

Review

① Moreau-Yosida Regularization of $f(x)$ is

$$f_\eta(x) = \min_u \left[f(u) + \frac{1}{2\eta} \|u - x\|^2 \right]$$

$$\text{Prox}_f^\eta(x) = (I + \eta \partial f)^{-1}(x) = \arg\min_u \left[f(u) + \frac{1}{2\eta} \|u - x\|^2 \right]$$

② $\begin{cases} \text{Prox}_f^\eta(x) = x - \eta \nabla f_\eta(x) \\ \nabla f_\eta \text{ is } L\text{-continuous with } L = \frac{1}{\eta} \end{cases}$

③
$$f_\eta(x) = \min_u \left[f(u) + \frac{1}{2\eta} \|u - x\|^2 \right]$$
$$\min_x f_\eta(x) = \min_{x, u} \left[\underbrace{f(u)}_{f(x_*)} + \frac{1}{2\eta} \underbrace{\|u - x\|^2}_0 \right] = f_*$$

$f_\eta(x)$ has the same minimizer as $f(x)$

④ So the proximal point method

$$x_{k+1} = (I + \eta \partial f)^{-1}(x_k)$$

is equivalent to gradient descent for $f_\eta(x)$

$$\text{step size } \eta \leq \frac{1}{L} \quad (L = \frac{1}{\eta})$$

⑤

Theorem [Prox is firmly nonexpansive] $f(x)$ is convex

$$\| \text{Prox}_f^\eta(x) - \text{Prox}_f^\eta(y) \|^2 \leq \langle \text{Prox}_f^\eta(x) - \text{Prox}_f^\eta(y), x - y \rangle$$

$$\leq \| \text{Prox}_f(x) - \text{Prox}_f(y) \| \cdot \| x - y \|$$

It implies $\| \text{Prox}_f^\eta(x) - \text{Prox}_f^\eta(y) \| \leq \| x - y \|$ (nonexpansive)
 $\hookrightarrow L$ -cont. with $L=1$

Theorem [Prox is a contraction for strongly convex function]

If $f(x)$ is strongly convex with $\mu > 0$

$$(1 + \eta\mu) \| \text{Prox}_f^\eta(x) - \text{Prox}_f^\eta(y) \|^2 \leq \langle \text{Prox}_f^\eta(x) - \text{Prox}_f^\eta(y), x - y \rangle$$

It implies $\| \text{Prox}_f^\eta(x) - \text{Prox}_f^\eta(y) \| \leq \frac{1}{1 + \eta\mu} \| x - y \|$

$$\text{Prox}_f^\eta(x_*) = x_* \Leftrightarrow (I + \eta\partial f)^{-1}(x_*) = x_* \Leftrightarrow x_* \in (I + \eta\partial f)(x_*)$$

Convergence Rate for nonsmooth problems

Subgradient Method

Convexity

$$O\left(\frac{1}{\sqrt{k}}\right)$$

Strong Convexity

$$O\left(\frac{1}{k}\right)$$

Proximal Point Method

$$O\left(\frac{1}{k}\right)$$

$$O\left(\frac{1}{[1 + \eta\mu]^2}\right)^k$$

Proof of the rate $f(x_k) - f(x_*) = O\left(\frac{1}{[1 + \eta\mu]^2}\right)^k$:

Step I: $\| x_{k+1} - x_* \| \leq \frac{1}{1 + \eta\mu} \| x_k - x_* \|$

$$\Rightarrow \|X_k - X_*\| \leq \frac{1}{(1+\eta\mu)^k} \|X_0 - X_*\|$$

Step II: Strong Convexity Only gives

$$f(X_k) \geq f(X_*) + \langle \partial f(X_*), X_k - X_* \rangle + \frac{\mu}{2} \|X_k - X_*\|^2$$

$$\text{Want } 0 < f(X_k) - f(X_*) \leq C \|X_k - X_*\|^2$$

Prox-Grad
Inequality

$$\bar{y} = \text{Prox}_f^{\eta}(y - \eta \nabla g(y))$$

$$F(x) - F(\bar{y}) \geq \frac{1}{2\eta} \|x - \bar{y}\|^2 - \frac{1}{2\eta} \|x - y\|^2 + g(x) - g(y) - \langle \nabla g(y), x - y \rangle$$

$$g(x) \equiv 0, \quad y = X_k$$

$$0 \geq 2\eta [F(X_*) - F(X_{k+1})] \geq \|X_* - X_{k+1}\|^2 - \|X_* - X_k\|^2$$

$$\begin{aligned} \Rightarrow f(X_{k+1}) - f(X_*) &\leq -\frac{1}{2\eta} \|X_* - X_{k+1}\|^2 + \frac{1}{2\eta} \|X_* - X_k\|^2 \\ &\leq \frac{1}{2\eta} \|X_k - X_*\|^2 \end{aligned}$$

Def An operator $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is

- 1) Firmly nonexpansive if $\|Tx - Ty\|^2 \leq \langle Tx - Ty, x - y \rangle$
- 2) Nonexpansive if $\|Tx - Ty\| \leq \|x - y\|$
- 3) Contraction if $\|Tx - Ty\| < \|x - y\|$

Theorem $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is an operator

$I: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the identity operator

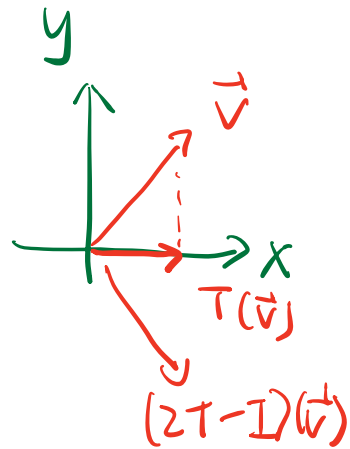
The following are equivalent:

- ① T is firmly nonexpansive
- ② $I - T$ is firmly nonexpansive

- ③ $2T - I$ is nonexpansive

- ④ $\|Tx - Ty\|^2 + \|(I - T)x - (I - T)y\|^2 \leq \|x - y\|^2$

T is nonexpansive $\Rightarrow \frac{T+I}{2}$ is firmly nonexpansive



Proof: ① $\Leftrightarrow$ ②: $\|Tx - Ty\|^2 \leq \langle Tx - Ty, x - y \rangle$

$$\| [x - Tx] - [y - Ty] \|^2$$

$$= \|x - y\|^2 + \|Tx - Ty\|^2 - 2 \langle Tx - Ty, x - y \rangle$$

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

$$= \langle [x - T(x)] - [y - T(y)], x - y \rangle$$

$$\textcircled{1} \Leftrightarrow \textcircled{3} \quad R = 2T - I$$

$$\|T(x) - T(y)\|^2 \leq \langle T(x) - T(y), x - y \rangle$$

$\Downarrow$

$$\begin{aligned} \|R(x) - R(y)\|^2 &= \|2(T(x) - T(y)) - (x - y)\|^2 \\ &= 4\|T(x) - T(y)\|^2 + \|x - y\|^2 - 4\langle T(x) - T(y), x - y \rangle \\ &\stackrel{\textcircled{3}}{\leq} \|x - y\|^2 \end{aligned}$$

$$\textcircled{1} \Leftrightarrow \textcircled{4}$$

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

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

$$\Leftrightarrow \|T(x) - T(y)\|^2 \leq \langle T(x) - T(y), x - y \rangle$$

Example: Assume ∇f is L -continuous

$\textcircled{1}$ $T(x) = \frac{1}{L} \nabla f(x)$ is nonexpansive

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \Leftrightarrow \left\| \frac{1}{L} \nabla f(x) - \frac{1}{L} \nabla f(y) \right\| \leq \|x - y\|$$

② If $f(x)$ is convex, then

$T(x) = \frac{1}{L} \nabla f(x)$ is firmly nonexpansive

$T(x) = x - \frac{1}{L} \nabla f(x)$ is firmly nonexpansive

2.3 Convergence for convex functions

Theorem 2.9. Assume $\nabla f(\mathbf{x})$ is Lipschitz-continuous with Lipschitz constant L and $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex. Then for any $\mathbf{x}, \mathbf{y}$:

1. $f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{1}{2L} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2$

2. $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2 \leq L \langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle.$

$\hookrightarrow \left\| \frac{1}{L} \nabla f(\mathbf{x}) - \frac{1}{L} \nabla f(\mathbf{y}) \right\|^2 \leq \left\langle \frac{1}{L} \nabla f(\mathbf{x}) - \frac{1}{L} \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \right\rangle$

Given an operator T , Example: 1) GD 2) Proximal

consider $x_{k+1} = T(x_k)$

1) Is there a fixed point $T(x_*) = x_*$?

2) $x_k \rightarrow x_*$?

Fixed Point Theorems

Theorem 6.8 (Banach Fixed Point Theorem). Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a contractive operator/mapping, then T has a unique fixed point $T(\mathbf{x}_*) = \mathbf{x}_*$.

Theorem 6.9 (Brouwer Fixed Point Theorem). Let B^n be the closed unit ball in $\mathbb{R}^n$. If $T : B^n \rightarrow B^n$ is continuous, then T has at least one fixed point $T(\mathbf{x}_*) = \mathbf{x}_*$.



nonexpansive

$$x \rightarrow y$$

$$\Rightarrow T(x) \rightarrow T(y)$$

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

A few quick examples:

1. If $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is nonexpansive, then T may not have a fixed point. Counterexample: the translation $T(\mathbf{x}) = \mathbf{x} + \mathbf{c}$ with $\mathbf{c} \neq \mathbf{0}$, which is an isometry.
2. If $T : B^n \rightarrow B^n$ is nonexpansive, then T has at least one fixed point but $\mathbf{x}_{k+1} = T(\mathbf{x}_k)$ may not converge. Counterexample: $T(\mathbf{x}) = -\mathbf{x}$.

$$\theta = \frac{1}{2}$$

$$x_{k+1} = \frac{I+T}{2}(x_k)$$

If only assume T is nonexpansive, $x_{k+1} = T(x_k)$ may NOT converge.

Want to show convergence of $x_{k+1} = \theta x_k + (1-\theta) T(x_k)$, $\theta \in (0,1)$

$$T = \frac{1}{2} \nabla f \text{ is nonexpansive} \Leftrightarrow I - T = I - \frac{1}{2} \nabla f \text{ is nonexp.}$$

Theorem Assume $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is nonexpansive, then T has at least one fixed point.

$$x_{k+1} = \theta x_k + (1-\theta) T(x_k), \theta \in (0,1) \text{ satisfies}$$

1) $\{x_k\}$ converges to one fixed point y of T

$$2) \|x_{k+1} - x_k\|^2 \leq \frac{1}{(k+1)} \left(\frac{1}{\theta} - 1\right) \|x_0 - y\|^2$$

→ This is not error!

Proof: ① $x_{k+1} - x_* = \theta [x_k - x_*] + (1-\theta) [T(x_k) - x_*]$

$$\|x_{k+1} - x_*\|^2 = \|\theta [x_k - x_*] + (1-\theta) [T(x_k) - x_*]\|^2$$

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

$$= \theta \|x_k - x_*\|^2 + (1-\theta)\|T(x_k) - x_*\|^2$$

$$- \theta(1-\theta)\|T(x_k) - x_k\|^2$$

$$\|T(x_k) - x_*\| = \|T(x_k) - T(x_*)\| \leq \|x_k - x_*\|$$

$$\leq \theta \|x_k - x_*\|^2 + (1-\theta)\|x_k - x_*\|^2$$

$$- \theta(1-\theta)\|T(x_k) - x_k\|^2$$

$$= \|x_k - x_*\|^2 - \theta(1-\theta)\|T(x_k) - x_k\|^2$$

$$x_{k+1} = S(x_k) := \theta x_k + (1-\theta)T(x_k)$$

$$S(x) - x = (1-\theta)[T(x) - x]$$

$$\|x_{k+1} - x_*\|^2 \leq \|x_k - x_*\|^2 - \frac{\theta}{1-\theta} \|S(x_k) - x_k\|^2$$

② We first get $\|x_{k+1} - x_*\|^2 \leq \|x_k - x_*\|^2$

Sum it up $S(x) = \theta x + (1-\theta) \cdot T(x)$

$$\sum_{k=0}^n \|S(x_k) - x_k\|^2 \leq \frac{1-\theta}{\theta} [\|x_0 - x_*\|^2 - \|x_{n+1} - x_*\|^2]$$

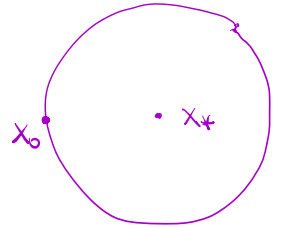
$$\sum_{k=0}^n \|x_{k+1} - x_k\|^2 \leq \frac{1-\theta}{\theta} [\|x_0 - x_*\|^2 - \|x_{n+1} - x_*\|^2]$$

$$\begin{aligned}
 \|X_{n+1} - X_n\| &= \|S(X_n) - S(X_{n-1})\| \\
 &\leq \theta \|X_n - X_{n-1}\| + (1-\theta) \|TX_n - TX_{n-1}\| \\
 &\leq \|X_n - X_{n-1}\|
 \end{aligned}$$

$$\Rightarrow (n+1) \|X_{n+1} - X_n\|^2 \leq \frac{1-\theta}{\theta} \|X_0 - X_*\|^2$$

$$\begin{aligned}
 \textcircled{3} \quad \|S(X) - X_*\| &= \|S(X) - S(X_*)\| \leq \|X - X_*\| \\
 \Rightarrow \|X_{k+1} - X_*\|^2 &\leq \|X_k - X_*\|^2
 \end{aligned}$$

$\Rightarrow \{X_k\}$ is in the ball centered at X_* with radius $\|X_0 - X_*\|$



Real Analysis Bounded Sequence in $\mathbb{R}^n$ has a convergent subsequence

$$\Rightarrow X_{k_j} \rightarrow y_*, j \rightarrow \infty$$

$$\|X_{k+1} - X_k\|^2 \leq \frac{1}{k+1} \frac{1-\theta}{\theta} \|X_0 - X_*\|^2$$

$$\Rightarrow \|X_{k+1} - X_k\|^2 \rightarrow 0$$

$$\Rightarrow \|S(X_k) - X_k\| \rightarrow 0$$

$$\Rightarrow (1-\theta) \|TX_k - X_k\| \rightarrow 0$$

$$\Rightarrow \|TX_{k_j} - X_{k_j}\| \rightarrow 0$$

$T(x)-x$ is continuous because $\|T(x)-x-T(y)+y\| \leq 2\|x-y\|$

$$\Rightarrow \|T(y_*) - y_*\| = 0$$

$$\Rightarrow y_* = T(y_*) \Rightarrow y_* = S(y_*)$$

$$\|S(x) - x_*\| = \|S(x) - S(x_*)\| \leq \|x - x_*\|$$

$$\Rightarrow \|x_{k+1} - x_*\|^2 \leq \|x_k - x_*\|^2$$



$$\left\{ \begin{array}{l} \|x_{k+1} - y_*\|^2 \leq \|x_k - y_*\|^2 \\ x_{k_j} \rightarrow y_* \end{array} \right.$$

$$\Rightarrow \{x_k\} \rightarrow y_*.$$
