Operator Properties

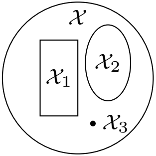
Pontus Giselsson

Today's lecture

- properties of *set-valued operators*:
  - monotonicity
  - maximal monotonicity
  - strong monotonicity
- properties of *single-valued operators*
  - Lipschitz continuity (contractiveness, nonexpansiveness)
  - averagedness
  - cocoercivity

Power set

- the *power set* of the set $\mathcal{X}$ is the set of all subsets of $\mathcal{X}$.
- our notation: $2^{\mathcal{X}}$
  - background: if number of elements in $\mathcal{X}$ is finite (n), then number of elements in the power set is 2^n
- other notations exist: $\mathcal{P}(\mathcal{X})$, $\wp(\mathcal{X})$, etc
- example:



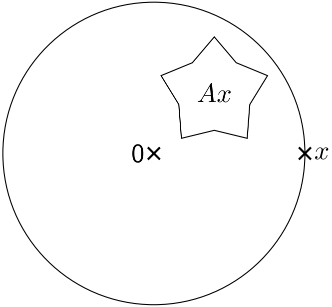
we have: $\mathcal{X}_1 \in 2^{\mathcal{X}}$, $\mathcal{X}_2 \in 2^{\mathcal{X}}$, $\mathcal{X}_3 \in 2^{\mathcal{X}}$, $\emptyset \in 2^{\mathcal{X}}$, $\mathcal{X} \in 2^{\mathcal{X}}$

Operators

- an operator $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ maps each point in $\mathbb{R}^n$ to a set in $\mathbb{R}^n$
- 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 $x \in \mathbb{R}^n$, then A *single-valued*
  - can construct $T : \mathbb{R}^n \to \mathbb{R}^n$ with $\{Tx\} = Ax$ for all $x \in \mathbb{R}^n$
  - with slight abuse of notation, we treat these to be the same
- examples:
  - the subdifferential operator ∂f is a set-valued operator
  - the gradient operator ∇f is a single-valued operator

Graphical representation

- a set-valued operator $A \ : \ \mathbb{R}^n \to 2^{\mathbb{R}^n}$



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

Graph

- the graph of an operator $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is defined as

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

- the graph consists of all pairs of points (x, y) such that $y \in Ax$
- $\mathrm{gph} A$ is a set, it is a subset of $\mathbb{R}^n \times \mathbb{R}^n$, i.e., $\mathrm{gph} A \subseteq \mathbb{R}^n \times \mathbb{R}^n$

Special operators

- the identity operator is denoted Id and is defined as

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

- inverse of an operator, defined through its graph:

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

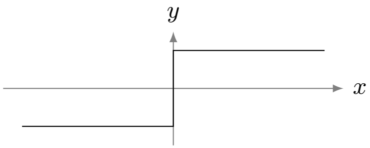
(therefore $y \in Ax$ if and only if $x \in A^{-1}y$)

Graphical representation – Inverse operators

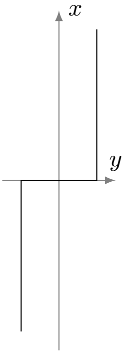
- we have the following equivalence

$$y \in Ax \quad \Leftrightarrow \quad x \in A^{-1}y$$

- therefore $y \in A(A^{-1}y)$ and $x \in A^{-1}(Ax)$
- A and A^{-1} are each others images under mapping $(x, y) \mapsto (y, x)$
- example: A in figure, draw A^{-1}



$y \in Ax$



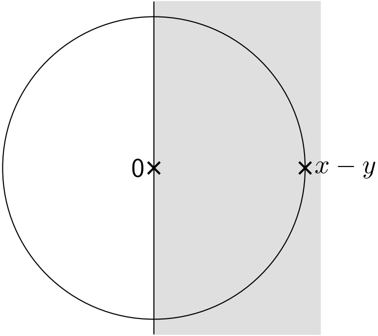
$x \in A^{-1}y$

Monotone operators

- an operator $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is monotone if

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

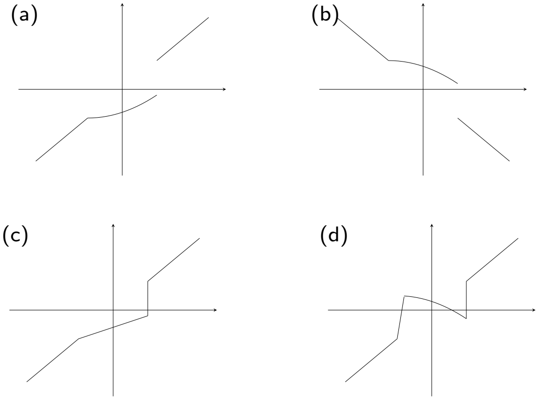
  for all $(x, u) \in \mathrm{gph}A$ and $(y, v) \in \mathrm{gph}A$
- graphical representation



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

Monotonicity 1D

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



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

Examples of monotone mappings

- the subdifferential ∂f of $f \ : \ \mathbb{R}^n \to \overline{\mathbb{R}}$
- proof: let $u \in \partial f(x)$ and $v \in \partial f(y)$ and subdifferential definitions

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

to get that

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

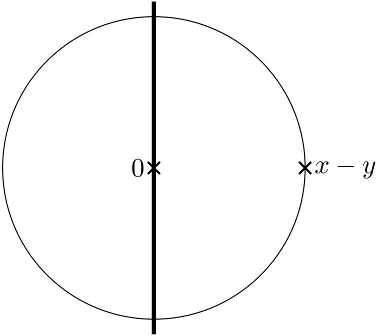
holds for all $(x, u), (y, v) \in \text{gph}\partial f$

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

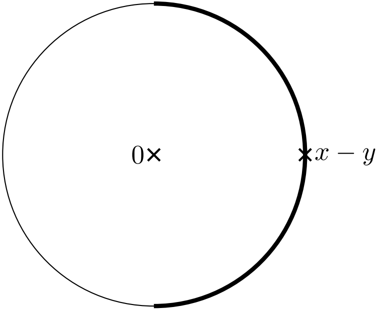
Examples of monotone mappings

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

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

$$= \left\langle \begin{bmatrix} z_1\cos\theta - z_2\sin\theta \\ z_1\sin\theta + z_2\cos\theta \end{bmatrix}, z \right\rangle = z_1^2\cos\theta + z_2^2\cos\theta \ge 0$$

- graphical representation



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

Maximal monotonicity

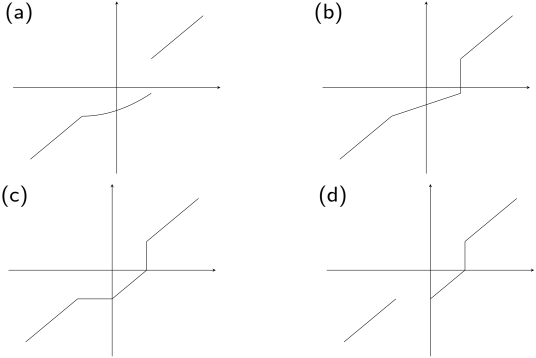
- let $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ be monotone
- A is maximal monotone if no pair $(\bar{x}, \bar{u}) \notin$ gph$A$ exists such that

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

  for all $(x, u) \in$ gph$A$
- equivalently: no monotone operator B exists with gph$A \subset$ gph$B$ (strict subset)

Graphical interpretation

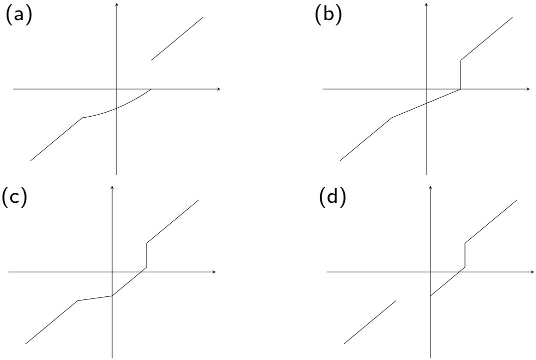
- which of the following $A : \mathbb{R} \to 2^{\mathbb{R}}$ are maximal monotone?



- (b) and (c) are maximally monotone

Minty's theorem

- let $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ be monotone
- A is maximal monotone iff ran$(A + \alpha \mathrm{Id}) = \mathbb{R}^n$ with $\alpha > 0$
- shifted previous figures with $0.2\mathrm{Id}$ (and re-scaled):



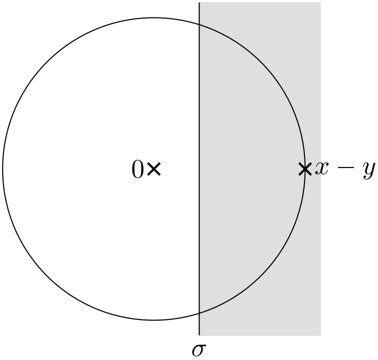
- "holes" in horizontal regions give holes in range due to $+\alpha \mathrm{Id}$
- "holes" in nonhorizontal regions give holes in range due to $+\alpha \mathrm{Id}$

Strongly monotone operators

- an operator A is σ-strongly monotone if

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

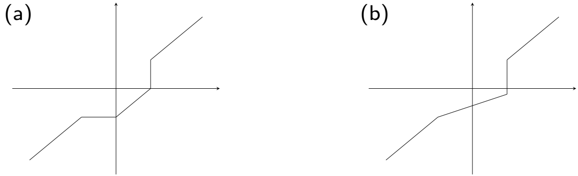
  for all $(x, u) \in \mathrm{gph}A$ and $(y, v) \in \mathrm{gph}A$
- 2D-graphical representation



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

1D Graphical interpretation

- strong monotonicity $\langle u - v, x - y \rangle \geq \sigma \|x - y\|^2$ $(\sigma > 0)$
- which of the following are strongly monotone?



- (b): $((u - v) \geq \sigma(x - y))$
- that is, slope is at least σ

Examples of strongly monotone operators

- assume that f is proper closed and σ-strongly convex
- then ∂f is σ-strongly monotone
- proof:
  - assumption equivalent to that $g = f - \frac{\sigma}{2}\|\cdot\|^2$ is convex
  - therefore $f = g + \frac{\sigma}{2}\|\cdot\|^2$
  - since g convex, $\partial f = \partial g + \sigma \mathrm{Id}$ and $\partial g(x) = \partial f(x) - \sigma x$
  - therefore, subgradients of g satisfy

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

where $u \in \partial f(x)$ and $v \in \partial f(y)$

  - add to get

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

and rearrange to get

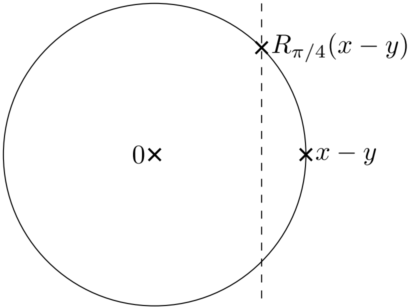
$$\langle u - v, x - y \rangle \geq \sigma \|x - y\|^2$$

Examples of strongly monotone operators

- 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 ($\theta = \frac{\pi}{4}$)



Single-valued operators

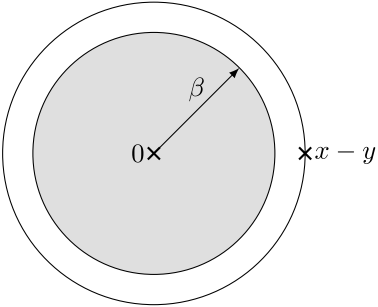
- so far, have considered set-valued operators $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$
    - monotonicity
    - maximal monotonicity
    - strong monotonicity
- now, we will consider single-valued operators $T : \mathcal{D} \to \mathbb{R}^n$
- we assume that $\mathcal{D} \subseteq \mathbb{R}^n$ is nonempty
- if $\mathcal{D} = \mathbb{R}^n$, then T has full domain
- a fixed-point y to the operator $T : \mathbb{R}^n \to \mathbb{R}^n$ satisfies $y = Ty$
- the set of fixed-points to $T : \mathbb{R}^n \to \mathbb{R}^n$ is denoted $\mathrm{fix}T$

Lipschitz continuous operator

- an operator $T : \mathcal{D} \to \mathbb{R}^n$ is β-Lipschitz continuous if

$$\|Tx - Ty\| \leq \beta\|x - y\|$$

  holds for all $x, y \in \mathcal{D}$
- T is single-valued (show by letting $y = x$ and use contradiction)
- graphical representation

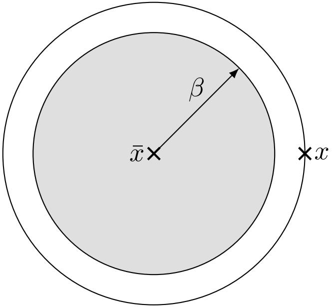


then $Tx - Ty$ is in gray area

Alternative graphical representation

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

$$\|Tx - \bar{x}\| = \|Tx - T\bar{x}\| \leq \beta \|x - \bar{x}\|$$



  then Tx in gray area

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

Examples of Lipschitz continuous mappings

- a rotation is 1-Lipschitz continuous (nonexpansive)
- a linear mapping $T : \mathcal{D} \to \mathbb{R}^n$ is $\|T\|$-Lipschitz continuous since

$$\|Tx - Ty\| = \|T(x - y)\| \leq \|T\|\|x - y\|$$

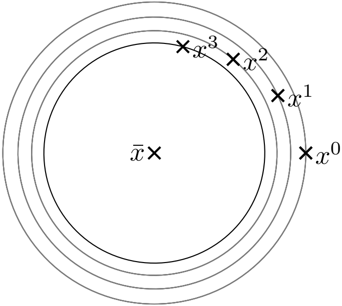
  by Cauchy-Schwarz inequality
- compositions: assume that $T_1, T_2 : \mathcal{D} \to \mathbb{R}^n$ are β_1, β_2-Lipschitz, then $T_1 T_2$ is $\beta_1 \beta_2$-Lipschitz

$$\|T_1 T_2 x - T_1 T_2 y\| \leq \beta_1 \|T_2 x - T_2 y\| \leq \beta_1 \beta_2 \|x - y\|$$

Iterating a contractive operator

- a contractive ($\beta < 1$) operator T has a unique fixed-point $\bar{x}$
- the iteration $x^{k+1} = Tx^k$ converges linearly to the fixed-point ($\bar{x}$):

$$\begin{aligned}\|x^{k+1} - \bar{x}\| &= \|Tx^k - T\bar{x}\| \leq \beta\|x^k - \bar{x}\| \\ &= \beta\|Tx^{k-1} - T\bar{x}\| \leq \cdots \leq \beta^{k+1}\|x^0 - \bar{x}\|\end{aligned}$$



Fixed-points of nonexpansive operator

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

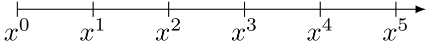
$$Tx = x + 2 = x$$

  does not hold for any $x \in \mathbb{R}$

- it is nonexpansive (1-Lipschitz continuous)

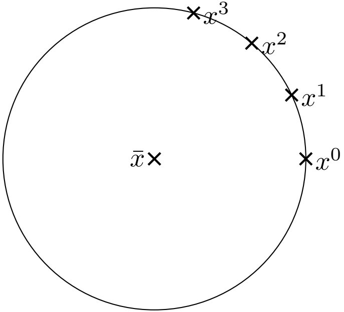
$$\|Tx - Ty\| = \|x + 2 - y - 2\| = \|x - y\|$$

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



Iteration of nonexpansive operator

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



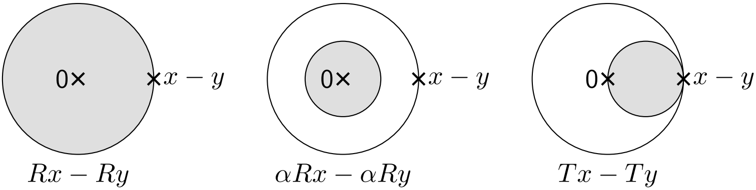
(however, the iterates are bounded)

Averaged operators

- let $\alpha \in (0,1)$ and $R : \mathcal{D} \to \mathbb{R}^n$ be some nonexpansive operator
- an operator $T : \mathcal{D} \to \mathbb{R}^n$ is α-averaged if:

$$T = (1-\alpha)\mathrm{Id} + \alpha R$$

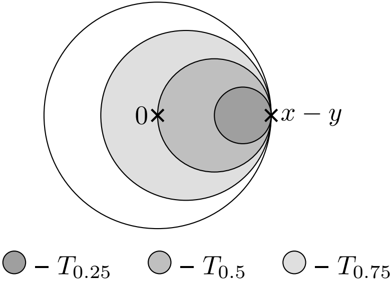
- graphical representation for $\alpha = \frac{1}{2}$:



- draw similar figures for $\alpha = 0.25$ and $\alpha = 0.75$

Averaged operators

- T_α is α-averaged with $\alpha = 0.25, 0.5, 0.75$
- graphical representation:



Fixed-points

- assume that fix$R$ is nonempty and that $\alpha \in (0,1)$
- the fixed-points of $T = (1-\alpha)\mathrm{Id} + \alpha R$ and R coincide
- proof:
  - a fixed point $\bar{x}$ to R is a fixed-point to T:

$$T\bar{x} = (1-\alpha)\bar{x} + \alpha R\bar{x} = (1-\alpha+\alpha)\bar{x} = \bar{x}$$

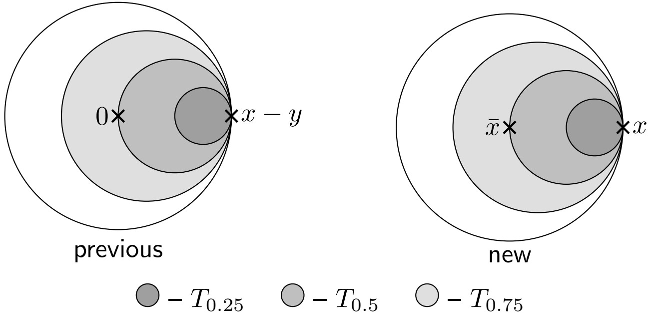
  - a fixed-point $\bar{x}$ to T is a fixed-point to R:

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

  (where $R = \frac{1}{\alpha}T - \frac{1-\alpha}{\alpha}\mathrm{Id}$ is used)

Additional graphical representation

- assume that T_α is α-averaged and that $\bar{x} \in \mathrm{fix} T_\alpha$
- then T_α for $\alpha = 0.25, 0.5, 0.75$ can be represented as:



- why?
  - figure on left holds for all y
  - let y be a fixed-point, i.e., $y = \bar{x}$
  - shift left figure by $y = \bar{x}$ to get right figure:
    ($0 \to \bar{x}$, $x - \bar{x} \to x$, $T_\alpha x - T_\alpha \bar{x} \to T_\alpha x - T_\alpha \bar{x} + \bar{x} = T_\alpha x$)
- for $x \notin \mathrm{fix} T_\alpha$, distance to fixed-point strictly decreased

Averaged operator formula

- let
  - $\alpha \in (0, 1)$
  - $R : \mathcal{D} \to \mathbb{R}^n$ be nonexpansive
  - $T = (1 - \alpha)\mathrm{Id} + \alpha R$
- the following are equivalent (show in exercise):
  - T is α-averaged
  - $(1 - 1/\alpha)\mathrm{Id} + \frac{1}{\alpha}T (= R)$ is nonexpansive
  - the following holds for all $x, y \in \mathcal{D}$

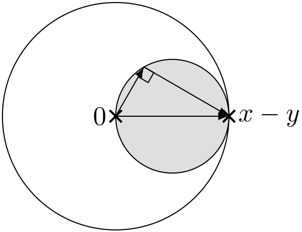
$$\tfrac{1-\alpha}{\alpha}\|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2 + \|Tx - Ty\|^2 \leq \|x - y\|^2$$

Averaged operator formula

- (previous slide) $T : \mathcal{D} \to \mathbb{R}^n$ is α-averaged iff for all $x, y \in \mathcal{D}$

$$\tfrac{1-\alpha}{\alpha}\|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2 + \|Tx - Ty\|^2 \le \|x - y\|^2$$

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



- $\frac{1}{2}$-averaged operators are also called *firmly nonexpansive*

Iterating averaged operators

- assume R is nonexpansive, want to find fixed-point $\bar{x} \in \text{fix}R \neq \emptyset$
- iterate the averaged map $T = (1-\alpha)\mathrm{Id} + \alpha R$ (α design param)
- the iteration $x^{k+1} = Tx^k$ converges to some $\bar{x} \in \text{fix}R = \text{fix}T$
- proof: note that

$$x^k - x^{k+1} = (\mathrm{Id} - T)x^k = (\mathrm{Id} - T)x^k - (\mathrm{Id} - T)\bar{x}$$

use α-averagedness formula with $x = x^k$ and $y = \bar{x}$:

$$\begin{aligned}\tfrac{1-\alpha}{\alpha}\|x^k - x^{k+1}\|^2 &= \tfrac{1-\alpha}{\alpha}\|(\mathrm{Id} - T)x^k - (\mathrm{Id} - T)\bar{x}\|^2 \\ &\leq \|x^k - \bar{x}\|^2 - \|Tx^k - T\bar{x}\|^2 \\ &= \|x^k - \bar{x}\|^2 - \|x^{k+1} - \bar{x}\|^2\end{aligned}$$

- what to do with this?

Iterating averaged operators cont'd

- multiply by $\frac{\alpha}{1-\alpha}$ and sum over $k = 0, 1, \ldots, n$:

$$\sum_{k=0}^{n} \|x^{k+1} - x^k\|^2 \leq \frac{\alpha}{1-\alpha} \sum_{k=0}^{n} \left(\|x^k - \bar{x}\|^2 - \|x^{k+1} - \bar{x}\|^2\right)$$
$$= \frac{\alpha\|x^0 - \bar{x}\|^2}{1-\alpha}$$

- since T is nonexpansive

$$\|x^{k+1} - x^k\| = \|Tx^k - Tx^{k-1}\| \leq \|x^k - x^{k-1}\|$$

i.e.

$$(n+1)\|x^{n+1} - x^n\|^2 \leq \sum_{k=0}^{n} \|x^{k+1} - x^k\|^2 \leq \frac{\alpha\|x^0 - \bar{x}\|^2}{(1-\alpha)}$$

or

$$\|x^{n+1} - x^n\|^2 \leq \sum_{k=0}^{n} \|x^{k+1} - x^k\|^2 \leq \frac{\alpha\|x^0 - \bar{x}\|^2}{(n+1)(1-\alpha)}$$

- not very informative since might not want $\|x^{n+1} - x^n\|$ small (compare to algorithm $x^{k+1} = x^k$)

Iterating averaged operators cont'd

- current distance to fixed-point of R is $\|Rx^n - x^n\|$
- can we bound this?
- yes, proof: (remember $T = (1-\alpha) + \alpha R$)

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

- therefore

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

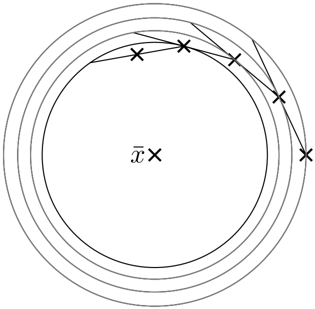
- that is $\|Rx^n - x^n\| \to 0$ (i.e., approach fixed-point) as $n \to \infty$
- optimal $\alpha = \frac{1}{2}$:

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

sublinear convergence

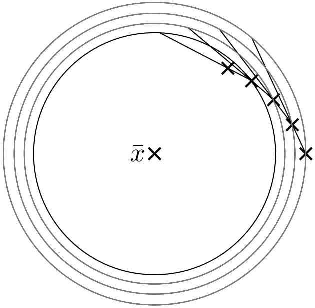
Iteration example - $\alpha = 0.5$

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



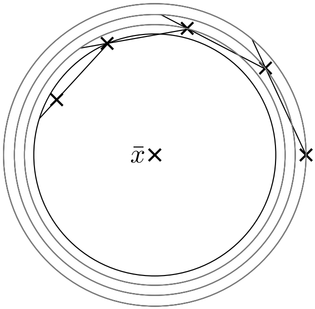
Iteration example - $\alpha = 0.25$

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



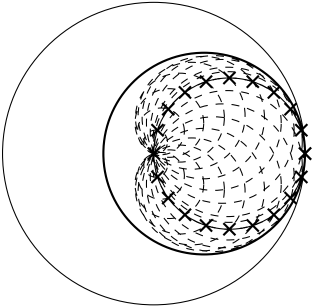
Iteration example - $\alpha = 0.75$

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



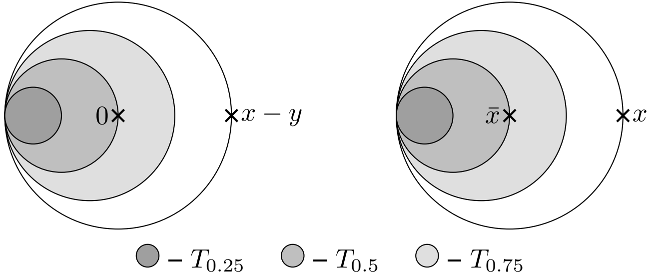
Composition of averaged operators

- composition of averaged opertors is averaged
- assume that T_1 is α_1-averaged and T_2 is α_2-averaged, $\alpha_i \in (0, 1)$
- then $T_1 T_2$ is $\frac{\alpha}{\alpha+1}$-averaged with $\alpha = \frac{\alpha_1}{1-\alpha_1} + \frac{\alpha_2}{1-\alpha_2}$
- example $\alpha_1 = \alpha_2 = 0.5 \Rightarrow T_1 T_2$ is $\frac{2}{3}$-averaged



Negatively averaged operators

- let $T : \mathcal{D} \to \mathbb{R}^n$ and $\alpha \in (0,1)$
- then T is α-negatively averaged if $-T$ is averaged
- T_α are α-negatively averaged, $\alpha = 0.25, 0.5, 0.75$, $\bar{x} \in$ fixT_α
- then T_α for $\alpha = 0.25, 0.5, 0.75$ can be represented as:



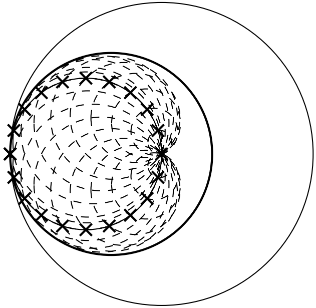
- averaged map of negatively averated operator

$$(1-\beta)\mathrm{Id} + \beta T_\alpha$$

is contractive (prove in exercise)

Composition of (negatively) averaged operators

- assume that $\alpha_1 \in (0,1)$ and $\alpha_2 \in (0,1)$
- assume that T_1 is α_1-negatively averaged and T_2 is α_2-averaged
- then $T_1 T_2$ is $\frac{\alpha}{\alpha+1}$-negatively averaged with $\alpha = \frac{\alpha_1}{1-\alpha_1} + \frac{\alpha_2}{1-\alpha_2}$
- example $\alpha_1 = \alpha_2 = 0.5 \Rightarrow T_1 T_2$ is $\frac{2}{3}$-negatively averaged



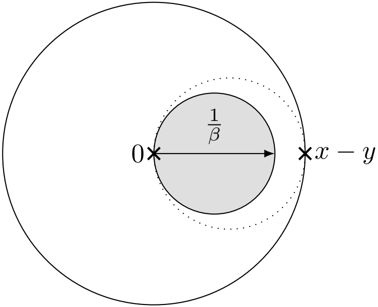
- what happens if T_1 and T_2 are negatively averaged?

Devise optimization algorithms

- look for mappings:
  - that are nonexpansive, averaged, or contractive
  - whose fixed-points can be used to solve optimization problem
- we know from previous disussion that we get:
  - linear convergence for contractive mappings
  - sublinear convergence for averaged mappings
  - sublinear convergence for nonexpansive mappings by iterating averaged map
- almost all algorithms in course boil down to this!

Cocoercive operators

- assume that $T : \mathcal{D} \to \mathbb{R}^n$
- T is β-cocoercive if βT is $\frac{1}{2}$-averaged
- draw a graphical representation in 2D?:



- $Tx - Ty$ in gray area
- (dotted area shows that βT is $\frac{1}{2}$-averaged, or firmly nonexpansive)

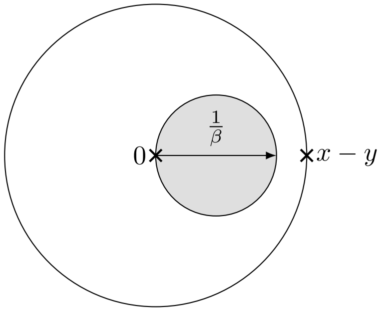
Cocoercive operator properties

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

$$\|(I - \beta T)x - (I - \beta T)y\|^2 + \|\beta Tx - \beta Ty\|^2 \leq \|x - y\|^2$$

- equivalently (by expanding the first square and div. by 2β)

$$\langle Tx - Ty, x - y\rangle \geq \beta\|Tx - Ty\|^2$$

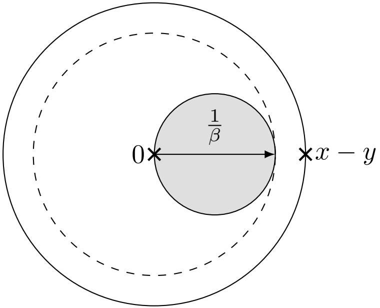


Properties

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

$$\beta\|Tx - Ty\|^2 \leq \langle Tx - Ty, x - y\rangle \leq \|x - y\|\|Tx - Ty\|$$

(then divide by $\beta\|Tx - Ty\|$)



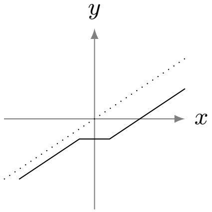
Graphical representation in 1D

- β-cocoercivity of T

$$\langle Tx - Ty, x - y \rangle \geq \beta \|Tx - Ty\|^2$$

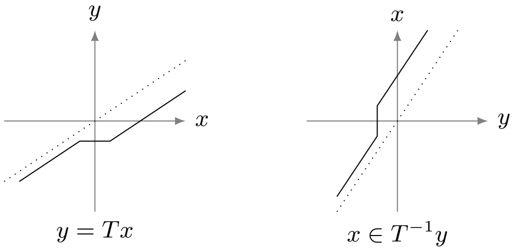
- what are bounds on slope in 1D?

$$(Tx - Ty)(x - y) \geq 0 \qquad \text{(nonnegative slope)}$$
$$|Tx - Ty| \leq \frac{1}{\beta}|x - y| \qquad \text{slope less than } \frac{1}{\beta}$$



$y = Tx$

Inverse strongly monotone



Relationship:

- maximum slope $\frac{1}{\beta}$ of $T \Leftrightarrow$ minimum slope β of T^{-1}
- nonnegative slope of $T \Leftrightarrow$ "at most" vertical slope of T^{-1}
- we have β-cocoercivity of $T \Leftrightarrow \beta$-strong monotonicity of T^{-1}

Inverse strong monotonicity

- proof:
  - β-cocoercivity:

$$\langle Tx - Ty, x - y\rangle \geq \beta \|Tx - Ty\|^2$$

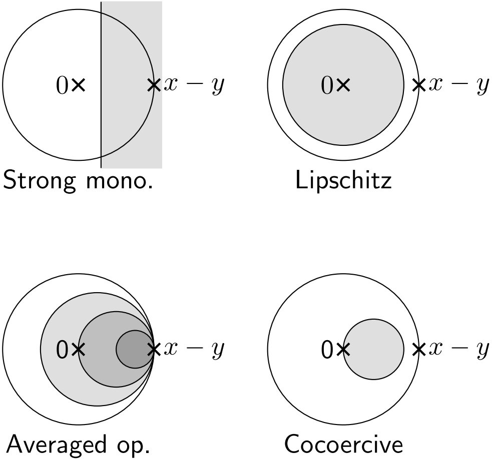
  - inverse: $u = Tx$ and $v = Ty$ iff $x \in T^{-1}u$ and $y \in T^{-1}u$:

$$\langle u - v, T^{-1}u - T^{-1}v\rangle \geq \beta \|u - v\|^2$$

  - i.e., T is β-cocoercive iff T^{-1} is β-strongly monotone
- sometimes β-cocoercivity is called β-inverse strong monotonicity

Summary

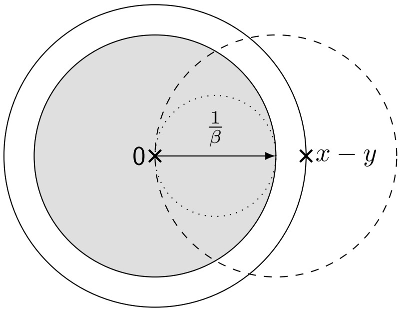
- we have discussed operators T with the following properties



- the set (or point) $Tx - Ty$ is in the respective gray areas

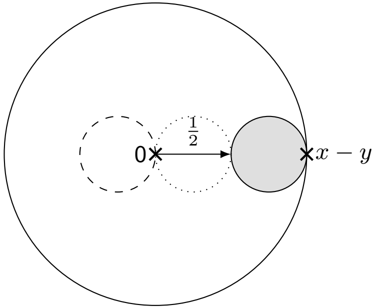
Exercise I

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



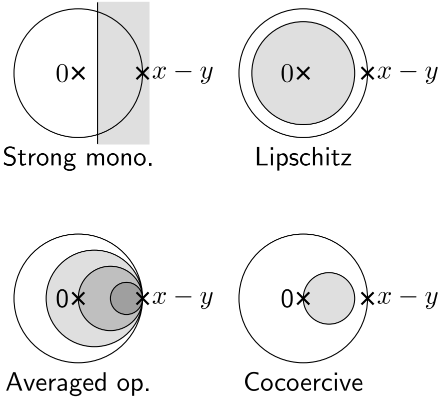
Exercise II

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



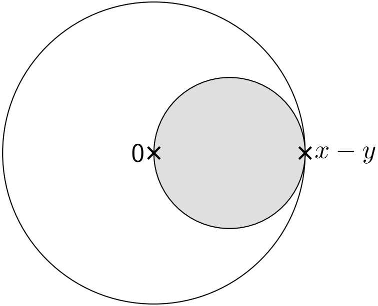
Relation to (strong) monotonicity?

- can relate Lipschitz continuity, cocoercivity, and averagedness by scaling and shifting (they are all circles)
- cannot directly relate to (strong) monotonicity
- since β-cocoercivity is β-inverse strong monotonicity, can relate to strong monotonicity via inverse



Exercise III

- T^{-1} is 1-strongly monotone
- T is α-averaged, compute α
- T is $\frac{1}{2}$-averaged
  "proof":
  1. since T^{-1} is 1-inverse strongly monotone, T is 1-cocoercive ($Tx - Ty$ 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