Review

$\boxed{\text{Def}}$ An operator $T: \mathbb{R}^n \to \mathbb{R}^n$ is called

1) a contraction if $\|T(x) - T(y)\| \leq c\,\|x-y\|$
$0 < c < 1$

2) firmly nonexpansive if

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

3) nonexpansive if $\|T(x) - T(y)\| \leq \|x-y\|$

$\boxed{\text{Theorem}}$ ① T is firmly nonexpansive

$\Leftrightarrow$ $2T - I$ is nonexpansive

② T is nonexpansive

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

Example: ① $f(x)$ is convex $\Rightarrow$

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

Prox_f is firmly nonexpansive $\Leftrightarrow$ $R_f = 2\mathrm{Prox}_f - I$ is nonexpansive

② $f(x)$ is strongly convex with $\mu > 0$ $\Rightarrow$

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

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

Consider $\min_x f(x) + g(x)$ f & g are convex
nonsmooth

$$\min_x \underbrace{\|x\|_1}_{f(x)} + \underbrace{\iota_{\{Ax=b\}}(x)}_{g(x)}$$

$$\boxed{A}\,\boxed{x} = \boxed{b}$$

$$\mathrm{Prox}^{\eta}_{f}(x)_i = \begin{cases} x_i - \eta, & x_i > \eta \\ x_i + \eta, & x_i < -\eta \\ 0, & x_i \in [-\eta, \eta] \end{cases}$$

$$\mathrm{Prox}^{\eta}_{g}(x) = x + A^T(AA^T)^{-1}(b - Ax)$$

Splitting for finding $0 \in \partial f(x_*) + \partial g(x_*)$

$\boxed{\text{Lemma}}$ Assume $f(x)$ and $g(x)$ are convex.
and x_* is a minimizer of $f(x) + g(x)$.
Let $y_* \in x_* + \eta\,\partial g(x_*)$, then

$$0 \in \partial f(x_*) + \partial g(x_*) \iff R_f R_g(y_*) = y_*$$

Remark: y_* is a fixed point of

$$\begin{cases} \text{Peaceman-Rachford} & y_{k+1} = R_f R_g(y_k) & \lambda = 2 \\ \text{Douglas-Rachford} & y_{k+1} = \dfrac{I + R_f R_g}{2}(y_k) & \lambda = 1 \\ \text{Generalized DR} & y_{k+1} = (1-\lambda) y_k + \lambda \dfrac{I + R_f R_g}{2}(y_k) & \lambda \in (0, 2) \end{cases}$$

Definition of Douglas-Rachford splitting for minizing $\min_x f(x)+g(x)$

f & g are proper closed convex functions

① Conexity of $f \Rightarrow \mathrm{Prox}_f$ is firmly nonexpansive

$\Updownarrow$

$R_f = 2\,\mathrm{Prox}_f - I$ is nonexpansive

② Similary R_g is nonexpansive

③ $\|R_f R_g(x) - R_f R_g(y)\| \le \|R_g(x) - R_g(y)\|$

$\le \|x-y\|$

$\Rightarrow R_f R_g$ is nonexpansive

$\Updownarrow$

$\frac{I + R_f R_g}{2}$ is firmly nonexpansive

$$\begin{cases} \text{Peaceman-Rachford} & y_{k+1} = R_f R_g(y_k) & \lambda = 2 \\ \text{Douglas-Rachford} & y_{k+1} = \frac{I+R_f R_g}{2}(y_k) & \lambda = 1 \\ \text{General DR} & y_{k+1} = [(1-\theta)I + \theta R_f R_g](y_k) & \theta \in (0,1) \end{cases}$$

$\theta = \frac{\lambda}{2}$ $\qquad = (1-\lambda)y_k + \lambda \frac{I+R_f R_g}{2}(y_k) \quad \lambda \in (0,2)$

Peaceman-Rachford

① The fixed point iteration for convex $f(x)$ & $g(x)$

$y_{k+1} = R_f R_g(y_k)$ converges to y_*

if either $f(x)$ or $g(x)$ is strongly convex.

② $\begin{cases} y_{k+1} = R_f R_g(y_k) \\ x_{k+1} = \mathrm{Prox}_g^\eta(y_{k+1}) \end{cases}$ for convex $f(x)$ & $g(x)$

$g(x)$ is strongly convex $\Rightarrow \{x_k\} \to x_*$

Without strong convexity, $y_{k+1} = R_f R_g(y_k)$ may NOT converge!

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

$$\|R_f(x) - R_f(y)\|^2 + 4\mu\eta \|\mathrm{Prox}_f^\eta(x) - \mathrm{Prox}_f^\eta(y)\|^2 \le \|x-y\|^2$$

$R_f(x) = 2\,\mathrm{Prox}_f^\eta(x) - x$

Remark: with only convexity ($\mu = 0$), we get nonexpansiveness $\|R_f(x) - R_f(y)\| \le \|x-y\|$.

Proof:

$$\|R_f(x) - R_f(y)\|^2 = \|2\mathrm{Prox}_f^\eta(x) - x - (2\mathrm{Prox}_f^\eta(y) - y)\|^2$$

$$= 4\|\mathrm{Prox}_f^\eta(x) - \mathrm{Prox}_f^\eta(y)\|^2 + \|x-y\|^2$$

$$-4\langle \mathrm{Prox}_f^\eta(x) - \mathrm{Prox}_f^\eta(y), x-y\rangle$$

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

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

$$\Downarrow$$

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

$$\leq -4\mu\eta\|\mathrm{Prox}_f^\eta(x) - \mathrm{Prox}_f^\eta(y)\|^2 + \|x-y\|^2$$

$$\Rightarrow \|R_f(x) - R_f(y)\|^2 + 4\mu\eta\|\mathrm{Prox}_f^\eta(x) - \mathrm{Prox}_f^\eta(y)\|^2$$

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

$\boxed{\text{Theorem}}$ (Peaceman-Rachford splitting with strong convexity)

Assume $f(x)$ is convex and $g(x)$ is strongly convex with $\mu > 0$. Then $f(x) + g(x)$ has a unique minimizer x_*. The iteration

$$\begin{cases} y_{k+1} = R_f R_g (y_k) \\ x_k = \mathrm{Prox}_g^\eta (y_k) \end{cases}, \quad \forall \eta > 0 \text{ converges and}$$

$$\min_{0 \le k \le N} \|x_k - x_*\|^2 \le \frac{1}{N+1} \cdot \frac{1}{4\mu\eta} \|y_0 - y_*\|^2$$

Remark: ① $R_f R_g$ is nonexpansive $\Rightarrow \|y_k - y_*\| \searrow$

But $\|x_k - x_*\|$ is not $\searrow$

② This is the worst case rate.

Proof:

$$\|R_f[R_g(y_k)] - R_f[R_g(y_*)]\|^2 + 4\mu\eta \|\mathrm{Prox}_g^\eta(y_k) - \mathrm{Prox}_g^\eta(y_*)\|^2$$

$$\le \|R_g(y_k) - R_g(y_*)\|^2 + 4\mu\eta \|\mathrm{Prox}_g^\eta(y_k) - \mathrm{Prox}_g^\eta(y_*)\|^2$$

$$\le \|y_k - y_*\|^2$$

$$\Rightarrow \|y_{k+1} - y_*\|^2 + 4\mu\eta \|x_k - x_*\|^2 \le \|y_k - y_*\|^2$$

$$\Rightarrow \quad 4\mu\eta \sum_{k=0}^{N} \|x_k - x_*\|^2 \le \|y_0 - y_*\|^2 - \|y_{N+1} - y_*\|^2$$

$$\le \|y_0 - y_*\|^2$$

Theorem (Peaceman-Rachford splitting with strong convexity)

Assume $f(x)$ is convex and $g(x)$ is strongly convex with $\mu>0$. Then $f(x)+g(x)$ has a unique minimizer x_*. The iteration

$$\begin{cases} z_{k+1} = R_g R_f (z_k) \\ w_{k+1} = \mathrm{Prox}_g^{\eta}(R_f(z_k)) \end{cases}, \quad \forall \eta>0 \text{ converges and}$$

$$\min_{0\le k\le N} \|w_k - x_*\|^2 \le \frac{1}{N}\cdot\frac{1}{4\mu\eta}\|R_f(z_0) - y_*\|^2$$

Proof:

$$\cdots\ \underbrace{R_f \overbrace{R_g \underbrace{R_f \overbrace{R_g \underbrace{R_f \overbrace{R_g(y_0)}^{z_0}}_{y_1}}^{z_1}}_{y_2}}^{z_2}}_{y_3}$$

$$\begin{cases} R_f z_k = y_{k+1} \\ R_g y_k = z_k \end{cases}$$

Theorem [rate of General Douglas-Rachford]

Assume ① $f(x)$ and $g(x)$ are convex,

$0<\theta<1$

② $f(x)+g(x)$ has a minizer $\quad (1-\theta)I + \theta R_f R_g$

then the iteration

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

$$\begin{cases} y_{k+1} = (1-\lambda)y_k + \lambda \dfrac{I+R_fR_g}{2}(y_k) \\ x_{k+1} = \mathrm{Prox}_g^{\eta}(y_{k+1}) \end{cases} \quad \text{converges,}$$

$\forall \eta>0$

$\lambda\in(0\ 2$

on $\{x_k\}$ converges to one minimizer of $f(x)+g(x)$ with

1) $\|y_{k+1}-y_k\|^2 \le \frac{1}{(k+1)}\left(\frac{1}{1-\lambda/2}-1\right)\|y_0-y_*\|^2$

2) $\|x_{k+1}-x_k\|^2 \le \frac{1}{(k+1)}\left(\frac{1}{1-\lambda/2}-1\right)\|y_0-y_*\|^2$

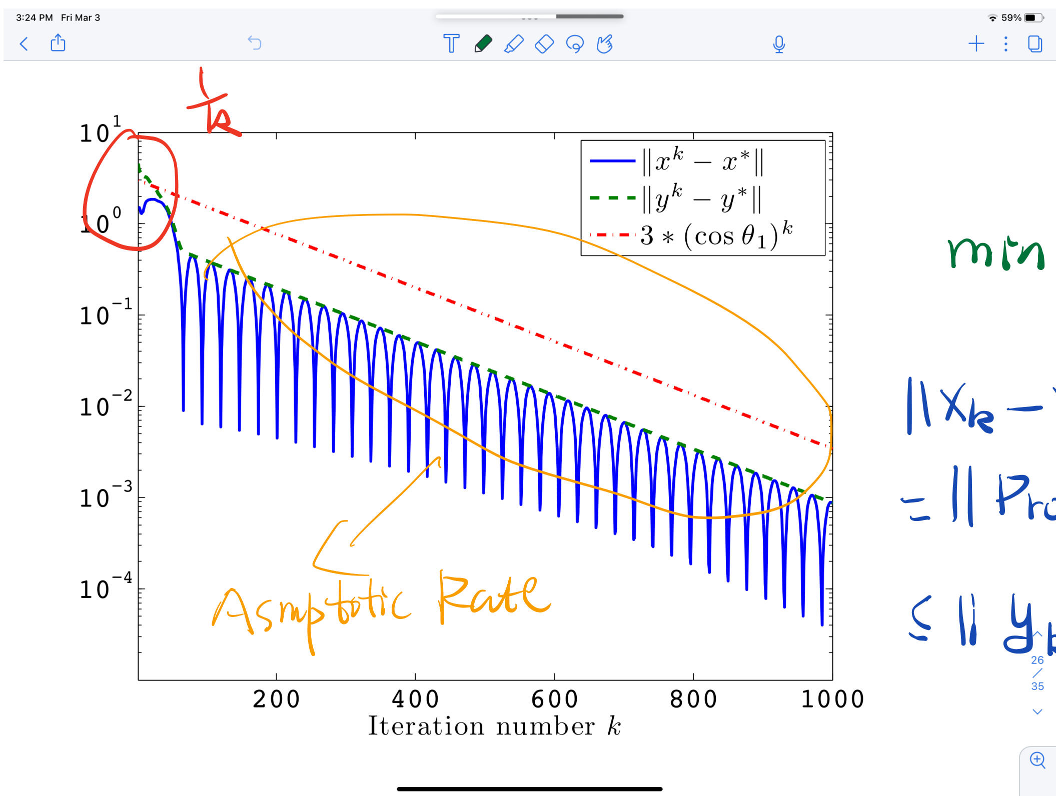
Proof: It is trivially implied by the fixed point iteration:

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



$\min \|x\|_1 + \iota_{\{Ax=b\}}$

$\|x_k - x_*\|$

$= \|\mathrm{Prox}_g(y_k) - \mathrm{Prox}_g(y_*)\|$

$\le \|y_k - y_*\|$

Provable linear rate if $\begin{cases} f(x) \text{ is strongly convex with } \mu>0 \\ \nabla f(x) \text{ is L-cont with } L \end{cases}$

$$\|\nabla f(x) - \nabla f(y)\|^2 \leq L^2 \|x-y\|^2$$

$$\Rightarrow \|(I+\eta\nabla f)(x) - (I+\eta\nabla f)(y)\|^2$$

$$= \|x-y\|^2 + \eta^2\|\nabla f(x) - \nabla f(y)\|^2 + 2\eta\langle x-y, \nabla f(x) - \nabla f(y)\rangle$$

$$\leq \|x-y\|^2 + \eta^2 L^2 \|x-y\|^2 + 2\eta\|x-y\| \cdot \|\nabla f(x) - \nabla f(y)\|$$

$$\leq [1+\eta^2 L^2 + 2\eta L] \|x-y\|^2$$

$(I+\eta\nabla f)(x) = u$

$(I+\eta\nabla f)(y) = v$

$$\Rightarrow \|\mathrm{Prox}_f^\eta(u) - \mathrm{Prox}_f^\eta(v)\|^2 \geq \frac{1}{[1+\eta L]^2}\|x-y\|^2$$

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

$$\|R_f(x) - R_f(y)\|^2 + 4\mu\eta \|\mathrm{Prox}_f^\eta(x) - \mathrm{Prox}_f^\eta(y)\|^2 \leq \|x-y\|^2$$

$R_f(x) = 2\,\mathrm{Prox}_f^\eta(x) - x$

$\Downarrow$

$$\|R_f(x) - R_f(y)\|^2 + 4\mu\eta \|\mathrm{Prox}_f^\eta(x) - \mathrm{Prox}_f^\eta(y)\|^2 - \|x-y\|^2 \leq 0$$

$$\Rightarrow \|R_f(x) - R_f(y)\|^2 + \left[\frac{4\mu\eta}{[1+\eta L]^2} - 1\right]\|x-y\|^2 \leq 0$$

$$\Rightarrow \|R_f(x) - R_f(y)\|^2 \leq \left[1 - \frac{4\mu\eta}{[1+\eta L]^2}\right]\|x-y\|^2$$

Recall $\mu \le L$, So $1 - \frac{4\mu\eta}{[1+\eta L]^2} \geq 1 - \frac{4\eta L}{[1+\eta L]^2} = \frac{[1-\eta L]^2}{[1+\eta L]^2}$

Lemma Assume $\begin{cases} f(x) \text{ is strongly convex with } \mu > 0 \\ \nabla f(x) \text{ is } L\text{-cont with } L \end{cases}$

$$\|R_f(x) - R_f(y)\|^2 \leq \left[1 - \frac{4\mu\eta}{[1+\eta L]^2}\right] \|x-y\|^2$$

where $1 - \frac{4\mu\eta}{[1+\eta L]^2} \geq \frac{[1-\eta L]^2}{[1+\eta L]^2}$

So R_f is a contraction

$\Rightarrow \begin{cases} R_f R_g \text{ is a contraction} \\ \frac{I + R_f R_g}{2} \text{ is a contraction} \\ (1-\lambda) I + \lambda \frac{I + R_f R_g}{2} \text{ is a contraction}, \lambda \in (0,2) \end{cases}$

$\boxed{\text{Theorem}}$ (Peaceman-Rackford splitting Lions & Mercier 1979)

Assume ① $f(x)$ and $g(x)$ are convex

② $g(x)$ is strongly convex with $\mu > 0$

③ $\nabla g(x)$ is Lipschitz continuous with $L > 0$.

then the iteration

$$\begin{cases} y_{k+1} = R_f R_g (y_k) \\ x_{k+1} = \mathrm{Prox}_g^{\eta}(y_{k+1}) \end{cases}$$

converges, $\forall \eta > 0$

and $\{x_k\}$ converges to the minimizer of $f(x)+g(x)$.

1) $\|y_{k+1} - y_*\| \leq \left(\sqrt{1 - \frac{4\mu\eta}{[1+\eta L]^2}}\right) \|y_k - y_*\|$

2) $\|x_k - x_*\|^2 \leq \left(1 - \frac{4\mu\eta}{[1+\eta L]^2}\right)^k \|y_0 - y_*\|^2$

Remark: $\{\|y_k - y_*\|\}$ is $\searrow$

$\{\|x_k - x_*\|\}$ is not monotone!

Proof:

$C = \sqrt{1 - \frac{4\mu\eta}{(1+\eta L)^2}}$

① $\|R_g(x) - R_g(y)\| \leq C \|x - y\|$

② R_g is a contraction $\Rightarrow$ So is $R_f R_g$

$\Rightarrow \|R_f R_g (y_k) - R_f R_g (y_*)\| \leq C \|y_k - y_*\|$

$\Rightarrow \|y_{k+1} - y_*\| \leq C \|y_k - y_*\|$

$\Rightarrow \|y_k - y_*\| \leq C^k \|y_0 - y_*\|$

③ $\|x_k - x_*\| = \|\mathrm{Prox}_g^{\eta}(y_k) - \mathrm{Prox}_g^{\eta}(y_*)\|$

$\leq \|y_k - y_*\| \leq C^k \|y_0 - y_*\|$

$\boxed{\text{Theorem}}$ (Linear Rate of Generalized Douglas-Rachford)

Assume ① $f(x)$ and $g(x)$ are convex

② $g(x)$ is strongly convex with $\mu > 0$

③ $\nabla g(x)$ is Lipschitz continuous with $L > 0$.

then the iteration

$$\begin{cases} y_{k+1} = (1-\lambda) y_k + \lambda \dfrac{I + R_f R_g}{2}(y_k) \\ x_{k+1} = \operatorname{Prox}_g^{\eta}(y_{k+1}) \end{cases}$$

converges, $\forall \eta > 0$, $\lambda \in (0, 2]$

and $\{x_k\}$ converges to one minimizer of $f(x)+g(x)$:

1) $\|y_{k+1} - y_*\| \le \left[1 - \frac{\lambda}{2} + \frac{\lambda}{2} c\right]^k \|y_k - y_*\|$ $\quad c = \sqrt{1 - \frac{4\mu\eta}{(1+\eta L)^2}}$

2) $\|x_k - x_*\|^2 \le \left(\left[1 - \frac{\lambda}{2} + \frac{\lambda}{2} c\right]^2\right)^k \|y_0 - y_*\|^2$

Remark: Be aware that actual convergence rate can be faster than this provable rate.

In other words, $\lambda = 2$ is not necessarily the best in practice.

Proof: $T = R_f R_g$ is a contraction with

$$\|T(x) - T(y)\| \le c\|x - y\|,$$

$$S = (1-\lambda) I + \lambda \frac{I+T}{2}, \quad \lambda \in (0, 2]$$

$$= \left(1 - \frac{\lambda}{2}\right) I + \frac{\lambda}{2} T$$

$$\|S(x) - S(y)\| \le \left(1 - \frac{\lambda}{2}\right)\|x - y\| + \frac{\lambda}{2}\|T(x) - T(y)\|$$

$$\le \left(1 - \frac{\lambda}{2} + \frac{\lambda}{2} c\right)\|x - y\|$$

$$\boxed{\text{Convergence rate of either } F(x_k)-F(x_*) \text{ or } \|x_k-x_*\|^2}$$

		Convexity	Strong Convexity	
$\min_x f(x)$	∇f			
Gradient Descent		$O(\frac{1}{k})$	$O\left(\left(\left[\frac{L-\mu}{L+\mu}\right]^2\right)^k\right)$	$\eta=\frac{2}{L+\mu}$
Accelerated GD		$O(\frac{1}{k^2})$	$O\left(\left(1-\sqrt{\frac{\mu}{L}}\right)^k\right)$	$\eta=\frac{1}{L}$
$\min_x f(x)$	∂f			
Subgradient Method		$O(\frac{1}{\sqrt{k}})$	$O(\frac{1}{k})$	$\eta\leq\cdots$
Proximal Point Method		$O(\frac{1}{k})$	$O\left(\left[\left(\frac{1}{1+\eta\mu}\right)^2\right]^k\right)$	$\forall\eta>0$
$\min_x f(x)+g(x)$	∂f ∇g			
Proximal Gradient		$O(\frac{1}{k})$	$O\left(\left(1-\frac{\mu}{L}\right)^k\right)$	$\eta=\frac{1}{L}$
Accelerated Prox Grad		$O(\frac{1}{k^2})$	$O\left(\left(1-\sqrt{\frac{\mu}{L}}\right)^k\right)$	$\eta=\frac{1}{L}$
$\min_x f(x)+g(x)$	∂f ∂g			
Generalized Douglas-Rachford	$\lambda\in(0,2)$	$O(\frac{1}{k})$	$O(\frac{1}{k})$	$\forall\eta>0$
Peaceman-Rachford	$\lambda=2$	divergent	$O(\frac{1}{k})$	$\forall\eta>0$
$\min_x f(x)+g(x)$	∂f ∇g			
Generalized Douglas-Rachford	$\lambda\in(0,2]$	$O(\frac{1}{k})$	$O\left(\left[1-\frac{\lambda}{2}+\frac{\lambda}{2}c\right]^{2k}\right)$	$\forall\eta>0$

$c=\sqrt{1-\frac{4\mu\eta}{[1+\eta L]^2}}$ $\forall\eta>0$