Review

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

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

$$\mathrm{Prox}_f^\eta(x) = (I+\eta\partial f)^{-1}(x) = \operatorname*{argmin}_u \left[f(u) + \frac{1}{2\eta}\|u-x\|^2 \right]$$

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

③ $$f_\eta(x) = \min_u \left[f(u) + \frac{1}{2\eta}\|u-x\|^2 \right]$$

$$\min_x f_\eta(x) = \min_{x,u} \Big[\underbrace{f(u)}_{f(x_*)} + \underbrace{\frac{1}{2\eta}\|u-x\|^2}_{0} \Big] = f_*$$

$f_\eta(x)$ 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 $f_\eta(x)$

but it only gives convergence rate for small $\eta > 0$.

(5)

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

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

It implies $\|\mathrm{Prox}_f^\eta(x) - \mathrm{Prox}_f^\eta(y)\| \leq \|x-y\|$ (nonexpansive)

$\hookrightarrow$ L-cont. with $L=1$

Theorem [Prox is a contraction for strongly convex function]

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

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

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

Convergence Rate for nonsmooth problems

	Convexity	Strong Convexity
Subgradient Method	$O(\frac{1}{\sqrt{k}})$	$O(\frac{1}{k})$
Proximal Point Method	$O(\frac{1}{k})$	$O\left(\frac{1}{[1+\eta\mu]^2}\right)^k$

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

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

$$\Rightarrow \|x_k - x_*\| \le \frac{1}{(1+\eta\mu)^k}\|x_0 - x_*\|$$

Step II: Strong Convexity Only gives

$$f(x_k) \ge f(x_*) + \langle \partial f(x_*), x_k - x_* \rangle + \frac{\mu}{2}\|x_k - x_*\|^2$$

Want $0 < f(x_k) - f(x_*) \le C\|x_k - x_*\|^2$

Prox-Grad Inequality

$$\begin{cases} \bar{y} = \mathrm{Prox}_f^{\eta}(y - \eta \nabla g(y)) \\ F(x) - F(\bar{y}) \ge \frac{1}{2\eta}\|x-\bar{y}\|^2 - \frac{1}{2\eta}\|x-y\|^2 \\ \qquad + g(x) - g(y) - \langle \nabla g(y), x-y \rangle \end{cases}$$

$g(x) \equiv 0$, $y = x_k$

$$0 \ge 2\eta[F(x_*) - F(x_{k+1})] \ge \|x_* - x_{k+1}\|^2 - \|x_* - x_k\|^2$$

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

$I: \mathbb{R}^n \to \mathbb{R}^n$ 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 $\|Tx-Ty\|^2 + \|(I-T)x-(I-T)y\|^2 \leq \|x-y\|^2$

Proof: ① ⇔ ② : $\|T(x)-T(y)\|^2 \leq \langle T(x)-T(y), x-y\rangle$

$$\|[x-T(x)]-[y-T(y)]\|^2$$

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

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

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

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

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

$$\Updownarrow$$

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

$$= 4\|T(x) - T(y)\|^2 + \|x - y\|^2 - 4\langle T(x) - T(y), x - y \rangle$$

$$\overset{③}{\leq} \|x - y\|^2$$

① $\Leftrightarrow$ ④

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

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

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

Example: Assume ∇f is L-continuous

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

③ 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 2.2. *Assume $\nabla f(\mathbf{x})$ is Lipschitz-continuous with Lipschitz constant L and $f(\mathbf{x}) : \mathbb{R}^n \longrightarrow \mathbb{R}$ 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. $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2 \leq L\langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle$.

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) $x_k \to x_*$?

Fixed Point Theorems

Theorem 6.4 (Banach Fixed Point Theorem). *Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a contractive operator/mapping, then T has a unique fixed point $T(\mathbf{x}_*) = \mathbf{x}_*$.*

Theorem 6.5 (Brouwer Fixed Point Theorem). *Let S^n be the unit ball in $\mathbb{R}^n$. If $T : S^n \longrightarrow S^n$ is continuous, then T has at least one fixed point $T(\mathbf{x}_*) = \mathbf{x}_*$.*

A few quick examples:

1. If $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is nonexpansive, then T may not have a fixed point. Counter example, $T(x, y) = \begin{pmatrix} \cos\theta & -\sin\theta \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.
2. If $T : S^n \longrightarrow S^n$ is nonexpansive, then T has at least one fixed point but $\mathbf{x}_{k+1} = T(\mathbf{x}_k)$ may not converge. Counter example: $T(\mathbf{x}) = -\mathbf{x}$.

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

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

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

$x_{k+1} = \theta x_k + (1-\theta) T(x_k)$, $\theta \in (0,1)$ 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 $\qquad S(x) = \theta x + (1-\theta)\cdot T(x)$

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

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

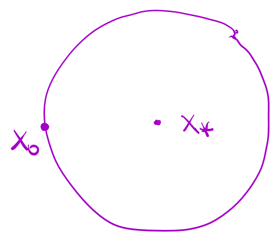
$$\|x_{n+1} - x_n\| = \|S(x_n) - S(x_{n-1})\|$$
$$\leq \theta \|x_n - x_{n-1}\| + (1-\theta)\|Tx_n - Tx_{n-1}\|$$
$$\leq \|x_n - x_{n-1}\|$$

$$\Rightarrow \quad (n+1)\|x_{n+1} - x_n\|^2 \leq \frac{1-\theta}{\theta}\|x_0 - x_*\|^2$$

③ $$\|S(x) - x_*\| = \|S(x) - S(x_*)\| \leq \|x - x_*\|$$

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

$\Rightarrow$ $\{x_k\}$ is in the ball centered at x_* with radius $\|x_0 - x_*\|$



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

$$\Rightarrow \quad x_{k_j} \to y_*, \; j \to \infty$$

$$\|x_{k+1} - x_k\|^2 \leq \frac{1}{k+1}\frac{1-\theta}{\theta}\|x_0 - x_*\|^2$$

$$\Rightarrow \|x_{k+1} - x_k\|^2 \to 0$$

$$\Rightarrow \|S(x_k) - x_k\| \to 0$$

$$\Rightarrow (1-\theta)\|T(x_k) - x_k\| \to 0$$

$$\Rightarrow \|T(x_{k_j}) - x_{k_j}\| \to 0$$

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

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

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

$$\boxed{\begin{aligned} &\|S(x)-x_*\| = \|S(x)-S(x_*)\| \leq \|x-x_*\| \\ \Rightarrow \quad &\|x_{k+1}-x_*\|^2 \leq \|x_k - x_*\|^2 \end{aligned}}$$

$$\Downarrow$$

$$\left\{\begin{aligned} &\|x_{k+1} - y_*\|^2 \leq \|x_k - y_*\|^2 \\ &x_{k_j} \to y_* \end{aligned}\right.$$

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