

Previously on nonsmooth case:

$f(x)$ and $g(x)$ are convex
 $g(x)$ is differentiable

$$\min_x f(x)$$

Example: $f(x) = \|x\|$,

① Subgradient Method

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

② Proximal Point Method

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

$$\min_x f(x) + g(x)$$

Example: $\|x\|_1 + \frac{\lambda}{2} \|Ax - b\|^2$

③ Proximal Gradient Method

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

④ Fast Proximal Gradient

II

$$\min_x f(x) + g(x)$$

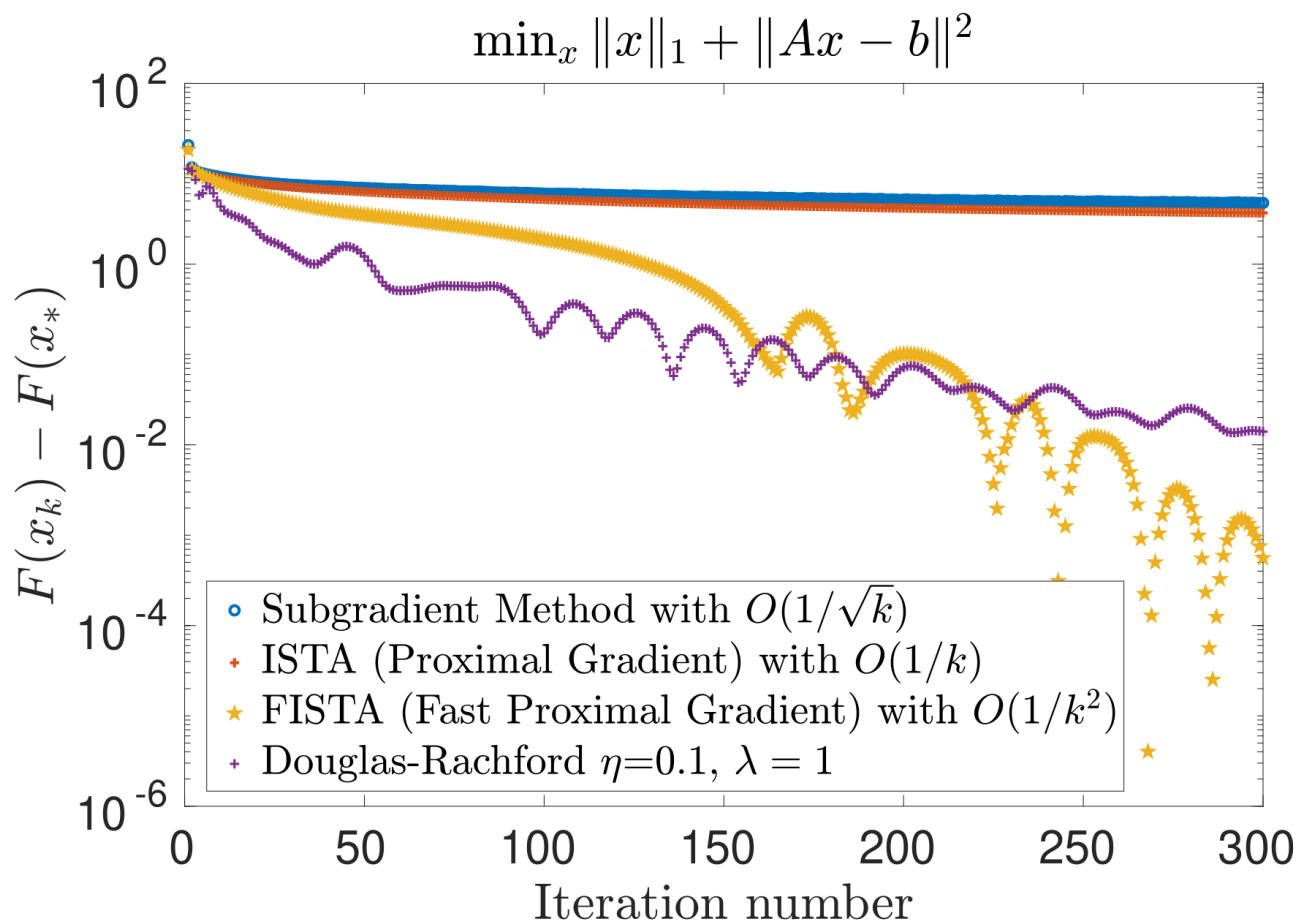
f & g are both nonsmooth

Example: $\min_x \|x\|_1 + \mathbb{1}_{\{x: Ax=b\}}$

⑤ Douglas-Rachford splitting (1979)

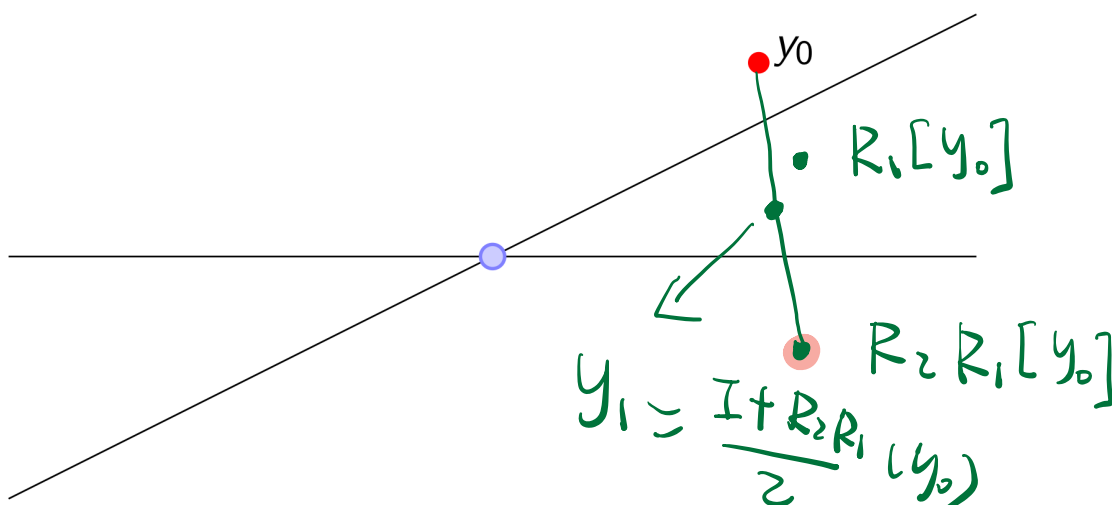
A simple
version

$$\begin{cases} y_{k+1} = \frac{I + R_f R_g}{2} (y_k) \\ R_f = 2 \cdot \text{Prox}_f - I \\ R_g = 2 \cdot \text{Prox}_g - I \end{cases}$$



(a) Douglas-Rachford splitting converges for any step size $\eta > 0$ but proximal operator for $g(\mathbf{x}) = \|A\mathbf{x} - b\|^2$ requires $(I + \eta 2A^T A)^{-1}$.

For $\min_x f(x) + g(x)$, if $f(x)$ and $g(x)$ are indicator functions of two lines respectively, then the minimizers are the intersection point.



Theorem If $f(x)$ & $g(x)$ are convex, the
simple Douglas-Rachford splitting iteration converges.

Proof:

Theorem The following are equivalent:

- ① T is firmly nonexpansive
- ② $I - T$ is firmly nonexpansive
- ③ $2T - I$ is nonexpansive

$f(x)$ is convex $\Rightarrow \text{Prox}_f^\eta$ is firmly nonexpansive

$\Leftrightarrow R_f = 2\text{Prox}_f^\eta - I$ is nonexpansive

$\Rightarrow T = R_f R_g$ is nonexpansive $\|R_f(x) - R_f(y)\| \leq \|x - y\|$

$\Rightarrow Y_{k+1} = [(1-\theta)I + \theta T](Y_k)$ converges

$\theta \in (0, 1)$

$\theta = \frac{1}{2}$ gives simplest DR splitting

$S = \frac{1}{2}I + \frac{1}{2}R_f R_g$ is firmly nonexpansive

$\Uparrow$

$2S - I = R_f R_g$ is nonexpansive

The **general** Douglas-Rachford splitting
is written as

$$y_{k+1} = [\theta I + (1-\theta) R_f R_g] (y_k)$$

$\theta \in (0, 1)$

Theorem

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

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

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

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

Sometimes, general DR is written as

$$y_{k+1} = \left[(1-\lambda)I + \lambda \frac{I + R_f R_g}{2} \right] (y_k), \quad \lambda \in (0, 2)$$

Theorem General DR converges for any convex $f(x)$ & $g(x)$, and any $\eta > 0$.

Proof: $T = \frac{I + R_f R_g}{2}$ T is firmly nonexpansive
 $\Downarrow$

$2T - I$ is nonexpansive

$$S = (1-\lambda)I + \lambda T = (1-2\theta)I + 2\theta T, \quad \theta \in (0, 1) \\ = (1-\theta) \cdot I + \theta \cdot (2T - I)$$

10:39 AM Wed Mar 1

Developments of Douglas-Rachford Splitting $u_t = u_{xx} + u_{yy}$

1. [Peaceman and Rachford, 1955; Douglas and Rachford, 1956](#): implicit finite difference in solving heat equations. $\rightarrow$ ADI
2. [Lions and Mercier, 1979](#): extension to maximal monotone operators; DR is firmly nonexpansive thus convergent.
3. [Glowinski and Marroco, 1975; Gabay, 1983](#): the alternating direction method of multipliers (ADMM) is equivalent to DR. ADMM has been widely used in nonlinear mechanics and convex optimization.
4. [Goldstein and Osher, 2009](#): split Bregman method widely used in image processing problems, which is the same as ADMM.
5. [Bauschke, Combettes, and Luke, 2002](#): the widely used Fienup's Hybrid Input-Output (1982) algorithm for phase retrieval problem can be viewed as Douglas-Rachford splitting.

The following three algorithms are **exactly the same**:

- Douglas-Rachford Splitting on $\min_x f(x) + g(x)$.
- ADMM on Fenchel Dual $\min_y f^*(y) + g^*(-y)$.
- Split Bregman on Fenchel Dual $\min_y f^*(y) + g^*(-y)$.

20 / 31

③ Douglas-Rachford Splitting (Lions & Mercier 1979)

$f(x)$ is convex $\Rightarrow \text{Prox}_f^\eta$ is firmly nonexpansive

$\Rightarrow R_f = 2\text{Prox}_f^\eta - I$ is nonexpansive

So $T = R_f R_g$ is nonexpansive

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

Simple Douglas-Rachford Splitting

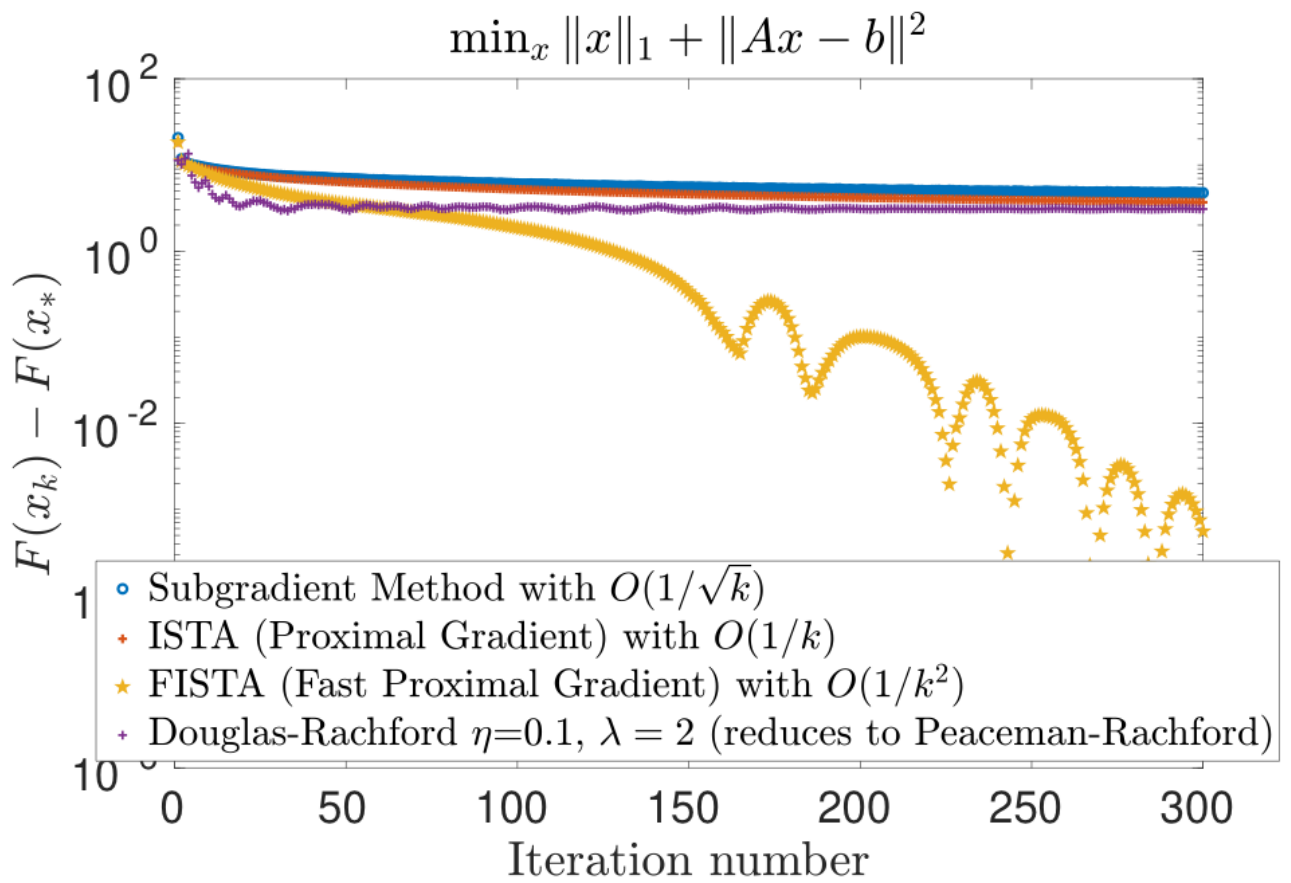
$$\textcircled{1} \quad y_{k+1} = \frac{I + R_f R_g}{2} (y_k) \quad \theta = \frac{1}{2}$$
$$= \theta y_k + (1-\theta) R_f R_g (y_k)$$

$x_{k+1} = \theta x_k + (1-\theta) T(x_k)$ converges if T
 $0 < \theta < 1$ is nonexpansive

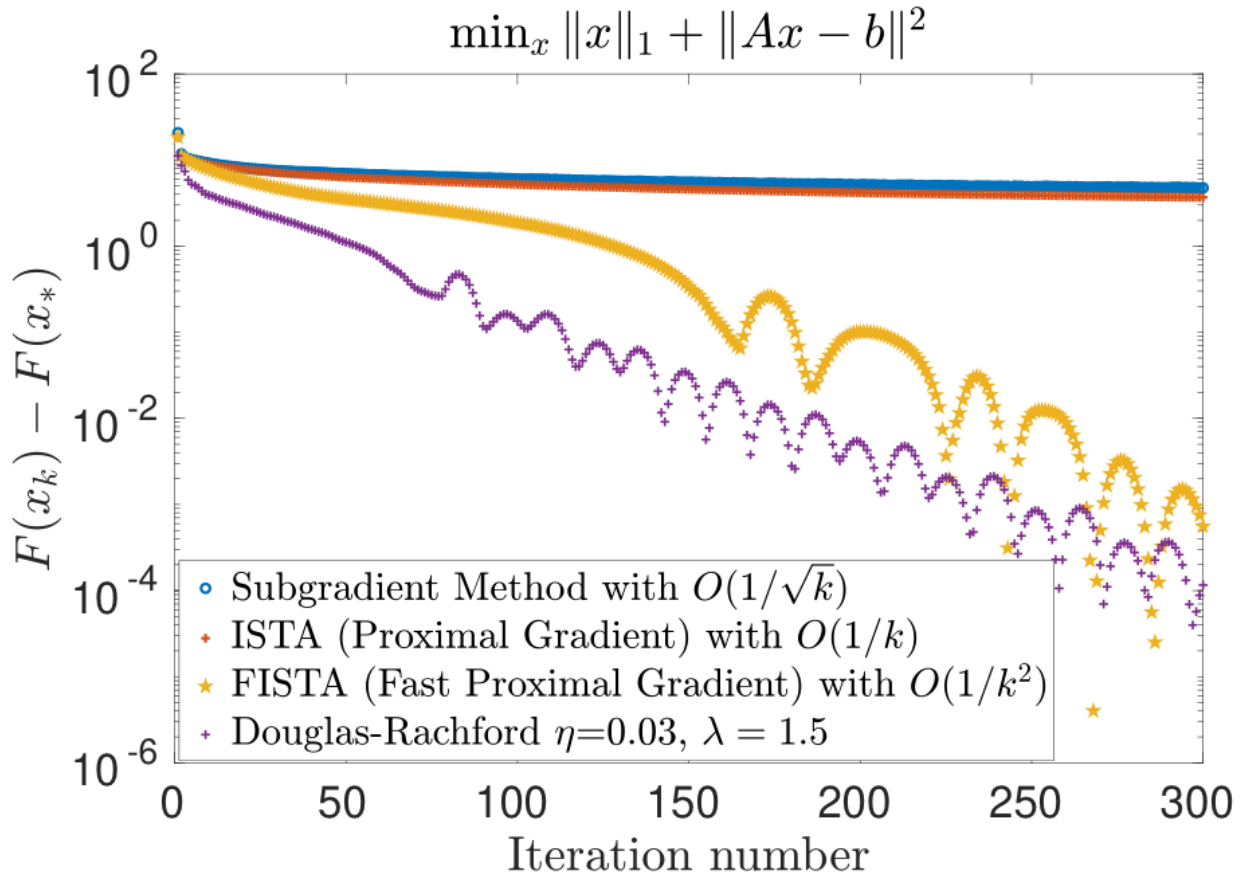
② { General Douglas-Rachford Splitting

$$y_{k+1} = \theta y_k + (1-\theta) R_f R_g (y_k) \quad \theta \in (0, 1)$$

③ $\theta = 0$, $y_{k+1} = R_f R_g (y_k)$ is called
Peaceman-Rachford splitting



(b) If $\lambda = 2$ in (6.1), then it does not converge. For this example $A \in \mathbb{R}^{40 \times 1000}$ thus no strong convexity for $g(\mathbf{x})$.



(c) If tuning parameters, Douglas-Rachford splitting can be faster than FISTA or restarted FISTA for certain accuracy threshold.

IV

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
 $x_* = \text{Prox}_g^\eta(y_*)$ ($I + \eta \partial g$) is multi-valued
($I + \eta \partial g$)⁻¹ is single-valued

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

Remark: y_* is a fixed point of

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

Proof: $0 \in \partial f(x_*) + \partial g(x_*)$
 $\Leftrightarrow 0 \in -\eta \partial f(x_*) - \eta \partial g(x_*)$

$$\Leftrightarrow 0 \in x_* - \eta \partial f(x_*) - \underbrace{[x_* + \eta \partial g(x_*)]}_{y_* \in}$$

$$\Leftrightarrow y_* \in x_* - \eta \partial f(x_*)$$

$$\Leftrightarrow -y_* \in -x_* + \eta \partial f(x_*) \quad y_* \in x_* + \eta \partial g(x_*) = [I + \eta \partial g](x_*)$$

$$\Leftrightarrow 2x_* - y_* \in (I + \eta \partial f)(x_*)$$

$$\Leftrightarrow (I + \eta \partial f)^{-1}(2x_* - y_*) = x_* \quad \Leftrightarrow [I + \eta \partial f]^{-1}(y_*) = x_*$$

$$\Leftrightarrow (I + \eta \partial f)^{-1}(R_g(y_*)) = \text{Prox}_g^\eta(y_*)$$

$$\Leftrightarrow 2(I + \eta \partial f)^{-1}(R_g(y_*)) - [2\text{Prox}_g^\eta(y_*) - y_*] = y_*$$

$$\Leftrightarrow 2(I + \eta \partial f)^{-1}(R_g(y_*)) - R_g(y_*) = y_*$$

$$\Leftrightarrow R_f[R_g(y_*)] = y_*$$

$$y_{k+1} = \frac{I + R_f R_g}{2} (y_k) \text{ converges to a fixed}$$

$$\text{point of } S = \frac{I + R_f R_g}{2} \text{ which is NOT } x_*!$$

$$y_* = \frac{I + R_f R_g}{2} (y_*)$$

$$\Leftrightarrow 2y_* = y_* + R_f R_g(y_*)$$

$$\begin{aligned} \Leftrightarrow y_* &= R_f[R_g(y_*)] \\ &= 2\text{Prox}_f^\eta[R_g(y_*)] - R_g(y_*) \\ &= 2\text{Prox}_f^\eta[R_g(y_*)] - 2\text{Prox}_g^\eta(y_*) + y_* \end{aligned}$$

$$\begin{aligned} \Leftrightarrow \underbrace{\text{Prox}_g^\eta(y_*)}_z &= \underbrace{\text{Prox}_f^\eta[R_g(y_*)]}_{2\text{Prox}_g^\eta(y_*) - y_* = 2z - y_*} \\ (I + \eta \partial g)(z) &= y_* \end{aligned}$$

$$\begin{aligned} \Leftrightarrow z &= (I + \eta \partial f)^{-1}[2z - (I + \eta \partial g)(z)] \\ &= (I + \eta \partial f)^{-1}[(I - \eta \partial g)(z)] \end{aligned}$$

$$\Leftrightarrow 0 \in \partial f(z) + \partial g(z)$$

Theorem (Douglas-Rachford Splitting)

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

② $f(x) + g(x)$ has a minimizer

then the iteration

$$\begin{cases} y_{k+1} = \frac{I + R_f R_g}{2}(y_k) \\ x_{k+1} = \text{Prox}_g^{\eta}(y_{k+1}) \end{cases} \text{ converges, } \forall \eta > 0$$

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

Remark: $\{y_k\}$ is an auxiliary variable

Even if x_* is unique, y_* may be non-unique

$$\text{Remark: } y_{k+1} = \frac{I + R_f R_g}{2}(y_k)$$

$$= \frac{1}{2} y_k + \frac{1}{2} [2 \text{Prox}_f [R_g(y_k)] - R_g(y_k)]$$

$$= \text{Prox}_f [2 \text{Prox}_g(y_k) - y_k] - \text{Prox}_g(y_k) + y_k$$

$$= \text{Prox}_f [2x_k - y_k] - x_k + y_k$$

Example: $\min_x \|x\|_1 + i_{\{Ax=b\}}(x) \Leftrightarrow \min_x \|x\|_1$
 $f(x) + g(x)$ s.t. $Ax=b$

A

$$\text{Prox}_f^\eta = S_\eta(x) \quad S_\eta(x)_i = \begin{cases} x_i - \eta & , x_i > \eta \\ x_i + \eta & , x_i < -\eta \\ 0 & , x_i \in [-\eta, \eta] \end{cases}$$

$$\begin{aligned} \text{Prox}_g^\eta &= P(x) \\ &= x + A^T(AA^T)^{-1}(b - Ax) \end{aligned}$$

$$x_k = \text{Prox}_g^\eta(y_k) = y_k + A^T(AA^T)^{-1}(b - Ay_k)$$

$$\begin{aligned} y_{k+1} &= \text{Prox}_f[2x_k - y_k] - x_k + y_k \\ &= S_\eta[2x_k - y_k] - x_k + y_k \end{aligned}$$

Theorem (General Douglas-Rachford)

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

② $f(x) + g(x)$ has a minimizer

then the iteration

$$\begin{cases} y_{k+1} = (1-\lambda)y_k + \lambda \frac{I + R_f R_g}{2}(y_k) \\ x_{k+1} = \text{Prox}_g^\eta(y_{k+1}) \end{cases} \quad \begin{array}{l} \text{converges,} \\ \forall \eta > 0 \\ \lambda \in (0, 2) \end{array}$$

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

- Remark: 1) λ is called Relaxation parameter
- 2) each λ gives a different algorithm
- 3) $\lambda=1$ is Douglas-Rachford
- 4) $\lambda=2$: $y_{k+1} = R_f R_g(y_k)$
is Peaceman-Rachford
which may not converge!