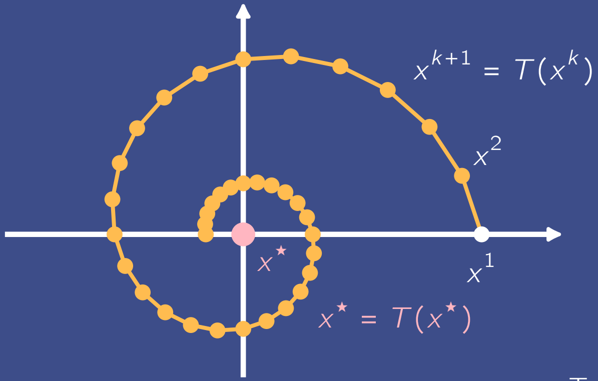
Krasnosel'skiĭ-Mann Iteration

an elegant and powerful abstraction



Howard Heaton

Typal Academy

Background

Observation

Many engineering problems reduce to finding a steady state or an equilibrium

Goal

Model problems and solution techniques, abstracting away unnecessary details

Approach

For an operator T, consider the fixed point problem[†]

$$\text{find } x^\star \text{ such that } T(x^\star) = x^\star$$

[†] Throughout, vectors are assumed to be in $\mathbb{R}^n$, although results apply more generally

Operator Definitions

Norm of a vector x is given by $||x|| = \sqrt{x \cdot x}$, where $\cdot$ is used for dot products

An operator Q is nonexpansive provided it is 1-Lipschitz, *i.e.*

$$||Q(x) - Q(y)|| \leq ||x - y||, \quad \text{for all } x \text{ and } y$$

An operator T is averaged is there is $\alpha \in (0, 1)$ and nonexpansive Q such that

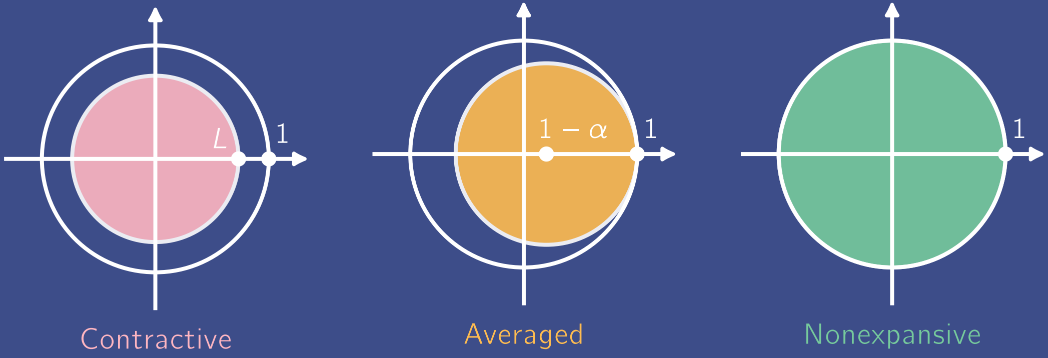
$$T(x) = (1 - \alpha)x + \alpha Q(x) = x + \alpha\big(Q(x) - x\big) \quad \text{for all } x \text{ and } y$$

An operator F is a contraction if there is $L \in [0, 1)$ such that

$$||F(x) - F(y)|| \leq L||x - y||, \quad \text{for all } x \text{ and } y$$

Operator Illustrations

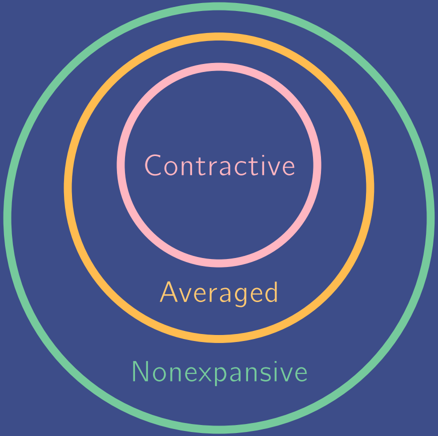
Each operator class corresponds to a circle in the plane*



*Formal definition of the shown scaled relative graphs is beyond scope of talk

Operator Relations

{Contractive Operators} ⊂ {Averaged Operators} ⊂ {Nonexpansive Operators}



proofs in appendix

Banach Fixed Point Theorem

If F is a contraction, then it has a unique fixed point $x^\star$ and the iteration

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

generates a sequence $\{x^k\}$ converging to $x^\star$ with

$$\|x^{k+1} - x^\star\| \leq L^k \|x^1 - x^\star\| \quad \text{for all } k$$

Each contraction has a unique fixed point

Nonexpansive operator might not have fixed points (*e.g.* $T(x) = x + 1$ for $x \in \mathbb{R}$)

Proof – Banach Fixed Point Theorem

Convergence Given $\varepsilon > 0$, it suffices to show $\{x^k\}$ is Cauchy, *i.e.* for some index N

$$\|x^n - x^m\| \leq \varepsilon \quad \text{for all } m, n \geq N$$

For $n \geq m$, repeated use of triangle inequality and L-Lipschitz property yields

$$\begin{aligned} \|x^n - x^m\| &\leq \|x^n - x^{n-1}\| + \cdots + \|x^{m+1} - x^m\| \\ &\leq L^{n-2}\|x^2 - x^1\| + \cdots + L^{m-1}\|x^2 - x^1\| \end{aligned}$$

This implies

$$\|x^n - x^m\| \leq \|x^2 - x^1\| \cdot \sum_{k=m-1}^{n-2} L^k = L^{m-1}\|x^2 - x^1\| \cdot \sum_{k=0}^{n-m-1} L^k \leq \frac{L^{m-1}\|x^2 - x^1\|}{1 - L}$$

Convergence follows by picking N large enough to ensure $L^{N-1}\|x^2 - 1\| \leq (1 - L)\varepsilon$

Proof – Banach Fixed Point Theorem

Limit is Fixed Point Since F is Lipschitz, it is continuous and

$$x^\star = \lim_{k\to\infty} x^{k+1} = \lim_{k\to\infty} F(x^k) = F\left(\lim_{k\to\infty} x^k\right) = F(x^\star)$$

Uniqueness If $x^\star$ and $y^\star$ were distinct fixed points, then

$$0 < \|x^\star - y^\star\| = \|F(x^\star) - F(y^\star)\| \leq L\|x^\star - y^\star\| \quad \Longrightarrow \quad 1 < L$$

a contradiction

Convergence Rate Inductively use L-Lipschitz and fixed point properties to get

$$\|x^{k+1} - x^\star\| = \|F(x^k) - F(x^\star)\| \leq L\|x^k - x^\star\| \leq \cdots \leq L^k\|x^1 - x^\star\|$$

Problem – Contractions are Too Restrictive

Loosely, in optimization we can map

$$\text{strongly convex function} \implies \text{contractive operator} \qquad \text{(Special Case)}$$

$$\text{convex function} \implies \text{nonexpansive operator} \qquad \text{(General Case)}$$

In many applications, we have nonexpansive Q rather than contractive F

How can we compute fixed points of nonexpansive operators?

Banach's theorem does not apply

Krasnosel'skiĭ-Mann Iteration

With nonexpansive Q and positive step sizes $\{\alpha_k\}$, update via

$$x^{k+1} = (1-\alpha_k)x^k + \alpha_k Q(x^k) \tag{KM Update}$$

If Q has a fixed point, the sequence $\{x^k\}$ converges to a fixed point of Q provided the step sizes satisfy

$$\sum_{k=1}^{\infty} \alpha_k(1-\alpha_k) = \infty$$

Special Case[†]

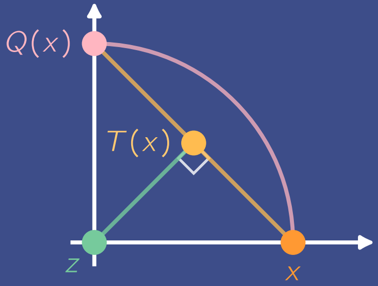
Set $\alpha \in (0,1)$ and $T(x) = (1-\alpha)x + \alpha Q(x)$ to iterate via $x^{k+1} = T(x^k)$

[†] Remaining slides focus on this special case

Descent Lemma

If T is α-averaged and $z = T(z)$, then

$$\|T(x) - z\|^2 \leq \|x - z\|^2 - \frac{1-\alpha}{\alpha}\|T(x) - x\|^2 \quad \text{for all } x$$



Rotation Example

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

$$\|T(x) - z\|^2 = \|x - z\|^2 - \|T(x) - x\|^2$$

$$\|T(x) - z\|^2 \leq \|x - z\|^2 - \left(1 - \frac{1}{2}\right)\|T(x) - x\|^2$$

Here $Q(x)$ is a counterclockwise rotation of 90 degrees and $T(x) = \frac{1}{2}(x + Q(x))$

Descent Lemma Proof

Since T is α-averaged, there is nonexpansive Q such that

$$T(x) = (1 - \alpha)x + \alpha Q(x)$$

which yields the residual relation

$$||T(x) - x||^2 = ||(1 - \alpha)x + \alpha Q(x) - x||^2 = \alpha^2 ||Q(x) - x||^2$$

Step 1 – Obtain a dot product bound

Step 2 – Use this bound and residual relation to get result

Descent Lemma Proof – Step 1

$$
\begin{aligned}
& 2(Q(x) - x) \cdot (x - z) \\
&= \|(Q(x) - x) + (x - z)\|^2 - \|Q(x) - x\|^2 - \|x - z\|^2 && \text{(Complete Square)} \\
&= \|Q(x) - z\|^2 - \|Q(x) - x\|^2 - \|x - z\|^2 && \text{(Simplify)} \\
&= \|Q(x) - Q(z)\|^2 - \|Q(x) - x\|^2 - \|x - z\|^2 && \text{(Fixed Point)} \\
&\leq \|x - z\|^2 - \|Q(x) - x\|^2 - \|x - z\|^2 && \text{(Nonexpansive)} \\
&= -\|Q(x) - x\|^2 && \text{(Cancel Terms)}
\end{aligned}
$$

Descent Lemma Proof – Step 2

$$\|T(x) - z\|^2$$

$$= \|(x - z) + \alpha(Q(x) - x)\|^2 \qquad \text{(Substitute for } T\text{)}$$

$$= \|x - z\|^2 + 2\alpha(x - z) \cdot (Q(x) - x) + \alpha^2 \|Q(x) - x\|^2 \qquad \text{(Expand)}$$

$$\leq \|x - z\|^2 + (-\alpha + \alpha^2) \|Q(x) - x\|^2 \qquad \text{(Apply Step 1)}$$

$$= \|x - z\|^2 - (1 - \alpha)\alpha \|Q(x) - x\|^2 \qquad \text{(Factor Quadratic)}$$

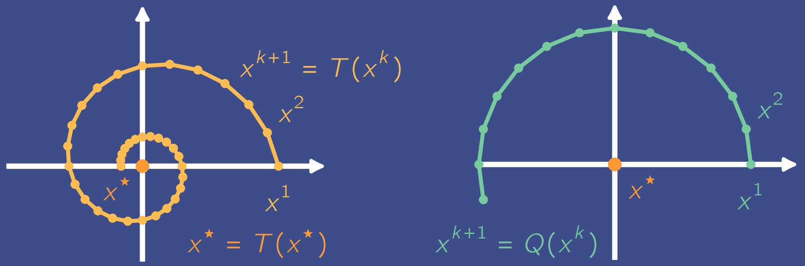
$$= \|x - z\|^2 - \frac{1 - \alpha}{\alpha} \|T(x) - x\|^2 \qquad \text{(Residual Relation)}$$

Convergence Theorem

If there is z such that $T(z) = z$ and T is averaged, then the iteration

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

generates a sequence $\{x^k\}$ converging a point $x^\star$ for which $x^\star = T(x^\star)$



Applying rotation yields cycling, but averaging it with identity converges

Proof – Subsequence Converges

By the descent lemma, for some $\alpha \in (0, 1)$,

$$\|x^{k+1} - z\|^2 = \|T(x^k) - z\|^2 \leq \|x^k - z\|^2 - \frac{1-\alpha}{\alpha}\|T(x^k) - x^k\|^2 \quad \text{for all } k$$

This implies $\{\|x^k - z\|\}$ is monotonically decreasing, and so

$$\|x^k\| = \|x^k - z + z\| \leq \|x^k - z\| + \|z\| \leq \|x^1 - z\| + \|z\| \quad \text{for all } k$$

Thus, $\{x^k\}$ is bounded and the Bolzano-Weierstrass theorem asserts there is a convergent subsequence $\{x^{n_k}\}$ with limit $x^\star$

Proof – Limit Point is Fixed Point

By the descent lemma,

$$\|T(x^k) - x^k\|^2 \le \frac{\alpha}{1-\alpha}\left(\|x^k - z\| - \|x^{k+1} - z\|\right)$$

This yields the telescoping series

$$\sum_{k=1}^{\infty} \|T(x^k) - x^k\|^2 \le \frac{\alpha}{1-\alpha}\left(\|x^1 - z\|^2 - \lim_{N\to\infty} \|x^N - z\|^2\right) \le \frac{\|x^1 - z\|^2}{1-\alpha}$$

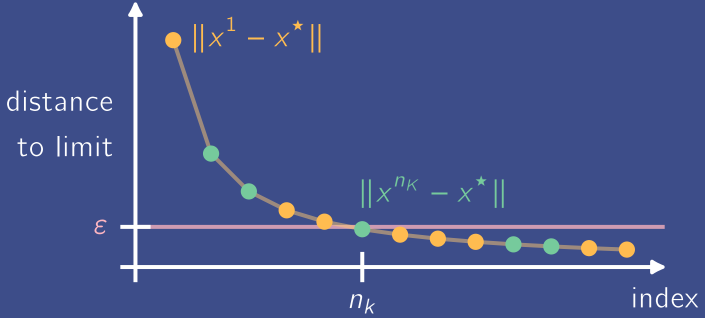
Since the series converges, its summands converge to zero, *i.e.*

$$0 = \lim_{k\to\infty} \|T(x^{n_k}) - x^{n_k}\| = \left\| T\left(\lim_{k\to\infty} x^{n_k}\right) - \lim_{k\to\infty} x^{n_k} \right\| = \|T(x^\star) - x^\star\|$$

where second equality holds as both T and norms are continuous

Proof – Entire Sequence Converges (Intuition)

Since $x^{n_k} \to x^\star$ and $\{\|x^k - x^\star\|\}$ is monotonically decreasing, we get $x^n \to x^\star$



Plot of $\{\|x^n - x^\star\|\}$, with green dots showing subsequence $\{\|x^{n_k} - x^\star\|\}$

Proof – Entire Sequence Converges (Formal)

Given $\varepsilon > 0$, the fact $x^{n_k} \to x^\star$ implies there is an index K such that

$$\|x^{n_k} - x^\star\| \leq \varepsilon \quad \text{for all } k \geq K$$

As $x^\star = T(x^\star)$, the sequence $\{\|x^k - x^\star\|\}$ is monotonically decreasing,[†] and so

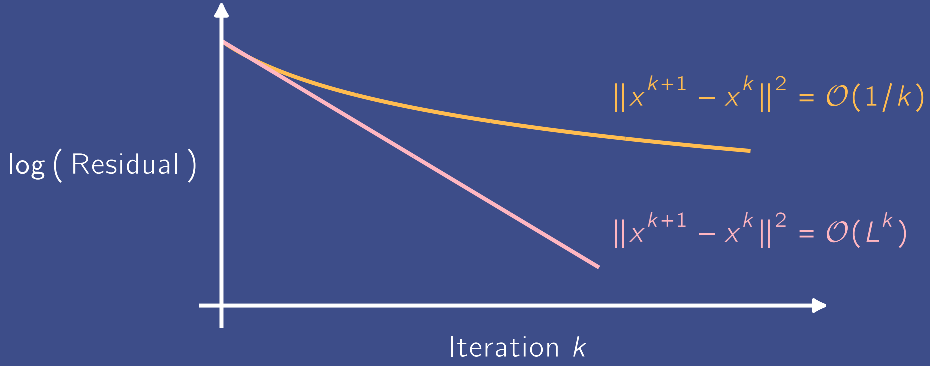
$$\|x^n - x^\star\| \leq \|x^{n_K} - x^\star\| \leq \varepsilon \quad \text{for all } n \geq n_K$$

Hence $x^k \to x^\star$ ■

[†] Apply first step in proof with $z = x^\star$, noting we only assumed $z = T(z)$

Monotonicity of Fixed Point Iteration

For averaged T and contractive F, the residual $\|x^{k+1} - x^k\|^2$ decays monotonically



Checking residual monotonicity can be useful for debugging code

proofs in appendix

Summary

- Three operator classes are of interest: contractive, averaged, nonexpansive
- Banach's theorem shows fast convergence with contractive fixed point updates
- Fixed point iteration with averaged operators converges
- Krasnosel'skiĭ-Mann iteration converges even with nonexpansive operators

References

- Ryu, Yin. *Large Scale Convex Optimization: Algorithms and Analysis via Monotone Operators*. 2023.
- Bauschke, Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. 2017.
- Cegielski. *Iterative Methods for Fixed Point Problems in Hilbert Spaces*. 2012.

Appendix – Averaged $\implies$ Nonexpansive

For an averaged operator T,

$$\begin{aligned}
\|T(x) - T(y)\| &= \|(1-\alpha)x + \alpha Q(x) - (1-\alpha)y - Q(y)\| && \text{(Substitute)} \\
&= \|(1-\alpha)(x-y) + \alpha(Q(x) - Q(y))\| && \text{(Rearrange)} \\
&\leq (1-\alpha)\|x-y\| + \alpha\|Q(x) - Q(y)\| && \text{(Triangle Inequality)} \\
&\leq (1-\alpha)\|x-y\| + \alpha\|x-y\| && \text{(Nonexpansive)} \\
&= \|x-y\| && \text{(Simplify)}
\end{aligned}$$

Appendix – Contractive $\implies$ Averaged

Consider a contraction F with $L \in [0, 1)$ and define

$$Q(x) = \frac{2}{1+L}\left(F(x) - \frac{1-L}{2}x\right)$$

Note Q is nonexpansive since, by the triangle inequality and fact F is L-Lipschitz,

$$\|Q(x) - Q(y)\| \leq \frac{2}{1+L}\left(L + \frac{1-L}{2}\right)\|x - y\| = \|x - y\|$$

Rearranging the formula for Q reveals

$$F(x) = \left(1 - \frac{1+L}{2}\right)x + \frac{1+L}{2}Q(x) \quad \implies \quad F \text{ is averaged}$$

Appendix – Rates of Convergence

Monotonicity holds for averaged T since

$$\|x^{k+1} - x^k\| = \|T(x^k) - T(x^{k-1})\| \leq \|x^k - x^{k-1}\|$$

Applying this with the telescoping series in proof of key result gives

$$k\|x^{k+1} - x^k\|^2 \leq \sum_{i=1}^{k} \|T(x^i) - x^i\|^2 \leq \frac{\|x^1 - z\|^2}{1-\alpha} \implies \|x^{k+1} - x^k\|^2 \leq \frac{\|x^1 - z\|^2}{(1-\alpha)k}$$

For contractive F,

$$\|x^{k+1} - x^k\| = \|F(x^k) - F(x^{k-1})\| \leq L\|x^k - x^{k-1}\| \leq \cdots \leq L^k \cdot \frac{\|x^2 - x^1\|}{L}$$