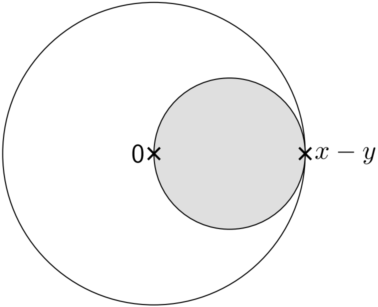


Exercise III

- A is 1-strongly convex
- A^{-1} is α-averaged, compute α

Exercise III

- A is 1-strongly convex
- A^{-1} is α-averaged, compute α
- A^{-1} is $\frac{1}{2}$-averaged
  "proof":
  1. (i)$\Rightarrow$(ii) gives that A^{-1} is 1-cocoercive ($A^{-1}x - A^{-1}y$ in gray)
  2. 1-cocoercivity defined as $\frac{1}{2}$-averagedness



Summary

- we have discussed the following operator properties
  1. (strong) monotonicity
  2. Lipschitz continuity (nonexpansiveness, contractiveness)
  3. averaged operators
  4. cocoercive operators
- 2., 3., and 4. are related to each other by scaling and translating
- 2., 3., and 4. are related to 1. through the inverse operator
- iteration of averaged operators converge (sublinearly)
- iteration of contractive operators converge linearly

Monotone inclusion problems

- we want to solve monotone inclusion problems of the form

$$0 \in A(x) + B(x)$$

  where A and B are maximal monotone operators
- special case:

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

  is equivalent to

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

- how to use the presented framework?

Creating algorithms

- state optimal point x as a fixed-point equation of some operator
- show that operator is either
  - α-averaged (sublinear convergence)
  - β-contractive (linear convergence)

Resolvent

- resolvent $J_A \ : \ \mathcal{D} \to \mathcal{H}$ to monotone operator is defined as

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

- if A maximally monotone, then $\mathcal{D} = \mathcal{H}$
  (important for algorithms involving the resolvent)

Resolvent

- resolvent $J_A \; : \; \mathcal{D} \to \mathcal{H}$ to monotone operator is defined as

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

- if A maximally monotone, then $\mathcal{D} = \mathcal{H}$
  (important for algorithms involving the resolvent)
- subdifferential case $A = \partial f$:

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

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

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

Properties of resolvent

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

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

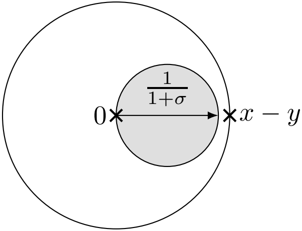
- properties of $J_A = (\mathrm{Id} + A)^{-1}$?

Properties of resolvent

- assume A 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



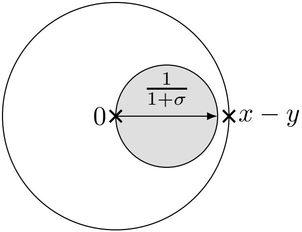
- $\sigma = 0$:

Properties of resolvent

- assume A 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



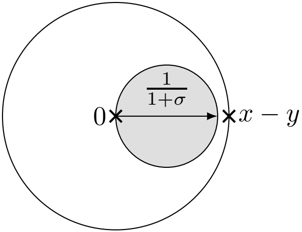
- $\sigma = 0$: J_A is $\frac{1}{2}$-averaged (or 1-cocoercive)
- $\sigma > 0$:

Properties of resolvent

- assume A 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)
- $\sigma > 0$: J_A is $\frac{1}{1+\sigma}$-contractive
- (iteration of the resolvent converges to a fixed-point)

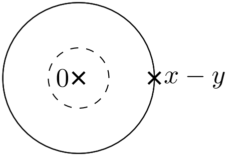
Further properties

- assume A is β-Lipschitz continuous
- $\mathrm{Id} + A$ is $(1+\beta)$-Lipschitz continuous

$$\|Ax - Ay + x - y\| \le \|Ax - Ay\| + \|x - y\| \le (1+\beta)\|x - y\|$$

- $J_A = (\mathrm{Id} + A)^{-1}$ satisfies (by definition of inverse operator)

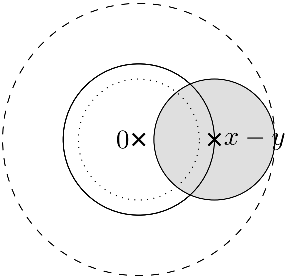
$$\|x - y\| \le (1+\beta)\|J_A x - J_A y\|$$



where $J_A x - J_A y$ *outside* dashed region (with radius $\frac{1}{1+\beta}$)

Suboptimal characterization

- still assume A is β-Lipschitz continuous
- previous characterization ($1 + \beta$-Lipschitz) of $\mathrm{Id} + A$ not tight!



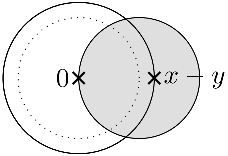
- dotted: $Ax - Ay$
- gray: $(\mathrm{Id} + A)x - (\mathrm{Id} + A)y$
- dashed: $(1 + \beta)$-Lipschitz continuity circle

Improved property

- still assume A is β-Lipschitz continuous
- property of $A + \beta \mathrm{Id}$?

Improved property

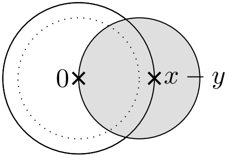
- still assume A is β-Lipschitz continuous
- property of $A + \beta\mathrm{Id}$?
- it is $\frac{1}{2\beta}$-cocoercive



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

Improved property

- still assume A is β-Lipschitz continuous
- property of $A + \beta\mathrm{Id}$?
- it is $\frac{1}{2\beta}$-cocoercive



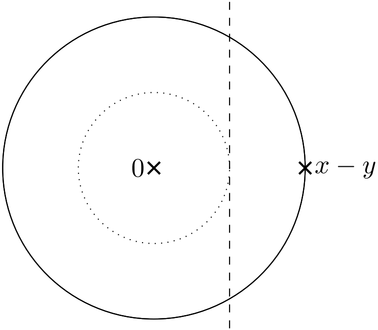
  - dotted: $Ax - Ay$
  - gray: $(\beta\mathrm{Id} + A)x - (\beta\mathrm{Id} + A)y$

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

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

Comparison of properties

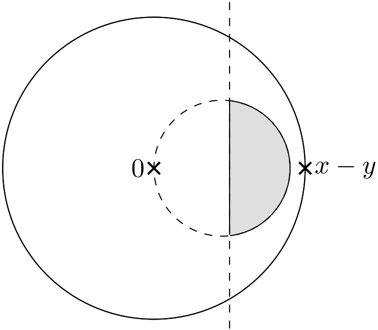
- assume A is β-Lipschitz continuous
- compare the two properties for J_A for $\beta = 1$



- first property: $J_A x - J_A y$ outside dotted region
- improved property: $J_A x - J_A y$ to the right of dashed line (J_A is $\frac{1}{2}$-monotone)

Combining properties

- let A be 1-Lipschitz and σ-strongly monotone (with $0 \le \sigma < 1$)
  - strong monotonity of A implies cocoerciveness of J_A
  - Lischitz continuity of A implies "improved property" 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 operator

- can properties be tighter when the resolvent is a prox operator?
- recall

$$J_{\partial f}(z) = \mathrm{prox}_f(z) = \underset{x}{\mathrm{argmin}} \left\{ f(x) + \tfrac{1}{2}\|x - z\|^2 \right\}$$

- define $h = \frac{1}{2}\| \cdot \|^2 + f$, properties:
  - f is σ-strongly convex implies h is $(1+\sigma)$-strongly convex
  - f is β-smooth implies h is $(1+\beta)$-smooth
  - we have $\partial h = (\mathrm{Id} + \partial f)$
- the prox operator satisfies

$$\mathrm{prox}_f(z) = (\mathrm{Id} + \partial f)^{-1} z = (\partial h)^{-1} z = \nabla h^*(z)$$

Proximal opertor properties

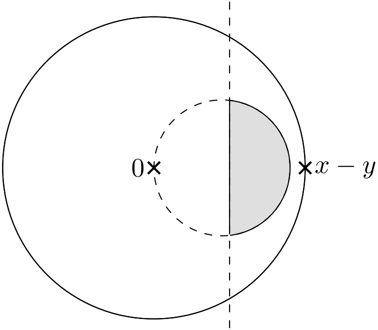
- we have $\mathrm{prox}_f(z) = \nabla h^*(z)$ where $h = \frac{1}{2}\|\cdot\|^2 + f$
- recall equivalent dual properties
  - (i) f is β-strongly convex
  - (ii) ∂f is β-strongly monotone
  - (iii) ∂f^* is β-cocoercive
  - (iv) ∂f^* is $\frac{1}{\beta}$-Lipschitz continuous
  - (v) f^* is $\frac{1}{\beta}$-smooth
- this gives

f	h	$\nabla h^* = \mathrm{prox}_f$
σ-str. cvx	$(1+\sigma)$-str. cvx.	$\frac{1}{1+\sigma}$-cocoercive
β-smooth	$(1+\beta)$-Lipschitz	$\frac{1}{1+\beta}$-str. mono.

- first property same and in general case, second property different

Graphical representation

- consider the case $\beta = 1$

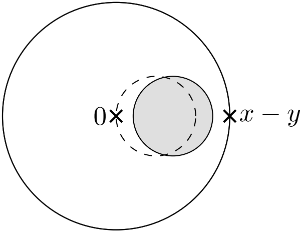


the same as in the general case

- can be improved

Improved property

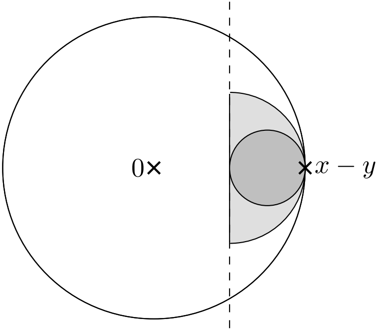
- assume that f is β-smooth (and σ-strongly convex $0 \le \sigma \le \beta$)
- then $\nabla h^* = \mathrm{prox}_f$ is $\frac{1}{1+\beta}$-strongly monotone and $\frac{1}{1+\sigma}$-Lipschitz
- further h^* is $\frac{1}{1+\beta}$-strongly convex and $\frac{1}{1+\sigma}$-smooth
- and $h^* - \frac{1}{2(1+\beta)}\|\cdot\|^2$ 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



- $\nabla h^* - \frac{1}{1+\beta}\mathrm{Id} = \mathrm{prox}_f - \frac{1}{1+\beta}\mathrm{Id}$ 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 general operator and that $B = \partial f$
- assume that A and ∂f are 1-Lipschitz and σ-strongly monotone
- the prox operator ends up in:



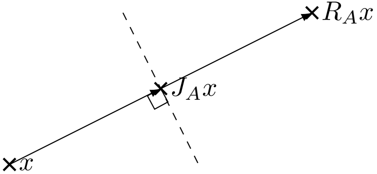
  where $J_A x - J_A y$ in light area and $J_{\partial f} x - J_{\partial f} y$ in darker area
- ($\sigma = 0$ in figure, i.e., only monotonicity is assumed)
- **conclussion**: under Lipschitz assumptions, the resolvent of subdifferentials are confined to smaller regions

Reflected resolvent

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

$$R_A := 2J_A - I$$

- it gives the reflection point (therefore its name)

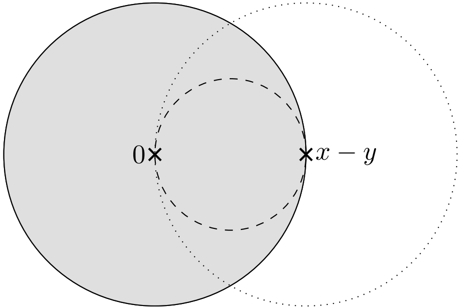


Properties of reflected resolvent

- in the general case, A monotone
- reflected resolvent R_A is β-Lipschitz, what is β?

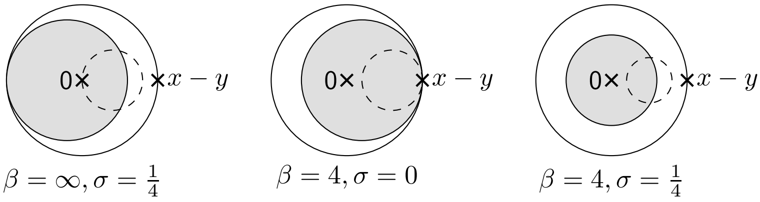
Properties of reflected resolvent

- 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 under different assumptions obtained by multiplying resolvent area by 2 (radially) and shifting by $-\mathrm{Id}$ $(-(x-y))$
- examples: subdifferential operator A is β-smooth and σ-strongly monotone



- contractive if $\beta < \infty$ and $\sigma > 0$

How to use these operators?

- how to use these operators to solve monotone inclusion problems

$$0 \in A(x) + B(x)$$

Optimality conditions

- inclusion problem with A and B maximally monotone

$$0 \in A(x) + B(x)$$

- x solves inclusion problem iff

$$z = R_{\gamma A} R_{\gamma B} z \qquad\qquad x = J_{\gamma A}(z)$$

  with $\gamma > 0$, i.e., z is a fixed-point to composition $R_{\gamma A} R_{\gamma B}$

- algorithm: find fixed-point to $R_{\gamma A} R_{\gamma B}$ to solve problem

(Generalized) Douglas-Rachford splitting

- iterate $R_{\gamma A}R_{\gamma B}$ to find fixed-point (Peaceman-Rachford splitting)

$$z^{k+1} = R_{\gamma A}R_{\gamma B}z^k$$

- $R_{\gamma A}$ and $R_{\gamma B}$ are nonexpansive in general case, so is composition
  $\Rightarrow$ algorithm not guaranteed to converge in general case
- need an averaged or contractive operator to converge
- introduce averaging with $\alpha \in (0,1)$:

$$z^{k+1} = ((1-\alpha)\mathrm{Id} + \alpha R_{\gamma A}R_{\gamma B})z^k$$

- $\alpha = \frac{1}{2}$ usually called Douglas-Rachford splitting (here for all α)

Convergence to fixed-point

- the Douglas-Rachford algorithm converges to fixed point of

$$(1-\alpha)\mathrm{Id} + \alpha R_{\gamma A} R_{\gamma B}$$

- fixed points coincide with fixed points of $R_{\gamma A} R_{\gamma B}$ (shown earlier)
- convergence is sublinear (shown earlier)

Linear convergence

- we get linear convergence if either of the following hold
  - A is σ-strongly monotone and β-Lipschitz
  - A is σ-strongly monotone and B is β-Lipschitz continuous
- reason: $(1-\alpha)\mathrm{Id} + \alpha R_{\gamma A} R_{\gamma B}$ contractive
- can choose γ and α to optimize rates
- different rates in general case and subdifferential case

ADMM

- ADMM $=$ the alternating direction method of multipliers
- consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) + g(y) \\ \text{subject to} & Ax = y \end{array}$$

- dual problem

$$\text{maximize} \quad \inf_{x,y} \big(f(x) + g(y) + \mu^T (Ax - y) \big)$$

- rewrite by identifying conjugates ($f^*(z) = \sup_x \{\langle z, x\rangle - f(x)\}$)

$$\text{minimize} \quad d(\mu) + g^*(\mu)$$

  where $d(\mu) = f^*(-A^T\mu)$
- apply DR to dual to get ADMM
- all convergence properties from DR translate to ADMM (use "Dual properties II" to infer properties of d and g^* from f and g)

Project

- provide linear convergence rates for Douglas-Rachford splitting in general case under assumptions
  - A is σ-strongly monotone and β-Lipschitz
  - A is σ-strongly monotone and B is β-Lipschitz
- optimize Douglas-Rachford algorithm parameters γ and α
- provide examples that achieve the rate exactly (if possible)

Summary

- introduced operators with different properties
  - (strong) monotonicity
  - Lipschitz continuity, nonexpansiveness, contractiveness
  - averaged operators
  - cocoercive operators
- dual properties
- stated monotone inclusion problems
- introduced resolvent and reflected resolvent
- described Douglas-Rachford splitting and "showed" convergence