Lecture 15

Last time
- Finish proof
- New topic
- Proximal point method

Today
- Fixed points
- Set-valued mappings
- Krasnoselskii-Mann iteration.

Fixed Point Iteration

The next result is key for our shift in perspective.

Proposition ⊠: A point x^* minimizes f if, and only if it is a fixed point of $T(x) := \operatorname{prox}_{\alpha f}(x)$ for some $\alpha > 0$.

(e.g., $x^* = T(x^*)$).

Proof: Damn it, another exercise! $\square$

Notice that PPM is just the fixed point iteration

$$(\diamond) \qquad x_{k+1} = T(x_k).$$

Thus, to study its convergence it suffices to understand when does $(\diamond)$ for general operators $T: E \to E$.

Def: A set-value mapping $F: E \rightrightarrows E$ is a mapping from E to 2^E (subsets of E). Its inverse $F^{-1}: E \rightrightarrows E$ is defined via

$$x \in F^{-1}(y) \Leftrightarrow y \in F(x).$$

Remark: $T(x)$ can be empty.

Example

- Notice that

$$\operatorname{prox}_{\alpha f}(x) = y \Leftrightarrow (x-y) \in \underbrace{\alpha \partial f}_{\text{set value map}}(y)$$

$$\Leftrightarrow y \in \underbrace{(I + \alpha \partial f)^{-1}}_{\text{Resolvant}}(x)$$

- In the particular case where

$f = \iota_C$ for some convex set C.

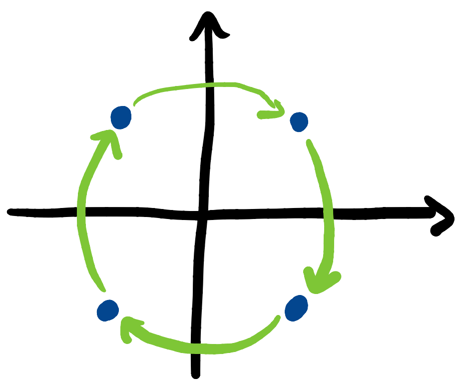
$$\operatorname{prox}_{\alpha \iota_C}(x) = y \iff (x-y) \in N_C(y)$$
$$\iff \operatorname{proj}_C(x) = y.$$

Def: A set-valued map F is nonexpansive if for all $y \in F(x)$, $y' \in F(x')$ we have
$$\|y - y'\| \leq \|x - x'\|.$$
We say that F is a contraction if the inequality is strict.

For instance projections are nonexpansive (HW 1).

From analysis we know that if $(I + \alpha \partial f)^{-1}$ is a contraction then by the Banach contraction Mapping Theorem, it converges linearly towards a unique fixed point. But, we rarely have a contraction, e.g., take $C = E$.

Warning Iterating noncontractive mappings might fail to converge. Take $F: \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T(x) = \mathrm{Rot}_{\pi/2}(x)$ by a $90°$ degree clockwise rotation.



Then, $T^k(1,1) \in \{(1,-1), (-1,-1), (-1,1), (1,1)\}$.

Even so we can overcome this issue if we average the operator

Theorem (Krasnoselskii-Mann iteration '53) Suppose $F: E \rightrightarrows E$ is averaged meaning

$$F = (1-\theta) I + \theta G$$

$\theta \in (0,1)$ nonexpansive

must converge to a fixed point

of F (or equivalently G) if one exists. $\dashv$

But, why do we care? Resolvants $(I+\alpha\partial f)^{-1}$ are always averaged.

Def: A set-valued map $T: E \rightrightarrows E$ is monotone if for all $y \in T(x)$, $y' \in T(x')$ we have

$$\langle y - y', x - x' \rangle \geq 0. \quad \dashv$$

Example

The subdifferential ∂f is monotone. Let $y \in \partial f(x)$, $y' \in \partial f(x')$, then

$$\langle y', x' - x \rangle \geq f(x') - f(x)$$

and

$$\langle y, x - x' \rangle \leq f(x) - f(x')$$

adding the two yields monotonicity. $\dashv$

Def: For a monotone operator $T: E \rightrightarrows E$ and $\alpha > 0$, we define

- The resolvent as $R := (I + \alpha T)^{-1}$.
- The Caley operator as $C := 2R - I$.

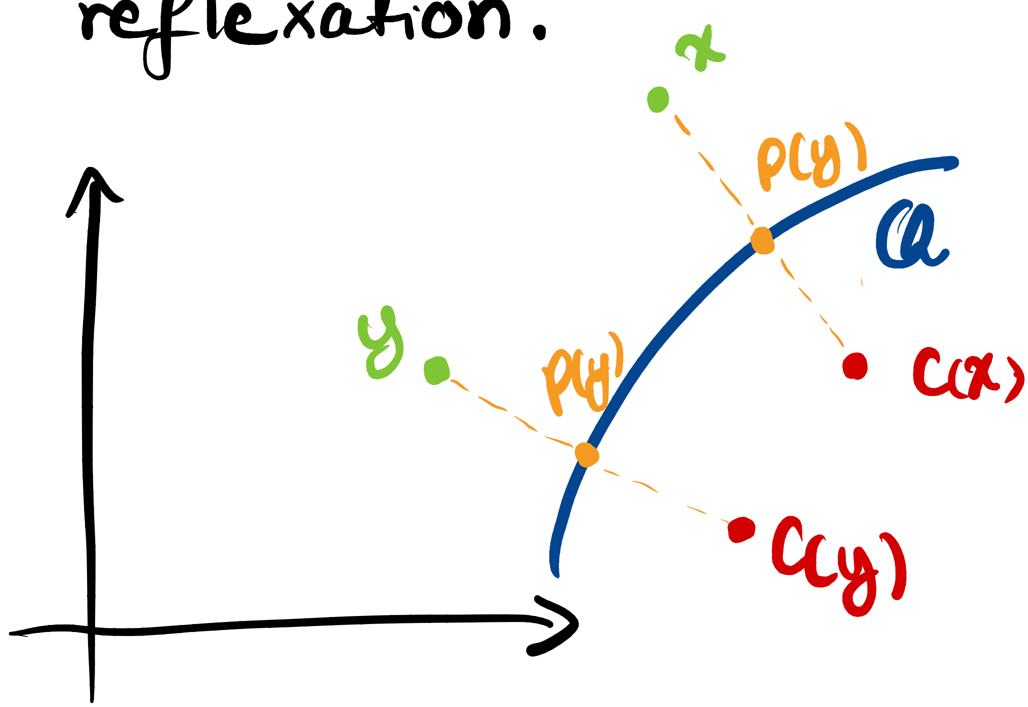
$\dashv$

Lemma: R and C are non-expansive and hence R is averaged.

Proof: C'mon, prove something. Exercise. $\square$

Example: Suppose $T = \iota_{\mathcal{C}}$. Then, $C = 2P - I$ is a reflexation.

convex closed set



$\dashv$

We are now ready to prove an enhanced version of the KM iteration theorem.

Theorem: Suppose $F: E \rightrightarrows E$ is averaged and that $\mathcal{X} = \{x \mid x = Fx\} \neq \emptyset$. Let $x_{k+1} = F(x_k)$ $\forall k \in \mathbb{N}$. Then,

(1) (Convergence) The iterates $x_k \to x^*$ to some $x^* \in \mathcal{X}$.

(2) (Fejér monotonicity) For all $k \in \mathbb{N}$ and $x \in \mathcal{X}$

$$\|x_{k+1} - x\| \leq \|x_k - x\|.$$

(3) (Rate) Let $F = (1-\theta)I + \theta G$, we have

$$\min_{i \in \{0,\ldots,k\}} \|x_k - G(x_k)\| \leq \frac{\operatorname{dist}(x_0, \mathcal{X})}{(1-\theta)\theta\sqrt{k+1}}.$$

Proof: Notice that (2) follows trivially from the fact that F is non-expansive. We prove a slightly stronger version. Take $\bar{x} \in X$, then

$$\|x_{k+1} - \bar{x}\|^2$$
$$= \|F(x_k) - \bar{x}\|^2$$
$$= \|(1-\theta)x_k + \theta G(x_k) - (1-\theta)\bar{x} + \theta G(\bar{x})\|^2.$$

Here we need a claim

Claim: For any $a, b \in E$ and $\theta \in \mathbb{R}$,

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

Proof of the claim: Both terms are quadratic in θ, it suffices to check three points: 0, $\frac{1}{2}$, 1. $\square$

Using this claim we get

$$\|x_{k+1} - \bar{x}\|^2$$
$$= (1-\theta)\|x_k - \bar{x}\|^2 + \theta\|G(x_k) - \bar{x}\|^2$$
$$- (1-\theta)\theta\|(x_k - \bar{x}) - (G(x_k) - G(\bar{x}))\|^2$$
$$\leq (1-\theta)\|x_k - \bar{x}\|^2 + \theta\|x_k - \bar{x}\|^2$$

$$- (1-\theta)\theta \, \|x_k - G(x_k)\|^2$$

$$= \|x_k - \bar{x}\|^2 - (1-\theta)\theta \, \|x_k - G(x_k)\|^2.$$

Then, taking a telescoping sum

$$(1-\theta)\theta \sum_{i=0}^{k} \|x_i - G(x_i)\|^2 \leq \|x_0 - \bar{x}\|^2.$$

Dividing by k+1 and using that a minimum is smaller than an average, yields (3).

To prove (1), notice that by (2) the sequence is bounded. Then, there is a convergent subsequence with limit point x^*. Notice that along this subsequence $\|x_k - G(x_k)\| \to 0$ and G is continuous, hence $x^* \in X$. Moreover, by (2) $\|x_k - x^*\| \downarrow 0$ along the full sequence, hence $x_k \to x^*$, completing the proof. $\square$