ORF522 – Linear and Nonlinear Optimization

14. Operator theory I

Bartolomeo Stellato — Fall 2024

Today's lecture
[LSMO][Chapter 4, FMO][PA]

Operator theory I

- Operators
- Monotone operators
- Fixed-point Iterations

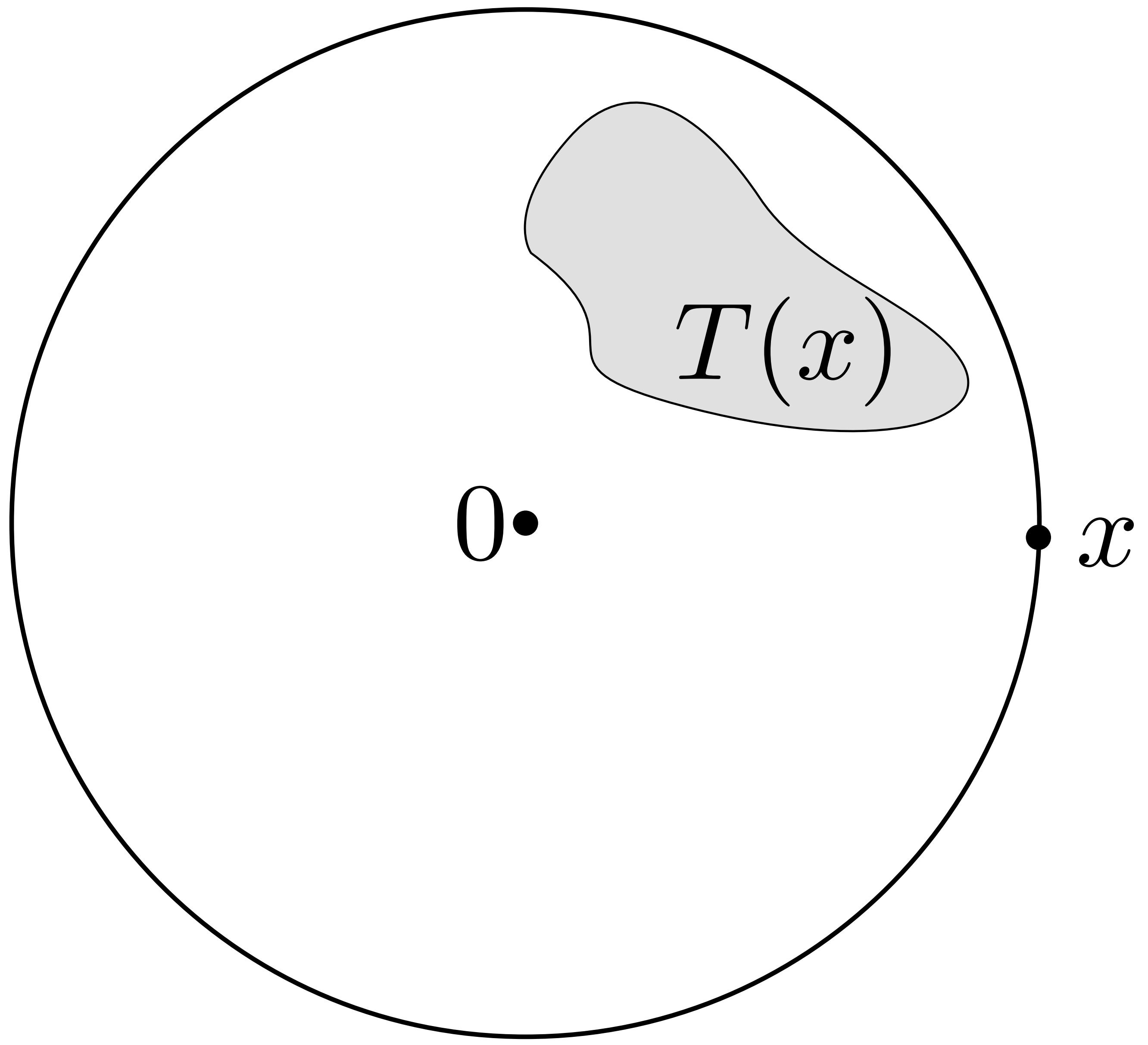
Operators

Operators

An operator T maps each point in $\mathbf{R}^n$ to a subset of $\mathbf{R}^n$

- **set valued** $T(x)$ returns a set
- **single-valued** $T(x)$ (function) returns a singleton

The **domain** of T is the set $\mathbf{dom}\, T = \{x \mid T(x) \neq \emptyset\}$



Example

- The subdifferential ∂f is a set-valued operator
- The gradient ∇f is a single-valued operator

Graph and inverse operators

Graph

The graph of an operator T is defined as

$$\mathbf{gph} T = \{(x, y) \mid y \in T(x)\}$$

In other words, all the pairs of points (x, y) such that $y \in T(x)$.

Inverse

The graph of the inverse operator T^{-1} is defined as

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

Therefore, $y \in T(x)$ if and only if $x \in T^{-1}(y)$.

Zeros

Zero

x is a **zero** of T if $\quad 0 \in T(x)$

Zero set

The set of all the zeros $\quad T^{-1}(0) = \{x \mid 0 \in T(x)\}$

Example
If $T = \partial f$ and $f : \mathbf{R}^n \to \mathbf{R}$, then
$0 \in T(x)$ means that x minimizes f

Many problems
can be posed as finding zeros
of an operator

Fixed points

$\bar{x}$ is a **fixed-point** of a single-valued operator T if

$$\bar{x} = T(\bar{x})$$

Set of fixed points $\quad \mathbf{fix}\, T = \{x \in \mathbf{dom}\, T \mid x = T(x)\} = (I - T)^{-1}(0)$

Examples

- **Identity** $T(x) = x$. Any point is a fixed point
- **Zero operator** $T(x) = 0$. Only 0 is a fixed point

Lipschitz operators

An operator T is L-Lipschitz if

$$\|T(x) - T(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathbf{dom}\, T$$

Fact If T is Lipschitz, then it is single-valued

Proof If $y \in T(x), z \in T(x)$, then $\|y - z\| \leq L\|x - x\| = 0 \Longrightarrow y = z$ ■

For $L = 1$ we say T is **nonexpansive**

For $L < 1$ we say T is **contractive** (with contraction factor L)

Lipschitz operators examples

Lipschitz affine functions

$$T(x) = Ax + b$$

$\longleftrightarrow$

maximum singular value

$$L = \|A\|_2 = \sqrt{\lambda_{\max}(A^T A)}$$

Lipschitz differentiable functions

T such that there exists derivative DT

$\longleftrightarrow$

derivative is bounded

$$\|DT\|_2 \leq L$$

Lipschitz operators and fixed points

Given a L-Lipschitz operator T and a fixed point $\bar{x} = T\bar{x}$,

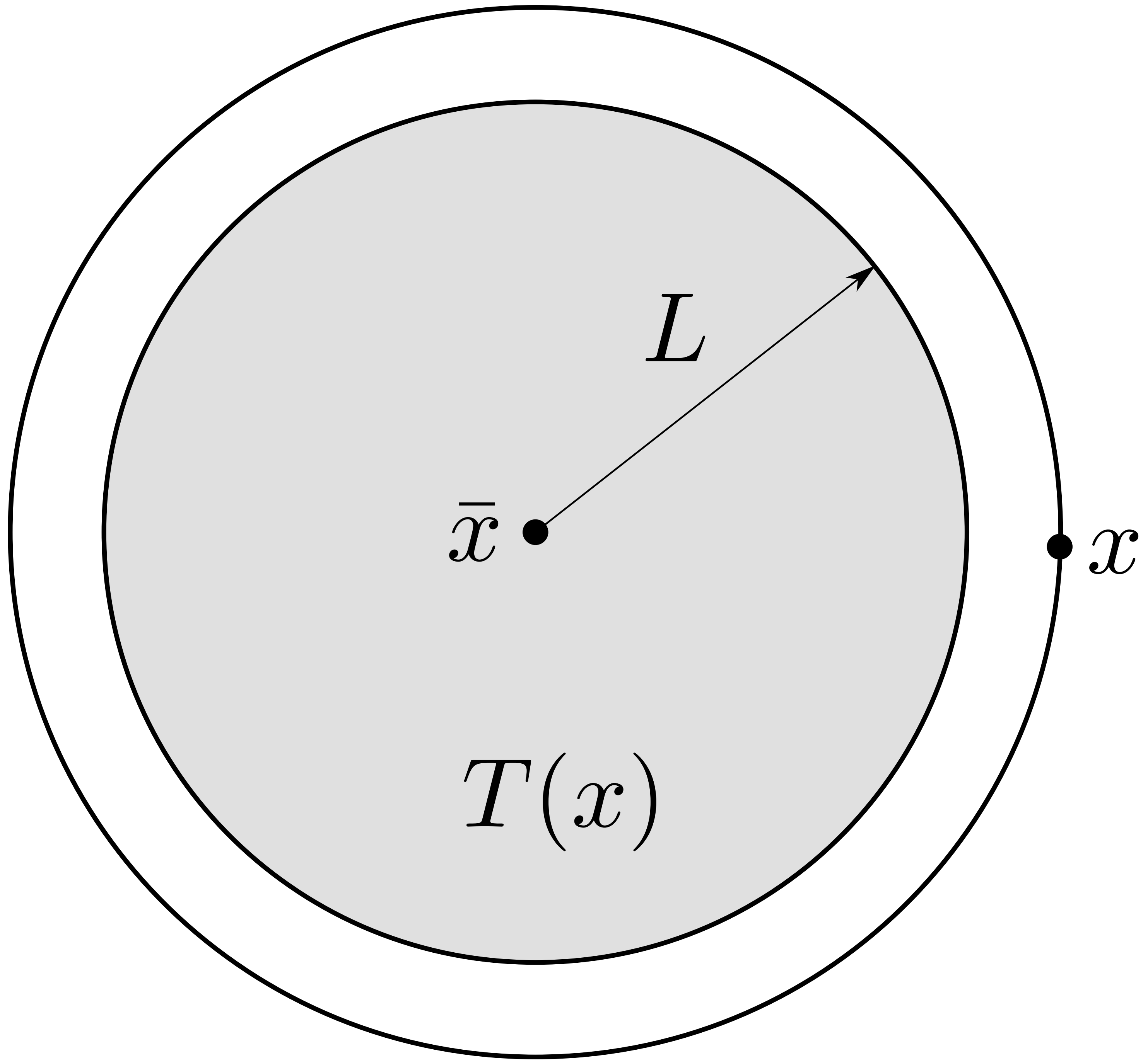
$$\|Tx - \bar{x}\| = \|Tx - T\bar{x}\| \leq L\|x - \bar{x}\|$$

A contractive operator ($L < 1$) can have at most one fixed point, i.e., $\mathbf{fix}\, T = \{\bar{x}\}$

Proof

If $\bar{x}, \bar{y} \in \mathbf{fix}\, T$ and $\bar{x} \neq \bar{y}$ then

$\|\bar{x} - \bar{y}\| = \|T(\bar{x}) - T(\bar{y})\| < \|\bar{x} - \bar{y}\|$ (contradiction) ■



A nonexpansive operator ($L = 1$) need not have a fixed point

Example $T(x) = x + 2$

Combining Lipschitz operators

T_1 is L_1-Lipschitz and T_2 is L_2-Lipschitz

The **composition** $T_1 T_2$ is $L_1 L_2$-Lipschitz

Proof $\|T_1 T_2 x - T_1 T_2 y\|_2 \leq L_1 \|T_2 x - T_2 y\|_2 \leq L_1 L_2 \|x - y\|_2$ $\blacksquare$

- Composition of *nonexpansive* is nonexpansive
- Composition of *nonexpansive* and *contractive* is contractive

The **weighted average** $\theta T_1 + (1-\theta) T_2,\ \theta \in (0, 1)$ is $(\theta L_1 + (1-\theta) L_2)$-Lipschitz

Proof (exercise)

- Weighted average of *nonexpansive* is nonexpansive
- Weighted average of *nonexpansive* and *contractive* is contractive

Monotone cocoercive operators

Monotone operators

An operator T on $\mathbf{R}^n$ is **monotone** if

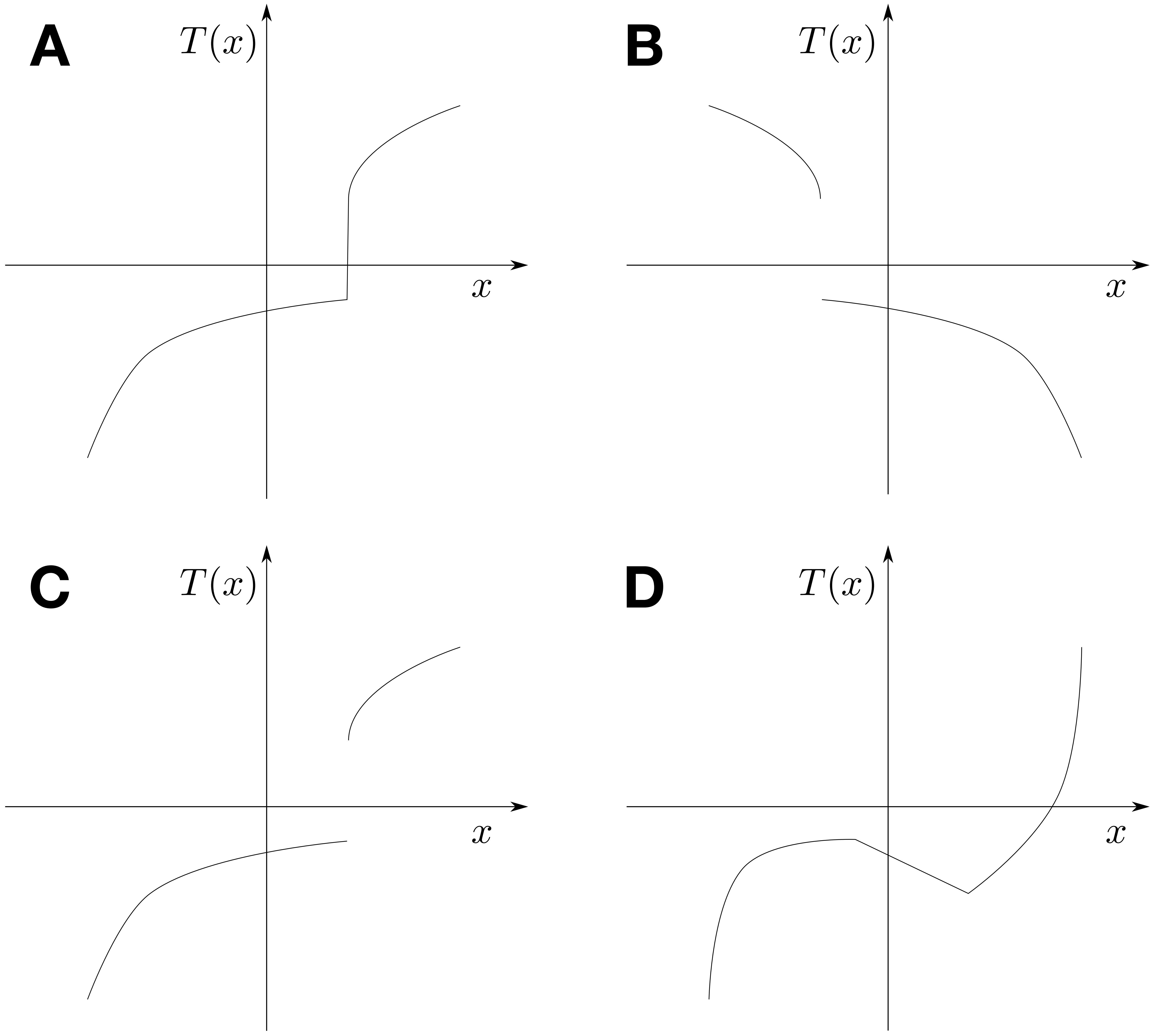
$$(u-v)^T(x-y) \geq 0, \quad \forall (x,u),(y,v) \in \mathbf{gph} T$$

T is **maximal monotone** if
$\nexists (\bar{x},\bar{u}) \notin \mathbf{gph} T$ such that

$$(\bar{u}-u)^T(\bar{x}-x) \geq 0, \quad \forall (x,u) \in \mathbf{gph} T$$

Equivalently: $\nexists$ monotone R
such that $\mathbf{gph} T \subset \mathbf{gph} R$

Monotone operators in 1D



Let's fill the table

	Monotone	Max Monotone
A		
B		
C		
D		

Monotonicity

$$y > x \quad \Rightarrow \quad T(y) \geq T(x)$$

Continuity

If T single-valued, continuous and monotone, then it's maximal monotone

Monotone operator properties

- **sum** $T + R$ is monotone
- **nonnegative scaling** αT with $\alpha \geq 0$ is monotone
- **inverse** T^{-1} is monotone
- **congruence** for $M \in \mathbf{R}^{n \times m}$, then $M^T T(Mz)$ is monotone on $\mathbf{R}^m$

Affine function $T(x) = Ax + b$ is maximal monotone

$$\iff \quad A + A^T \succeq 0$$

Strongly monotone operators

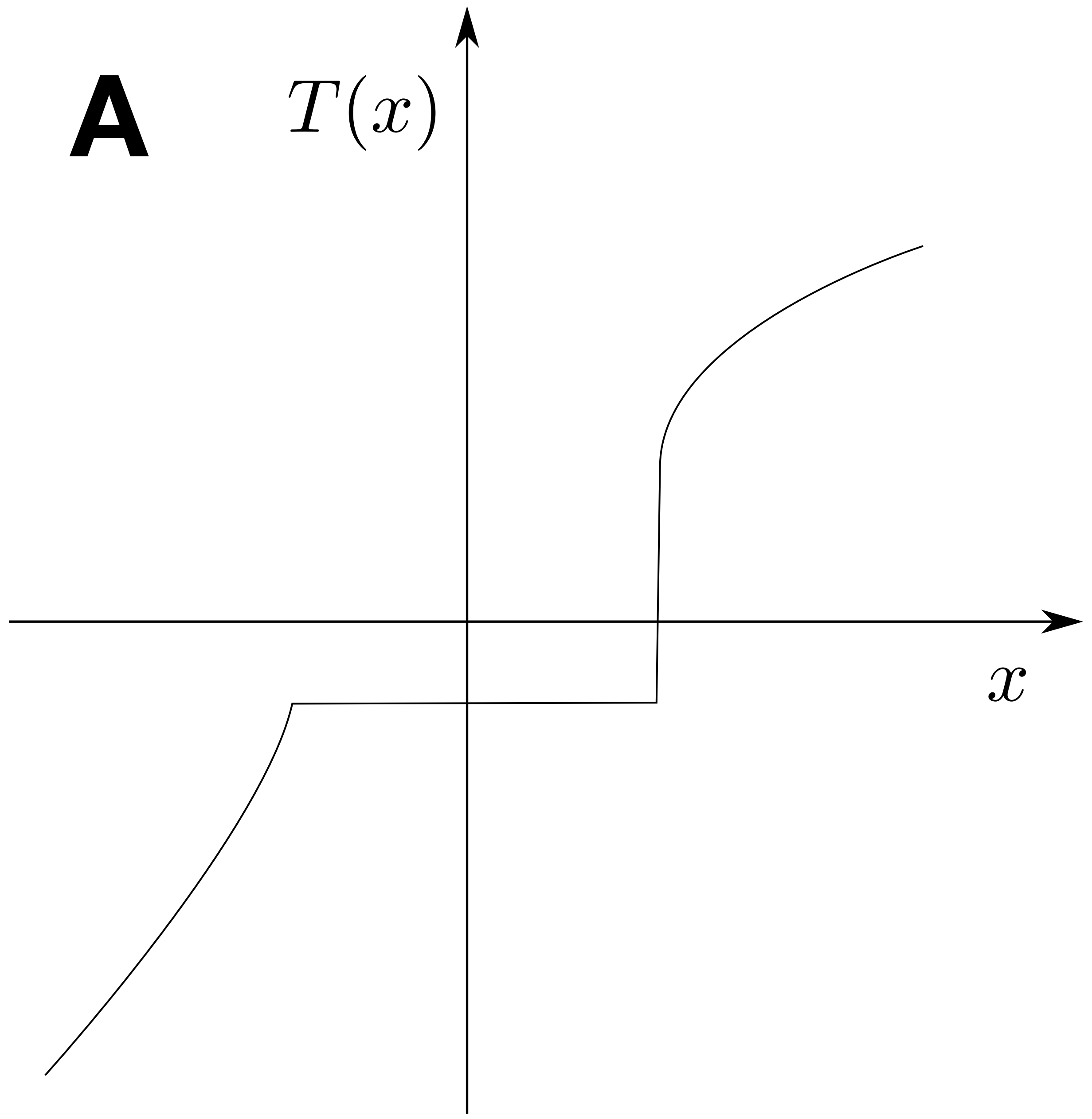
An operator T on $\mathbf{R}^n$ is μ-**strongly monotone** if

$$(u-v)^T(x-y) \geq \mu\|x-y\|^2, \quad \mu > 0$$

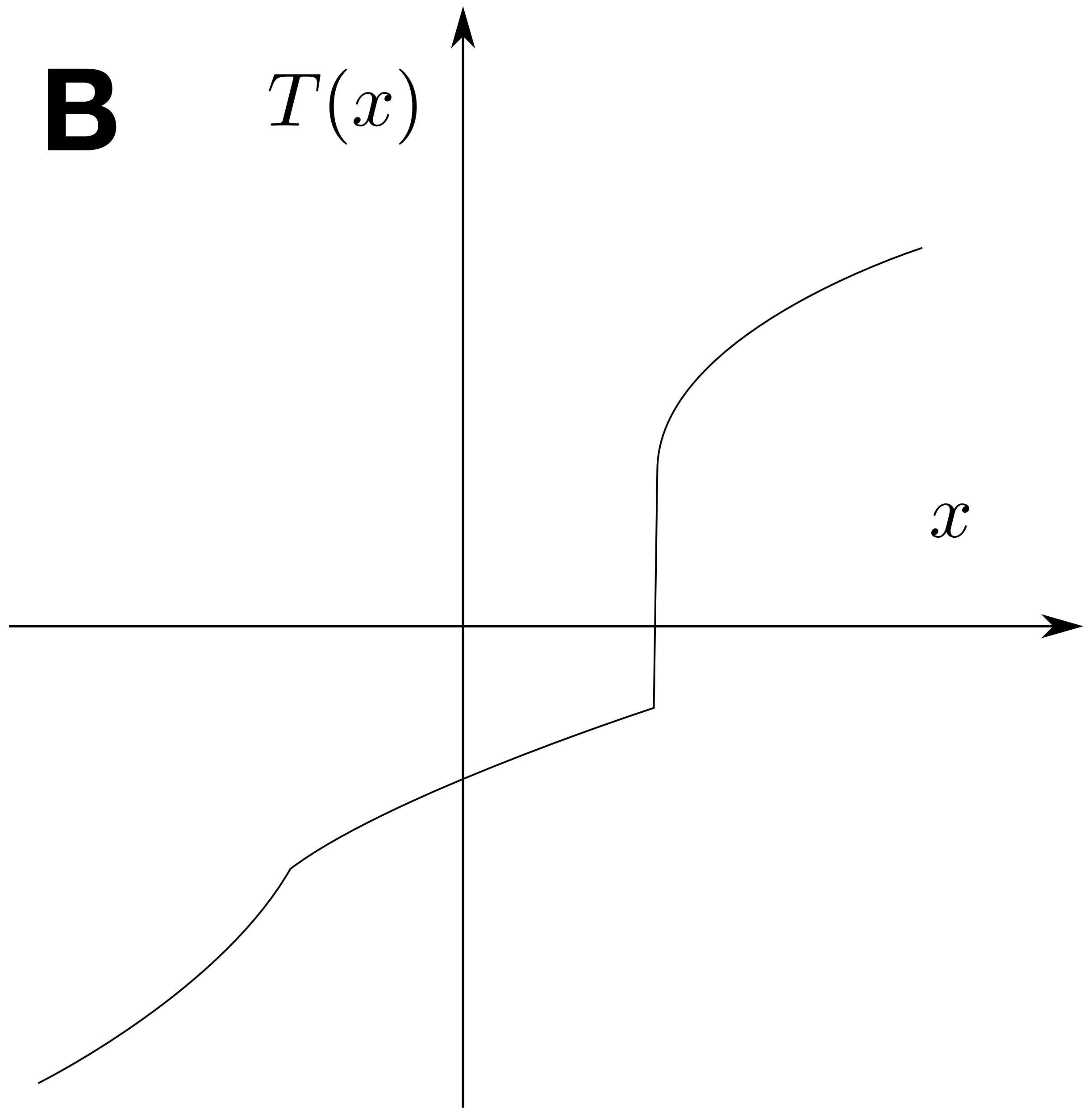
(also called μ-**coercive**)

$\forall (x,u),(y,v) \in \mathbf{gph} T$

A



B



Let's fill the table

	Monotone	Strongly Monotone
A		
B		

The slope is at least μ

Cocoercive operators

An operator T is β**-cocoercive**, $\beta > 0$, if

$$(T(x) - T(y))^T(x - y) \geq \beta\|T(x) - T(y)\|^2$$

If T is β**-cocoercive**, then T is $(1/\beta)$**-Lipschitz**

Proof $\beta\|T(x) - T(y)\|^2 \leq (T(x) - T(y))^T(x - y) \leq \|T(x) - T(y)\|\|x - y\|$

$$\implies \|T(x) - T(y)\| \leq (1/\beta)\|x - y\| \qquad \blacksquare$$

If T is μ**-strongly monotone** if and only if T^{-1} is μ**-cocoercive**

Proof $(T(x) - T(x))^T(x - y) \geq \mu\|x - y\|^2$

Inverse: $u = T(x)$ and $v = T(y)$ if and only if $x \in T^{-1}(u)$ and $y \in T^{-1}(v)$

$$(u - v)^T(T^{-1}(u) - T^{-1}(v)) \geq \mu\|T^{-1}(u) - T^{-1}(v)\|^2 \qquad \blacksquare$$

Cocoercive and nonexpansive operators

If T is β-**cocoercive** if and only if $I - 2\beta T$ is **nonexpansive**

Proof $\|(I{-}2\beta T)(y) - (I - 2\beta T)(x)\|^2 =$

$$= \|y - 2\beta T(y) - x + 2\beta T(x)\|^2$$

$$= \|y - x\|^2 - 4\beta(T(y) - T(x))^T(y - x) + 4\beta^2\|T(y) - T(x)\|^2$$

$$= |y - x\|^2 - 4\beta\left((T(y) - T(x))^T(y - x) - \beta\|T(y) - T(x)\|^2\right)$$

$$\leq \|y - x\|^2 \quad \blacksquare$$

(cocoercive)

Summary of monotone and cocoercive operators

Monotone

$$(T(x) - T(y))^T (x - y) \geq 0$$

↑ $\mu = 0$

Strongly monotone

$$(T(x) - T(y))^T (x - y) \geq \mu \|x - y\|^2$$

↔ $F = T^{-1}$

Lipschitz

$$\|F(x) - F(y)\| \leq L \|x - y\|$$

↑ $L = 1/\mu$

Cocoercive

$$(F(x) - F(y))^T (x - y) \geq \mu \|F(x) - F(y)\|^2$$

↕ $G = I - 2\mu F$

Nonexpansive

$$\|G(x) - G(y)\| \leq \|x - y\|$$

Fixed point iterations

Fixed point iteration

Apply operator

$$x^{k+1} = T(x^k)$$

until you reach $\bar{x} \in \mathbf{fix}\, T$

Main approach

1. Find a suitable T such that $\bar{x} \in \mathbf{fix}\, T$ solve your problem
2. Show that the fixed point iteration converges

Fixed point residual to terminate

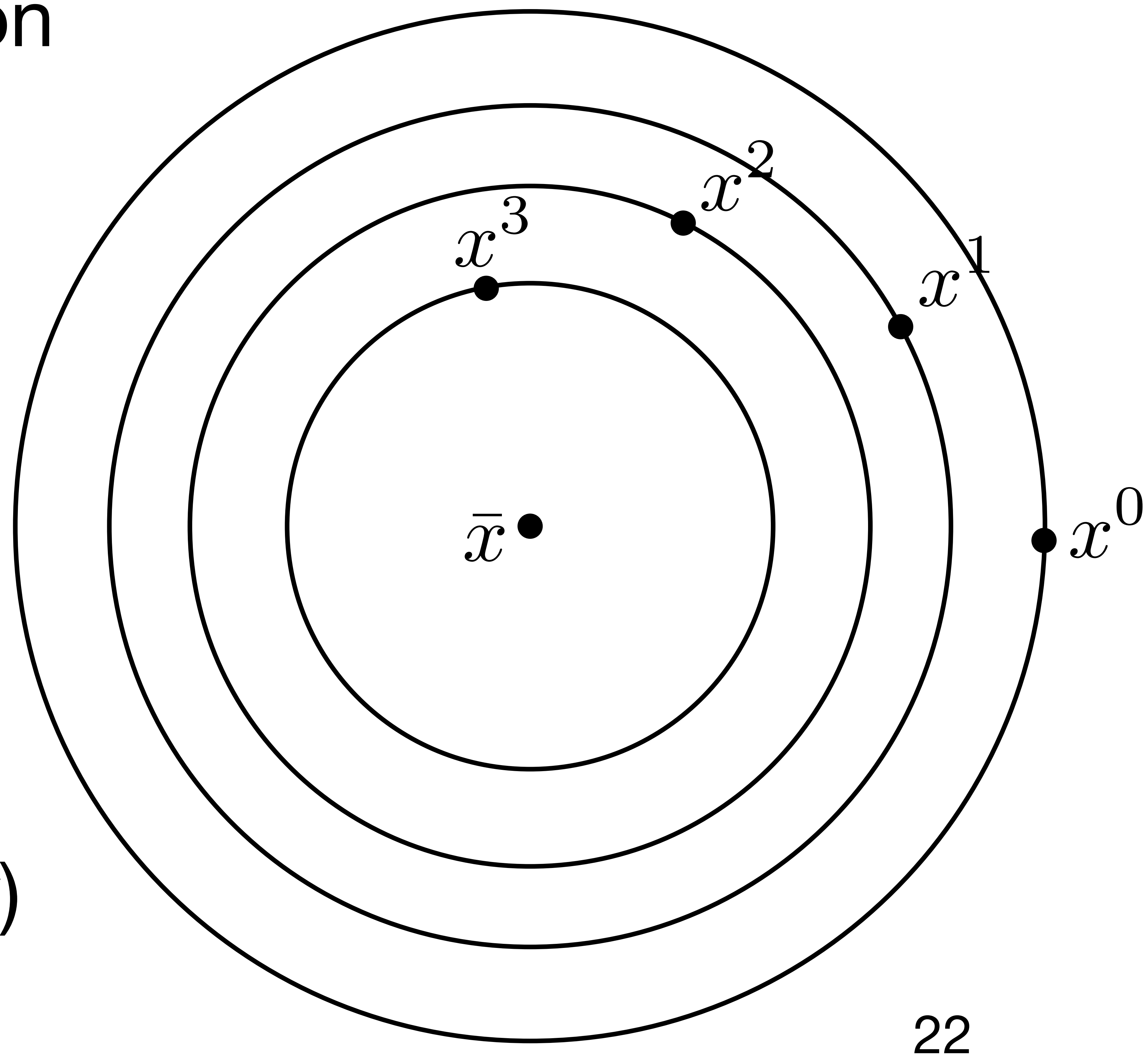
$$r^k = T(x^k) - x^k$$

Contractive fixed point iterations

Contraction mapping theorem
If T is L-Lipschitz with $L < 1$ (contraction), the iteration

$$x^{k+1} = T(x^k)$$

converges to $\bar{x}$, the unique fixed point of T



Properties

- Distance to $\bar{x}$ decreases at each step

$$\|x^{k+1} - \bar{x}\| \leq L\|x^k - \bar{x}\|$$

  (if it does not increase, we have **Fejer monotonicity**)

- Linear convergence rate L

Contraction mapping theorem

Proof

The sequence x^k is Cauchy

$$
\begin{aligned}
\|x^{k+\ell} - x^k\| &\leq \|x^{k+\ell} - x^{k+\ell-1}\| + \cdots + \|x^{k+1} - x^k\| \\
&\leq (L^{\ell-1} + \cdots + 1)\|x^{k+1} - x^k\| && \text{(Lipschitz constant)} \\
&\leq \frac{1}{1-L}\|x^{k+1} - x^k\| && \text{(geometric series)} \\
&\leq \frac{L^k}{1-L}\|x^1 - x^0\| && \text{(Lipschitz constant)}
\end{aligned}
$$

Therefore it converges to a point $\bar{x}$ which must be the (unique) fixed point of T

The convergence is linear (geometric) with rate L

$$\|x^k - \bar{x}\| = \|T(x^{k-1}) - T(\bar{x})\| \leq L\|x^{k-1} - \bar{x}\| \leq L^k\|x^0 - \bar{x}\| \quad \blacksquare$$

Nonexpansive fixed point iterations

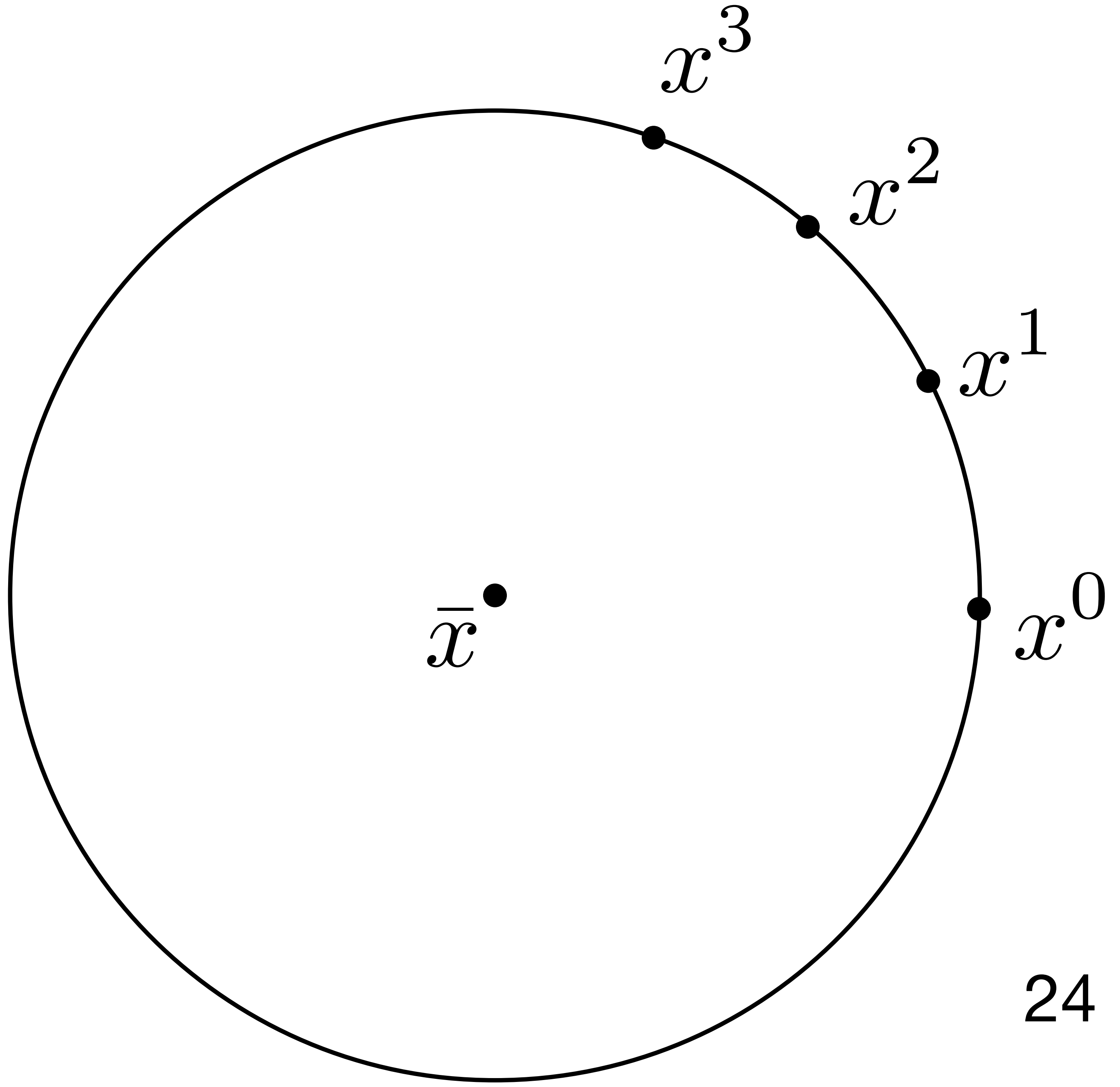
If T is L-Lipschitz with $L = 1$ (nonexpansive), the iteration

$$x^{k+1} = T(x^k)$$

need not converge to a fixed point, even if one exists.

Example

- Let T be a rotation around the origin
- T is nonexpansive and has a fixed point $\bar{x} = 0$
- $\|x^k\|$ never decreases

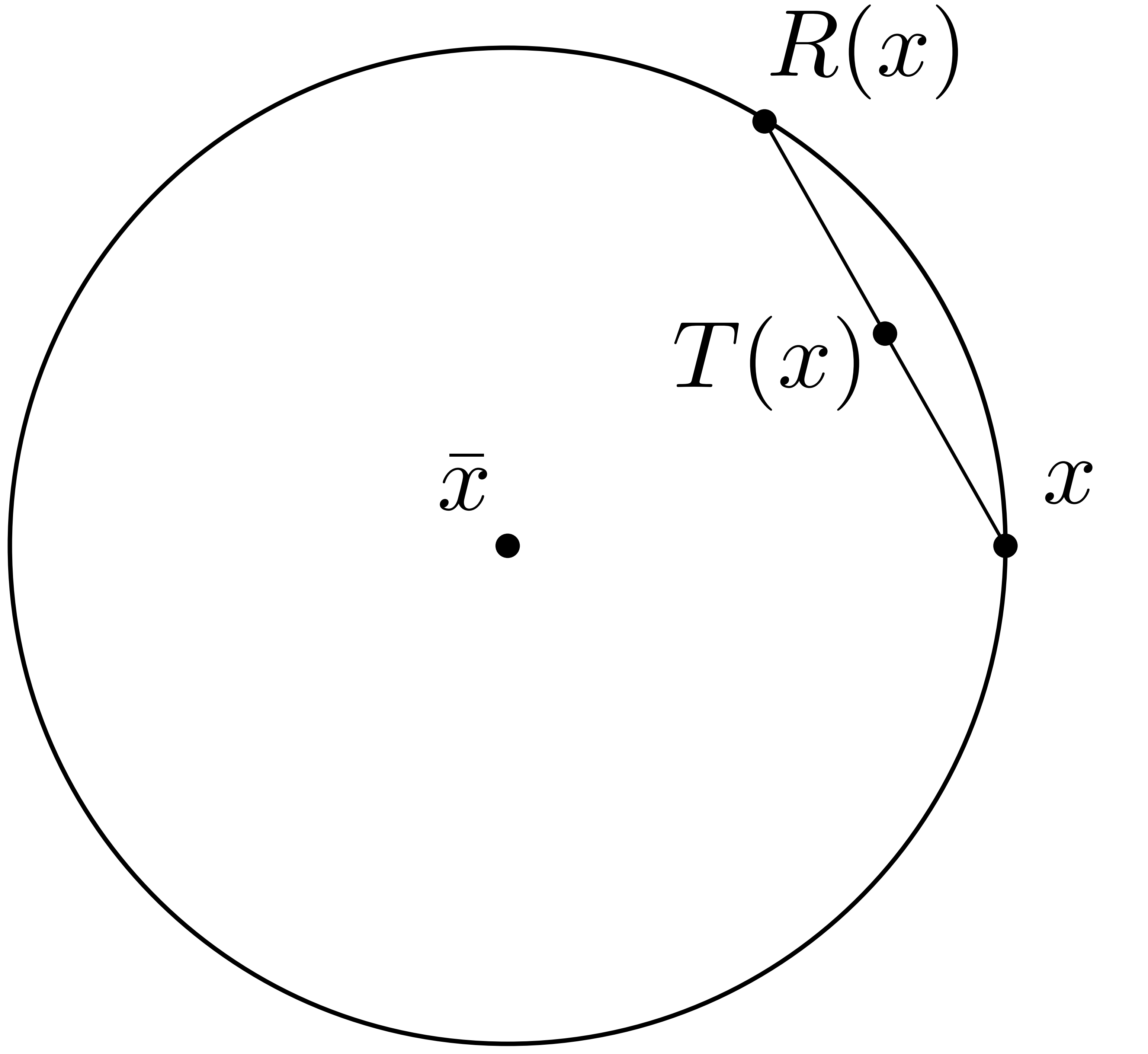


Averaged operators

We say that an operator T is $\alpha-$**averaged** with $\alpha \in (0,1)$ if

$$T = (1-\alpha)I + \alpha R$$

and R is nonexpansive.



Averaged operators fixed points

We say that an operator T is $\alpha-$**averaged** with $\alpha \in (0,1)$ if

$$T = (1-\alpha)I + \alpha R$$

Fact If T is α-averaged, then $\mathbf{fix}\, T = \mathbf{fix}\, R$

Proof

$$\begin{aligned} \bar{x} = T(\bar{x}) &= (1-\alpha)I(\bar{x}) + \alpha R(\bar{x}) \\ &= (1-\alpha)\bar{x} + \alpha R(\bar{x}) \\ \iff &\quad \alpha\bar{x} = \alpha R(\bar{x}) \\ \iff &\quad \bar{x} = R(\bar{x}) \end{aligned}$$

$\blacksquare$

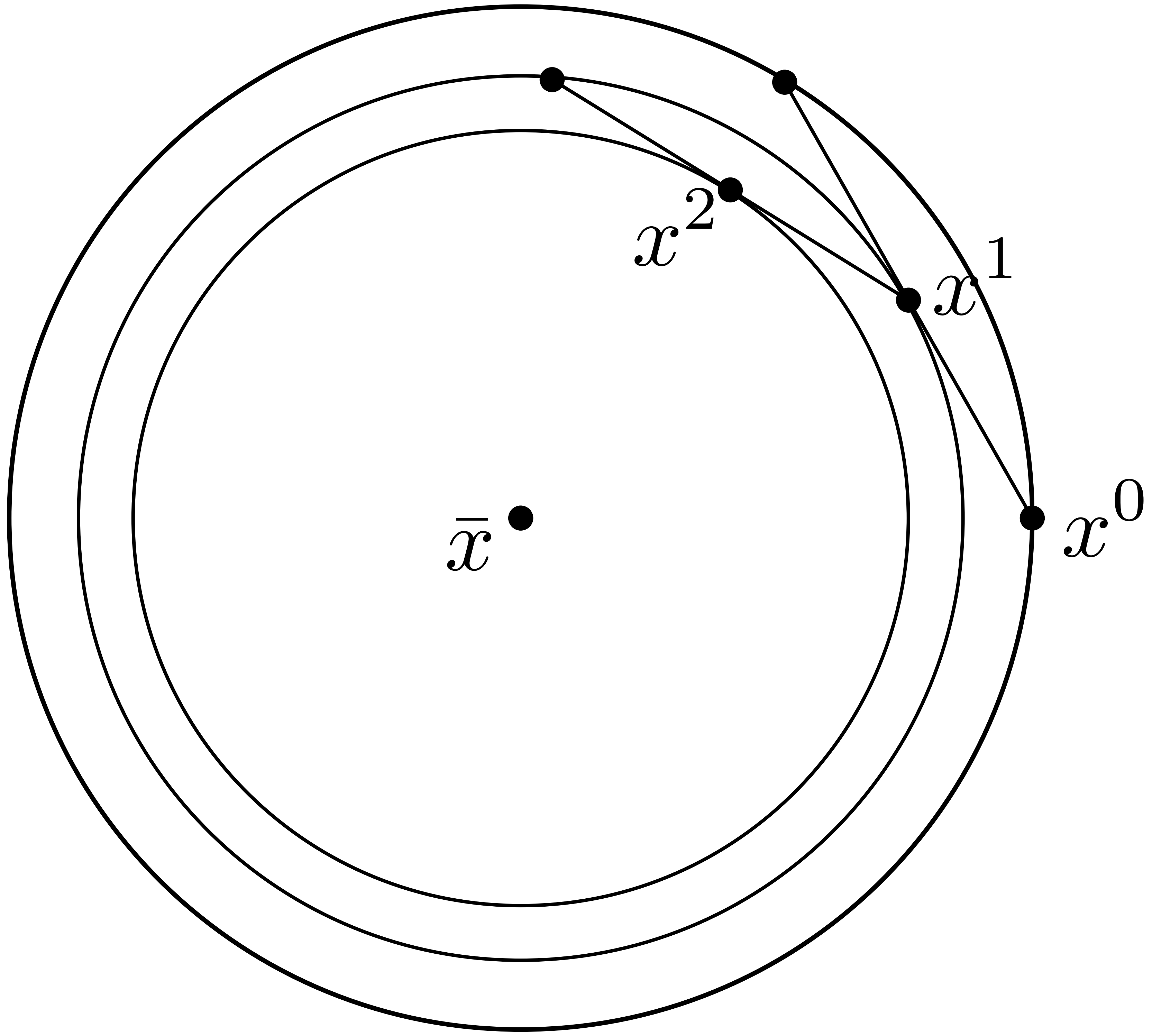
Averaged fixed point iterations

If $T = (1-\alpha)I + \alpha R$ is α-averaged ($\alpha \in (0,1)$ and R nonexpansive), the iteration

$$x^{k+1} = T(x^k)$$

converges to $\bar{x} \in \mathbf{fix}\, T$

(also called damped, averaged or Mann-Krasnosel'skii iteration)



Properties

- Distance to $\bar{x}$ does not increase at each step (**Fejer monotone**)
- Sublinear convergence to fixed-point residual

$$\|R(x^k) - x^k\| \leq \frac{1}{\sqrt{(k+1)\alpha(1-\alpha)}}\|x^0 - \bar{x}\|$$

Averaged fixed point iterations

Proof

Use the identity (proof by expanding)

$$\|(1-\alpha)a + \alpha b\|^2 = (1-\alpha)\|a\|^2 + \alpha\|b\|^2 - \alpha(1-\alpha)\|a-b\|^2$$

and apply it to

$$x^{k+1} - \bar{x} = (1-\alpha)(\underbrace{x^k - \bar{x}}_{a}) + \alpha(\underbrace{R(x^k) - \bar{x}}_{b})$$

obtaining

$$\begin{aligned}\|x^{k+1} - \bar{x}\|^2 &= (1-\alpha)\|x^k - \bar{x}\|^2 + \alpha\|R(x^k) - \bar{x}\|^2 - \alpha(1-\alpha)\|x^k - R(x^k)\|^2 \\ &\le (1-\alpha)\|x^k - \bar{x}\|^2 + \alpha\|x^k - \bar{x}\|^2 - \alpha(1-\alpha)\|x^k - R(x^k)\|^2 \quad \text{(nonexpansive)} \\ &= \|x^k - \bar{x}\|^2 \underbrace{- \alpha(1-\alpha)\|x^k - R(x^k)\|^2}_{\le 0}\end{aligned}$$

Iterations are Fejer monotone

Averaged fixed point iterations

Proof (continued)

iterate righthand side over k steps

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

Since $\|x^{k+1} - \bar{x}\|^2 \geq 0$, we have $$\sum_{i=0}^{k}\|x^i - R(x^i)\|^2 \leq \frac{1}{\alpha(1-\alpha)}\|x^0 - \bar{x}\|^2$$

Using $\displaystyle\sum_{i=0}^{k}\|x^i - R(x^i)\|^2 \geq (k+1)\min_{i=0,\dots,k}\|x^i - R(x^i)\|^2$, we obtain

$$\min_{i=0,\dots,k}\|x^i - R(x^i)\|^2 \leq \frac{1}{(k+1)\alpha(1-\alpha)}\|x^0 - \bar{x}\|^2$$

(R is nonexpansive $\rightarrow \min$ at k) $$\|x^k - R(x^k)\|^2 \leq \frac{1}{(k+1)\alpha(1-\alpha)}\|x^0 - \bar{x}\|^2 \quad \blacksquare$$

Average fixed point iteration convergence rates

$$\|R(x^k) - x^k\| \leq \frac{1}{\sqrt{(k+1)\alpha(1-\alpha)}}\|x^0 - \bar{x}\|$$

Righthand side minimized when $\alpha = 1/2$

$$\|R(x^k) - x^k\| \leq \frac{2}{\sqrt{k+1}}\|x^0 - \bar{x}\|$$

Iterations

$$x^{k+1} = (1/2)x^k + (1/2)R(x^k)$$

Remarks

- Sublinear convergence (same as subgrad method), in general not the actual rate
- $\alpha = 1/2$ is very common for averaged operators

How to design an algorithm

Problem

minimize $\quad f(x)$

Algorithm (operator) construction

1. Find a suitable T such that $\bar{x} \in \mathbf{fix}\, T$ solve your problem
2. Show that the fixed point iteration converges

If T is contractive $\quad \Longrightarrow \quad$ **linear convergence**
If T is averaged $\quad \Longrightarrow \quad$ **sublinear convergence**

Most first order algorithms can be constructed in this way

Operator theory

Today, we learned to:

- **Define** operators and fixed points
- **Define** operator properties such as monotonicity
- **Use operator theory** to construct general fixed-point iterations and prove their convergence

Next lecture

- Operators in optimization algorithms