Fast/Accelerated PDHG for $g(x)$ being strongly convex with $\mu>0$

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

$$\begin{cases} \tau_0\eta_0 < \frac{1}{\rho(K^*K)}, \quad \bar{x}_0 = x_0 \\ y_{k+1} = (I+\tau_k\partial f^*)^{-1}[y_k+\tau_k K\bar{x}_k] \\ x_{k+1} = (I+\eta_k\partial g)^{-1}[x_k-\eta_k K^* y_{k+1}] \\ \theta_k = \frac{1}{\sqrt{1+2\mu\eta_k}}, \quad \eta_{k+1}=\theta_k\eta_k, \quad \tau_{k+1}=\frac{\tau_k}{\theta_k} \\ \bar{x}_k = x_{k+1}+\theta_k(x_{k+1}-x_k) \end{cases}$$

Step size rule $\Leftrightarrow (1+2\mu\eta_n)\frac{\eta_{n+1}}{\eta_n} = \frac{\tau_{n+1}}{\tau_n} = \frac{1}{\theta_n} = \frac{\eta_n}{\eta_{n+1}}$

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

$$(*)\begin{cases} [\langle Kx, y_*\rangle - f^*(y_*)+g(x)] \\ -[\langle Kx_*, y\rangle - f^*(y)+g(x_*)] \\ = g(x)-g(x_*)-\langle -K^*y_*, x-x_*\rangle \\ +f^*(y)-f^*(y_*) - \langle Kx_*, y-y_*\rangle \geq \frac{\mu}{2}\|x-x_*\|^2 \\ \text{because} \begin{cases} -K^*y_* \in \partial g(x_*) \Leftarrow \frac{\partial L}{\partial x}(x_*,y_*)=0 \\ Kx_* \in \partial f^*(y_*) \Leftarrow \frac{\partial L}{\partial y}(x_*,y_*)=0 \end{cases} \\ L(x,y) = \langle Kx,y\rangle - f^*(y)+g(x) \end{cases}$$

Sketchy Proof on acceleration:

$$\begin{cases} x_{k+1} = (I+\eta\partial g)^{-1}[x_k-\eta K^*y_k] \\ y_{k+1} = (I+\tau\partial f^*)^{-1}[y_k+\tau K\bar{x}_{k+1}] \end{cases}$$

$$\begin{cases} x_{k+1} + \eta \partial g(x_{k+1}) \ni x_k - \eta K^* y_k \\ y_{k+1} + \tau \partial f^*(y_{k+1}) \ni y_k + \tau K \bar{x}_{k+1} \end{cases} \qquad \bar{x}_{k+1} = x_{k+1} + \theta(x_{k+1} - x_k)$$

$$\Rightarrow \begin{cases} \partial g(x_{k+1}) \ni \dfrac{x_k - x_{k+1}}{\eta} - K^* y_k \\ \partial f^*(y_{k+1}) \ni \dfrac{y_k - y_{k+1}}{\tau} + K \bar{x}_{k+1} \end{cases}$$

$$\Rightarrow \begin{cases} g(x) \geq g(x_{k+1}) + \left\langle \dfrac{x_k - x_{k+1}}{\eta} - K^* y_k, x - x_{k+1} \right\rangle + \dfrac{\mu}{2}\|x - x_{k+1}\|^2 \\ f^*(y) \geq f^*(y_{k+1}) + \left\langle \dfrac{y_k - y_{k+1}}{\tau} + K \bar{x}_{k+1}, y - y_{k+1} \right\rangle \end{cases}$$

$$\Rightarrow \begin{cases} g(x) \geq g(x_{k+1}) + \left\langle \dfrac{x_k - x_{k+1}}{\eta}, x - x_{k+1} \right\rangle - \langle K^* y_k, x - x_{k+1} \rangle + \dfrac{\mu}{2}\|x - x_{k+1}\|^2 \\ f^*(y) \geq f^*(y_{k+1}) + \left\langle \dfrac{y_k - y_{k+1}}{\tau}, y - y_{k+1} \right\rangle + \langle K \bar{x}_{k+1}, y - y_{k+1} \rangle \end{cases}$$

subgradient definition

$$\bar{x}_{k+1} = x_{k+1} + \theta_k (x_{k+1} - x_k)$$

Add two :

$$\frac{\|y - y_k\|^2}{2\tau} + \frac{\|x - x_k\|^2}{2\eta} \geq \left[\langle K x_{k+1}, y \rangle - f^*(y) + g(x_{k+1})\right] - \left[\langle K x, y_{k+1} \rangle - f^*(y_{k+1}) + g(x)\right]$$

$$+ \frac{\mu}{2}\|x - x_{k+1}\|^2 + \frac{\|y - y_{k+1}\|^2}{2\tau} + \frac{\|x - x_{k+1}\|^2}{2\eta} + \frac{\|y_k - y_{k+1}\|^2}{2\tau} + \frac{\|x_k - x_{k+1}\|^2}{2\eta}$$

$$+ \theta \langle K(x_k - x_{k+1}), y_{k+1} - y \rangle - \langle K(x_{k+1} - x), y_{k+1} - y_k \rangle$$

$$= \theta \langle K(x - x_{k+1}), y_{k+1} - y \rangle + \theta \langle K(x_k - x), y_k - y \rangle + \theta \langle K(x_k - x), y_{k+1} - y_k \rangle$$

$$\Rightarrow \quad L(x_{k+1}, y) - L(x, y_{k+1}) + \frac{\mu}{2}\|x - x_{k+1}\|^2$$

$$+ \frac{\|y - y_{k+1}\|^2}{2\tau} + \frac{\|x - x_{k+1}\|^2}{2\eta} - \theta\langle K(x - x_{k+1}), y - y_{k+1}\rangle$$

$$+ \frac{\|y_k - y_{k+1}\|^2}{2\tau} + \frac{\|x_k - x_{k+1}\|^2}{2\eta} - \langle K(x_{k+1} - x_k), y_{k+1} - y_k\rangle$$

$$+ (1-\theta)\langle K(x - x_k), y_{k+1} - y_k\rangle$$

$$\leq \frac{\|y - y_k\|^2}{2\tau} + \frac{\|x - x_k\|^2}{2\eta} - \theta\langle K(x - x_k), y - y_k\rangle$$

Plug in $(x, y) = (x_*, y_*)$, then

$$(*) \quad \Rightarrow \quad L(x_{k+1}, y_*) - L(x_*, y_{k+1}) \geq \frac{\mu}{2}\|x_* - x_{k+1}\|^2$$

$$\Rightarrow \quad \frac{\mu}{2}\|x_* - x_{k+1}\|^2 + \frac{\mu}{2}\|x_* - x_{k+1}\|^2$$

$$+ \frac{\|y_* - y_{k+1}\|^2}{2\tau_n} + \frac{\|x_* - x_{k+1}\|^2}{2\eta_n} - \theta_n\langle K(x_* - x_{k+1}), y_* - y_{k+1}\rangle$$

if $\tau_n \eta_n < \frac{1}{\|K\|^2}$

$$\underbrace{+ \frac{\|y_k - y_{k+1}\|^2}{2\tau_n} + \frac{\|x_k - x_{k+1}\|^2}{2\eta_n} - \langle K(x_{k+1} - x_k), y_{k+1} - y_k\rangle}_{\geq 0}$$

$$+ (1-\theta_n)\langle K(x_* - x_k), y_{k+1} - y_k\rangle$$

$$\leq \frac{\|y_* - y_k\|^2}{2\tau_n} + \frac{\|x_* - x_k\|^2}{2\eta_n} - \theta_n\langle K(x_* - x_k), y_* - y_k\rangle$$

$$\Rightarrow \quad \cdots \quad \Rightarrow$$

$$\|x_n - x_*\|^2 \leq \eta_n^2 \left(\frac{\|x_0 - x_*\|^2}{\eta_0^2} + \|K\|^2 \|y_0 - y_*\|^2 \right)$$

Then show $\eta_n \sim O(\frac{1}{n})$

Remark: ① Consider $\min_x F(Kx) + G(x)$

$$\Longleftrightarrow -\min_y F^*(y) + G^*(-K^* y)$$

If G is not strongly convex but F^* is strongly convex, then just apply Fast PDHG with $\begin{cases} g = F^* \\ f = G^* \end{cases}$

② If both f^* and g are strongly convex, similar to fast Proximal grad, one version can achieve linear rate.

Example: $\min_{x \in \mathbb{R}^n} \|x\|_1 + i_{\{x : Ax = b\}}(x)$

$$\begin{array}{cc} & n \\ m & \boxed{A} \end{array} \; x = b$$

$$f(Ax) = i_{\{x : Ax = b\}}(x)$$

$$f(y) = i_{\{y : y = b\}}(y) \qquad y \in \mathbb{R}^m$$

$$g(x) = \|x\|_1$$

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

1) PDHG is

$$\begin{cases} x_{k+1} = (I + \eta \partial g)^{-1} [x_k - \eta K^* y_k] \\ y_{k+1} = (I + \tau \partial f^*)^{-1} [y_k + \tau K(2x_{k+1} - x_k)] \end{cases}$$

$$\Rightarrow \begin{cases} x_{k+1} = S_\eta (x_k - \eta A^T y_k) \\ y_{k+1} = y_k + \tau A(2x_{k+1} - x_k) - \tau b \end{cases}$$

Moreau's: $\mathrm{Prox}_f^\eta(x) + \eta \, \mathrm{Prox}_{f^*}^{1/\eta}\left(\frac{x}{\eta}\right) = x$

2) $F(x) = f(Ax) = i_{\{x : Ax = b\}}(x)$

$G(x) = \|x\|_1$

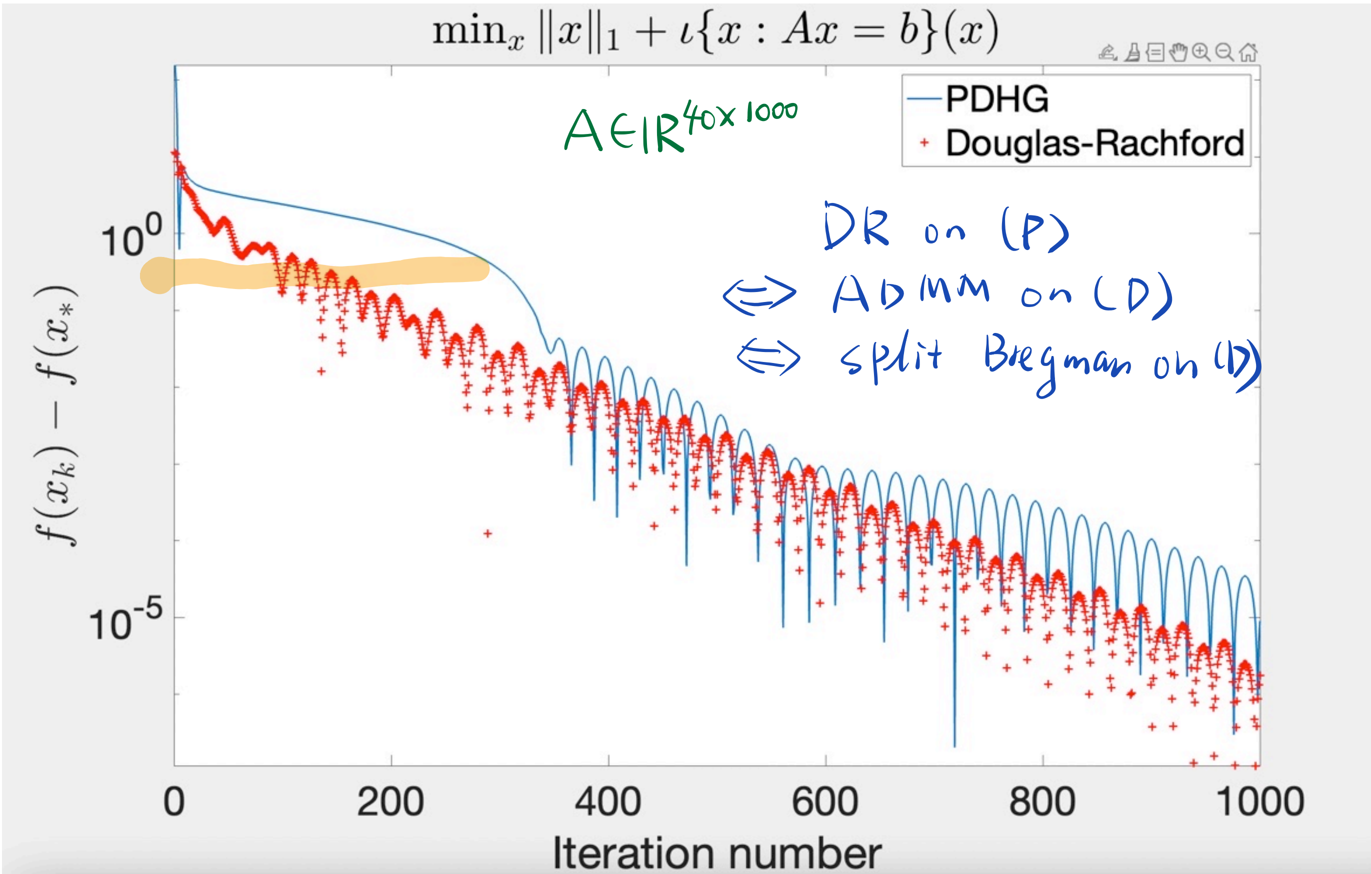
$\min_x F(x) + G(x) \quad (P)$

DR: $z_{k+1} = \dfrac{z_k + R_F R_G(z_k)}{2}$

$R_F = 2\mathrm{Prox}_F^\eta - I$

Douglas-Rachford on $(P) \Leftrightarrow$ ADMM on (D)

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



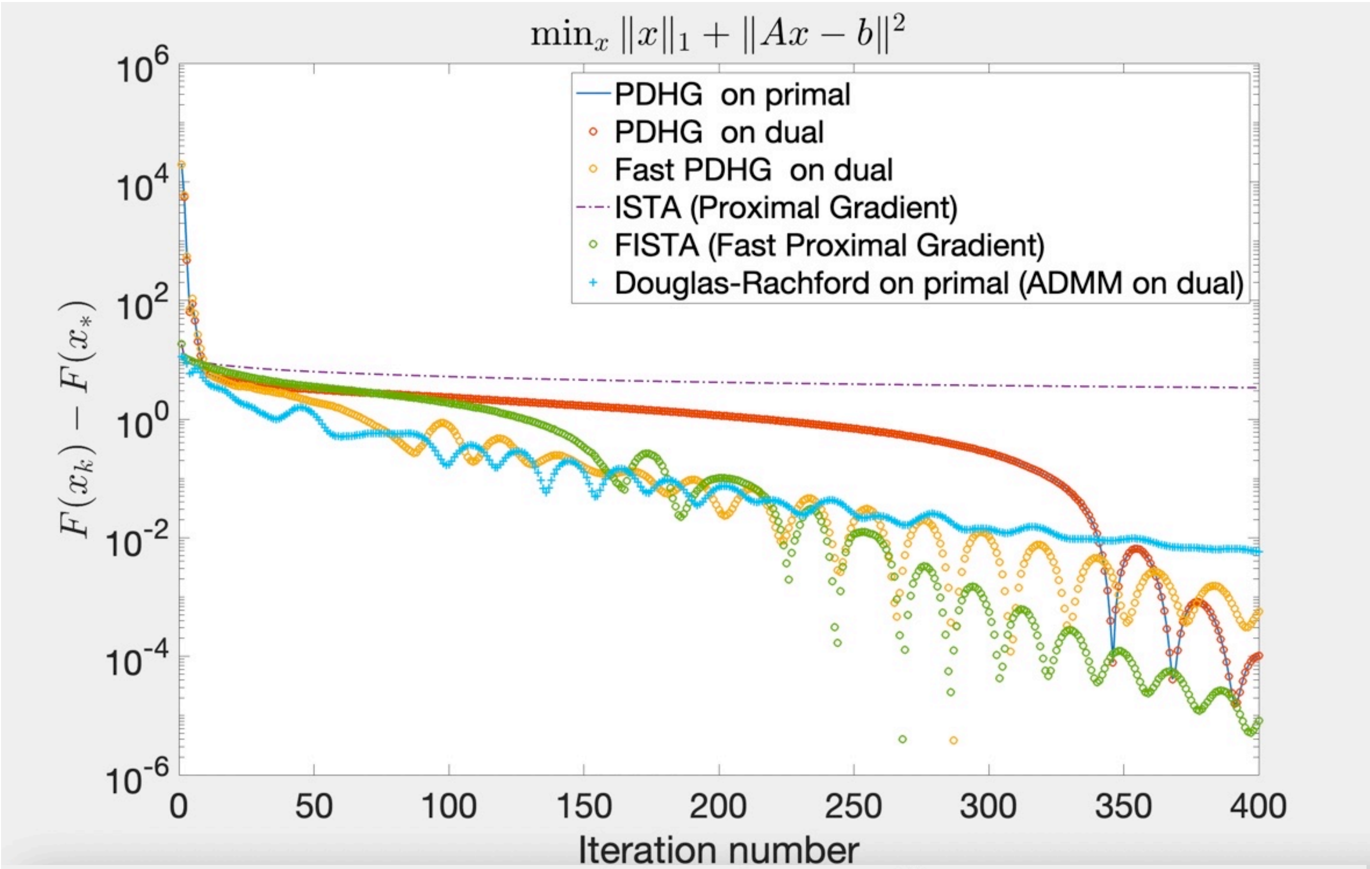
Question: which one is better?

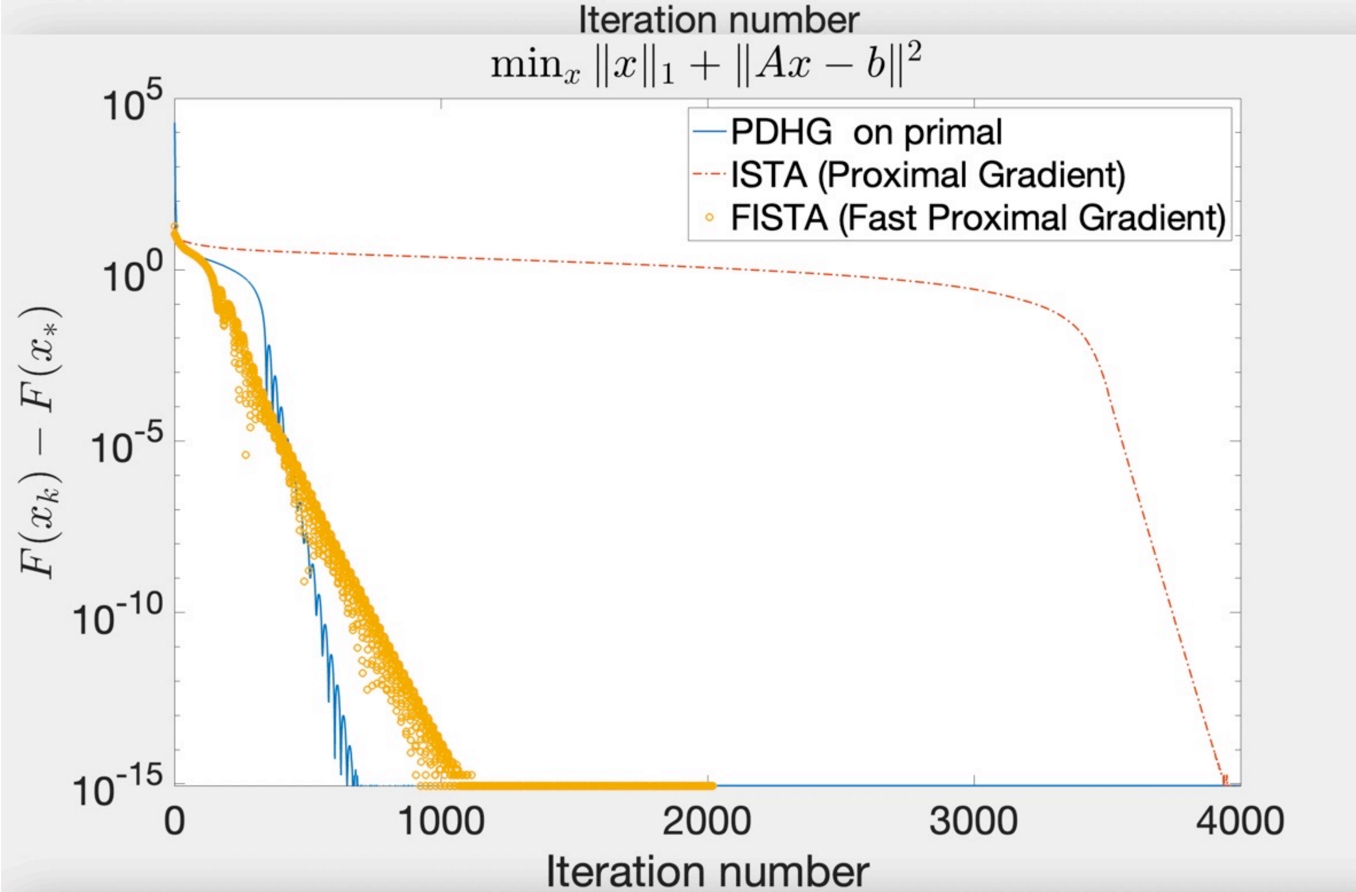
In what sense?

Remark: DR on (P) : $(AA^T)^{-1}$

DR on (D) : $(I + \eta A^T A)^{-1}$

A similar but easier problem





Questions & Remarks:

① which ones are easier to implement?

② which one has the most computational cost per iteration?

③ which ones has no parameter constraint? e.g., arbitrarily large step size can be used.

$$\|x\|_1 + \frac{1}{2}\|Ax-b\|^2 \qquad f(y) = \frac{1}{2}\|y-b\|^2$$

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

$$\Longleftrightarrow \; -\min_y \; f^*(y) + g^*(-A^T y)$$