

Theorem [$O(\frac{1}{k^2})$ rate of the Accelerated Method]

$$\min_x f(x) + g(x) \quad F(x) = f(x) + g(x)$$

Assume $\begin{cases} \textcircled{1} f(x) \text{ is convex} \\ \textcircled{2} g(x) \text{ is convex} \\ \textcircled{3} \nabla g(x) \text{ is } L\text{-cont.} \end{cases}$

Fast/Accelerated Proximal Gradient Method

$$\begin{cases} x_0 = y_0, t_0 = 1, t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2} \\ x_{k+1} = \text{Prox}_f^\eta(y_k - \eta \nabla g(y_k)), \quad \eta = \frac{1}{L} \\ y_{k+1} = x_{k+1} + \frac{t_k - 1}{t_{k+1}} (x_{k+1} - x_k) \end{cases}$$

$\begin{cases} t_k^2 - t_k \leq t_{k-1} \\ t_k \geq 1 \end{cases}$

$$F(x_k) - F(x_*) \leq \frac{1}{(k+1)^2} 2L$$

Proof:

[Prox-Grad Inequality]

Let $\bar{y} = \text{Prox}_f^\eta(y - \eta \nabla g(y))$, and $\eta \leq \frac{1}{L}$

$$\begin{aligned} F(x) - F(\bar{y}) &\geq \frac{L}{2} \|x - \bar{y}\|^2 - \frac{L}{2} \|x - y\|^2 \\ &\quad + \underbrace{g(x) - g(y) - \langle \nabla g(y), x - y \rangle}_{\geq 0} \end{aligned}$$

$\Downarrow$

$$F(x) - F(x_{k+1}) \geq \frac{L}{2} \|x - x_{k+1}\|^2 - \frac{L}{2} \|x - y_k\|^2$$

$$x_{k+1} = \text{Prox}_f^\eta(y_k - \eta \nabla g(y_k))$$

$$x = \frac{1}{t_k} x_* + (1 - \frac{1}{t_k}) x_k$$

$$F(\frac{1}{t_k} x_* + (1 - \frac{1}{t_k}) x_k) - F(x_{k+1})$$

$$\geq \frac{L}{2} \|\frac{1}{t_k} x_* + (1 - \frac{1}{t_k}) x_k - x_{k+1}\|^2 - \frac{L}{2} \|\frac{1}{t_k} x_* + (1 - \frac{1}{t_k}) x_k - y_k\|^2$$

$$= \frac{L}{2t_k^2} \|\underbrace{t_k x_{k+1} - [x_* + (t_k - 1)x_k]}_{u_{k+1}}\|^2 - \frac{L}{2t_k^2} \|\underbrace{t_k y_k - [x_* + (t_k - 1)x_k]}_{\parallel}\|^2$$

$$= \frac{L}{2t_k^2} \|\underbrace{t_k [x_k + \frac{t_{k-1} - 1}{t_k} (x_k - x_{k-1})] - [x_* + (t_k - 1)x_k]}_{\parallel}\|^2$$

$$= \frac{L}{2t_k^2} \|\underbrace{t_{k-1} x_k - [x_* + (t_{k-1} - 1)x_{k-1}]}_{u_k}\|^2$$

$$F \text{ is convex} \Rightarrow F(\frac{1}{t_k} x_* + (1 - \frac{1}{t_k}) x_k) \leq \frac{1}{t_k} F(x_*) + (1 - \frac{1}{t_k}) F(x_k)$$

$$\Rightarrow \frac{1}{t_k} F(x_*) + (1 - \frac{1}{t_k}) F(x_k) - F(x_{k+1}) \geq \frac{L}{2t_k^2} \|u_{k+1}\|^2 - \frac{L}{2t_k^2} \|u_k\|^2$$

$$\Rightarrow \underbrace{t_k F(x_*) + (t_k^2 - t_k) F(x_k) - t_k^2 F(x_{k+1})}_{\geq \frac{L}{2} \|u_{k+1}\|^2 - \frac{L}{2} \|u_k\|^2} \geq \frac{L}{2} \|u_{k+1}\|^2 - \frac{L}{2} \|u_k\|^2$$

$$-(t_k^2 - t_k) F_* + (t_k^2 - t_k) F_k + t_k^2 F_* - t_k^2 F_{k+1}$$

$$R_k = F_k - F_*$$

$$(t_k^2 - t_k) R_k - t_k^2 R_{k+1} \geq \frac{L}{2} \|u_{k+1}\|^2 - \frac{L}{2} \|u_k\|^2$$

$$\text{We only need } t_k^2 - t_k \leq t_{k-1}^2$$

$$\Rightarrow t_{k-1}^2 R_k - t_k^2 R_{k+1} \geq \frac{L}{2} \|u_{k+1}\|^2 - \frac{L}{2} \|u_k\|^2$$

$$\Rightarrow t_{k-1}^2 R_k + \frac{L}{2} \|U_k\|^2 \geq t_k^2 R_{k+1} + \frac{L}{2} \|U_{k+1}\|^2$$

$$\begin{aligned} \Rightarrow \frac{L}{2} \|U_k\|^2 + t_{k-1}^2 R_k &\leq \frac{L}{2} \|U_1\|^2 + t_0^2 R_1 \\ &= \frac{L}{2} \|X_1 - X_*\|^2 + (F_1 - F_*) \end{aligned}$$

$$\boxed{U_k = t_{k-1} X_k - [X_* + (t_{k-1} - 1) X_{k-1}] \quad t_0 = 1} \Rightarrow U_1 = X_1 - X_*$$

$$\Rightarrow t_{k-1}^2 R_k \leq \frac{L}{2} \|X_1 - X_*\|^2 + (F_1 - F_*)$$

$$\boxed{\begin{array}{l} \text{Prox-Grad Inequality} \\ F(\underset{X_*}{x}) - F(\underset{X_1}{\bar{y}}) \geq \frac{L}{2} \|\underset{X_* - X_1}{x} - \underset{X_* - y_0}{\bar{y}}\|^2 - \frac{L}{2} \|x - y\|^2 \\ \hspace{15em} X_0 = y_0 \end{array}}$$

$$\Rightarrow F_* - F_1 \geq \frac{L}{2} \|X_1 - X_*\|^2 - \frac{L}{2} \|X_0 - X_*\|^2$$

$$\Rightarrow F_1 - F_* \leq -\frac{L}{2} \|X_1 - X_*\|^2 + \frac{L}{2} \|X_0 - X_*\|^2$$

$$\Rightarrow t_{k-1}^2 R_k \leq \frac{L}{2} \|X_1 - X_*\|^2 + (F_1 - F_*) \leq \frac{L}{2} \|X_0 - X_*\|^2$$

$$\left. \begin{array}{l} t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2} \\ t_0 = 1 \end{array} \right\} \Rightarrow t_k \geq \frac{k+2}{2}$$

$$\Rightarrow R_k \leq 2L \|X_0 - X_*\|^2 \cdot \frac{1}{(k+1)^2}$$

Example: $\min_x f(x)$ such that $x_i \geq 0, \forall i$

Define $S = \{x \in \mathbb{R}^n : x_i \geq 0\}$.

$$\min_x f(x) + \mathcal{I}_S(x)$$

S is convex $\Rightarrow \mathcal{I}_S(x) = \begin{cases} 0 & x \in S \\ +\infty & x \notin S \end{cases}$ is convex

Projected Gradient Method (Proximal Gradient)

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

$$P_S(x)_i = \begin{cases} x_i & , x_i \geq 0 \\ 0 & , x_i < 0 \end{cases}$$

$$\textcircled{1} F(x_k) - F(x_{k+1}) = \frac{1}{k}$$

$$\textcircled{2} \|x_{k+1} - x_k\| \leq \|x_k - x_{k-1}\|$$

$\textcircled{3} x_k \rightarrow \text{one of minimizers}$

Fast Projected Gradient Method

$$\begin{cases} x_{k+1} = P_S(y_k - \eta \nabla f(y_k)) \\ y_{k+1} = x_{k+1} + \frac{k-1}{k+2} (x_{k+1} - x_k) \\ x_0 = y_0 \end{cases}$$

$$t_k = \frac{k+1}{2}$$

Theorem [The Fast Proximal-Gradient Method
for strongly convex functions]

$$\min_x F(x) := f(x) + g(x)$$

Assume $\begin{cases} \textcircled{1} f(x) \text{ is convex} \\ \textcircled{2} g(x) \text{ is strongly convex with } \mu > 0 \\ \textcircled{3} \nabla g(x) \text{ is } L\text{-cont.} \end{cases}$

Fast Proximal Gradient for strongly convex functions

$$\begin{cases} x_0 = y_0 \\ x_{k+1} = \text{Prox}_f^\eta (y_k - \eta \nabla g(y_k)) , \quad \eta = \frac{1}{L} \\ y_{k+1} = x_{k+1} + \frac{\sqrt{\sigma} - 1}{\sqrt{\sigma} + 1} (x_{k+1} - x_k) , \quad \sigma = \frac{L}{\mu} \end{cases}$$

Example: If $\mu I \leq \nabla^2 g(x) \leq L I$, then

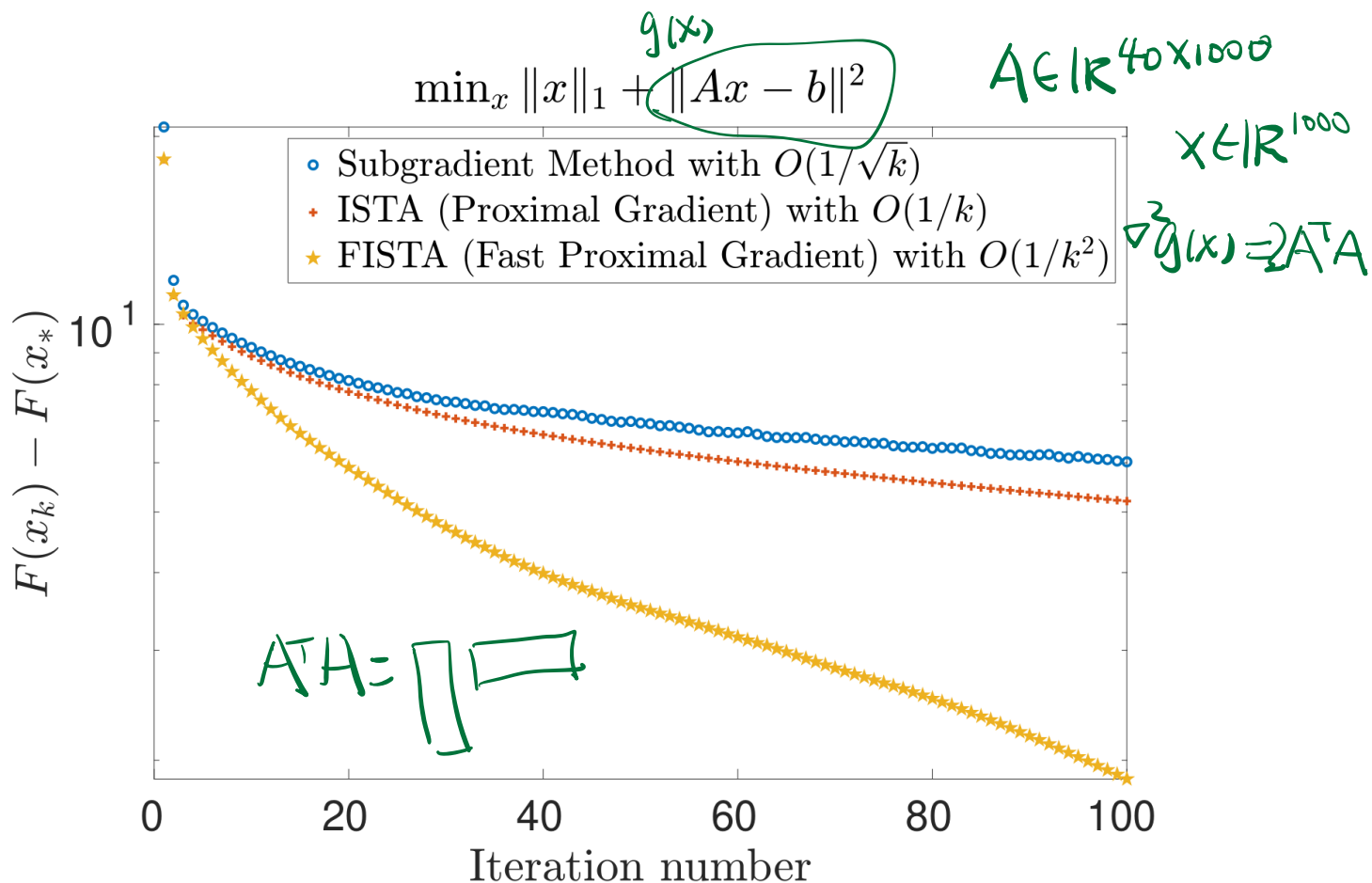
$\sigma = \frac{L}{\mu}$ is the condition number of $\nabla^2 g(x)$.

$$F(x_k) - F(x_*) \leq (1 - \sqrt{\frac{\mu}{L}})^k [F(x_0) - F(x_*) + \frac{\mu}{2} \|x_0 - x_*\|^2]$$

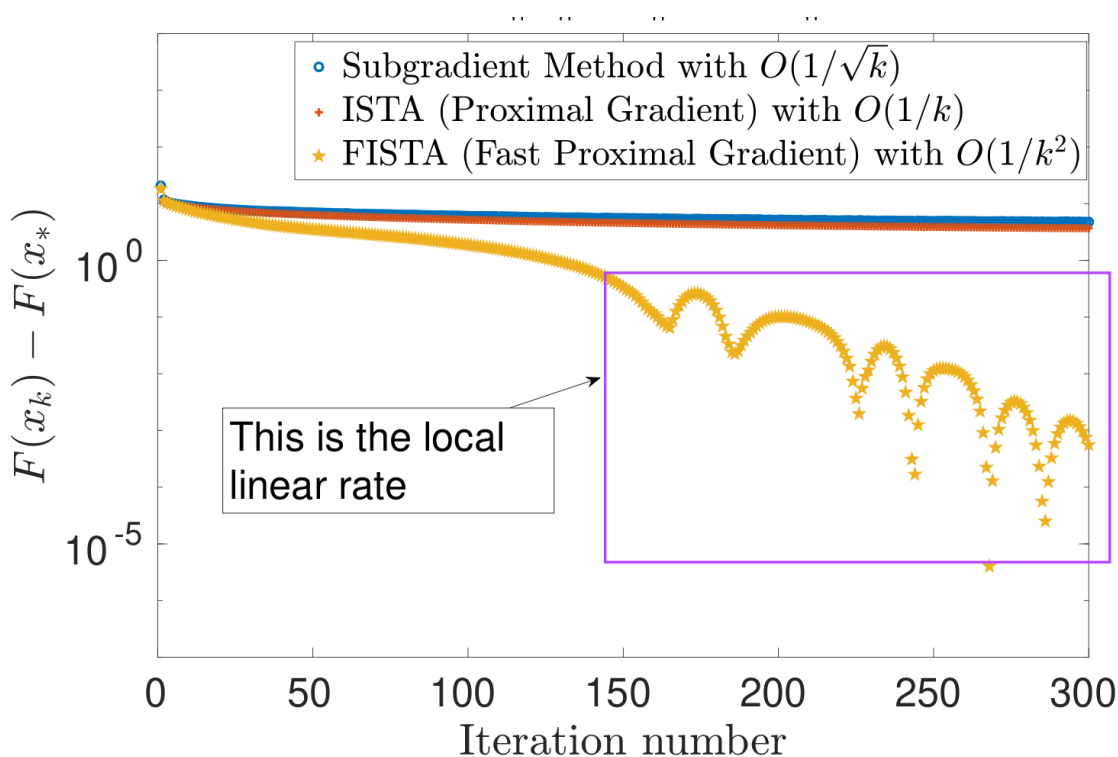
$F(x_k) - F(x_*)$	Convexity	Strong Convexity
Gradient Descent	$O(\frac{1}{k})$	$O\left(\left[\frac{L-\mu}{L+\mu}\right]^2\right)^k \quad \eta = \frac{2}{L+\mu}$
Accelerated GD	$O(\frac{1}{k^2})$	$O((1 - \sqrt{\frac{\mu}{L}})^k) \quad \eta = \frac{1}{L}$
① Subgradient Method	$O(\frac{1}{\sqrt{k}})$	$O(\frac{1}{k})$
② Proximal Point Method	$O(\frac{1}{k})$	$O\left(\left[\left(\frac{1}{1+\eta\mu}\right)^2\right]^k\right) \quad \forall \eta > 0$
③ Proximal Gradient	$O(\frac{1}{k})$	$O((1 - \frac{\mu}{L})^k) \quad \eta = \frac{1}{L}$
④ Accelerated Prox Grad	$O(\frac{1}{k^2})$	$O((1 - \sqrt{\frac{\mu}{L}})^k) \quad \eta = \frac{1}{L}$

$$1 - \sqrt{\frac{\mu}{L}} < \left(\frac{L-\mu}{L+\mu}\right)^2 \text{ if } \frac{\mu}{L} \leq 0.085$$

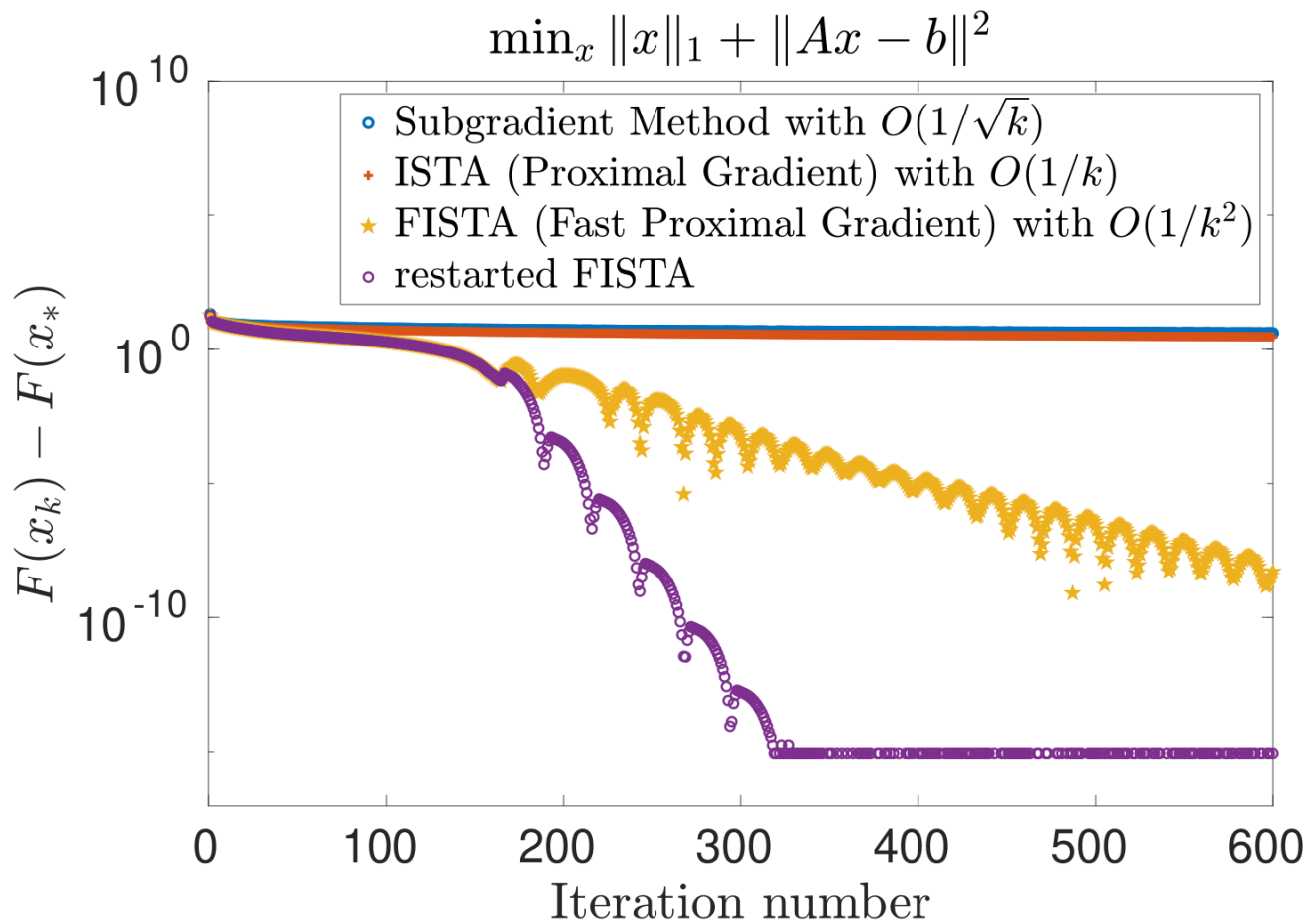
$$x_{k+1} = (I + \eta \partial f)^{-1}(x_k) = \arg\min_u F(u) + \frac{1}{2\eta} \|u - x_k\|^2$$



(a) Provable rates are the worst case rates, which are usually observed in the beginning.



(b) Even if the function has no strong convexity nor smoothness, a local linear rate may be observed: for large k , the iterates $\mathbf{x}_k$ stay on a lower dimensional set, and the function becomes smooth on this set. Such a set is often called active set.

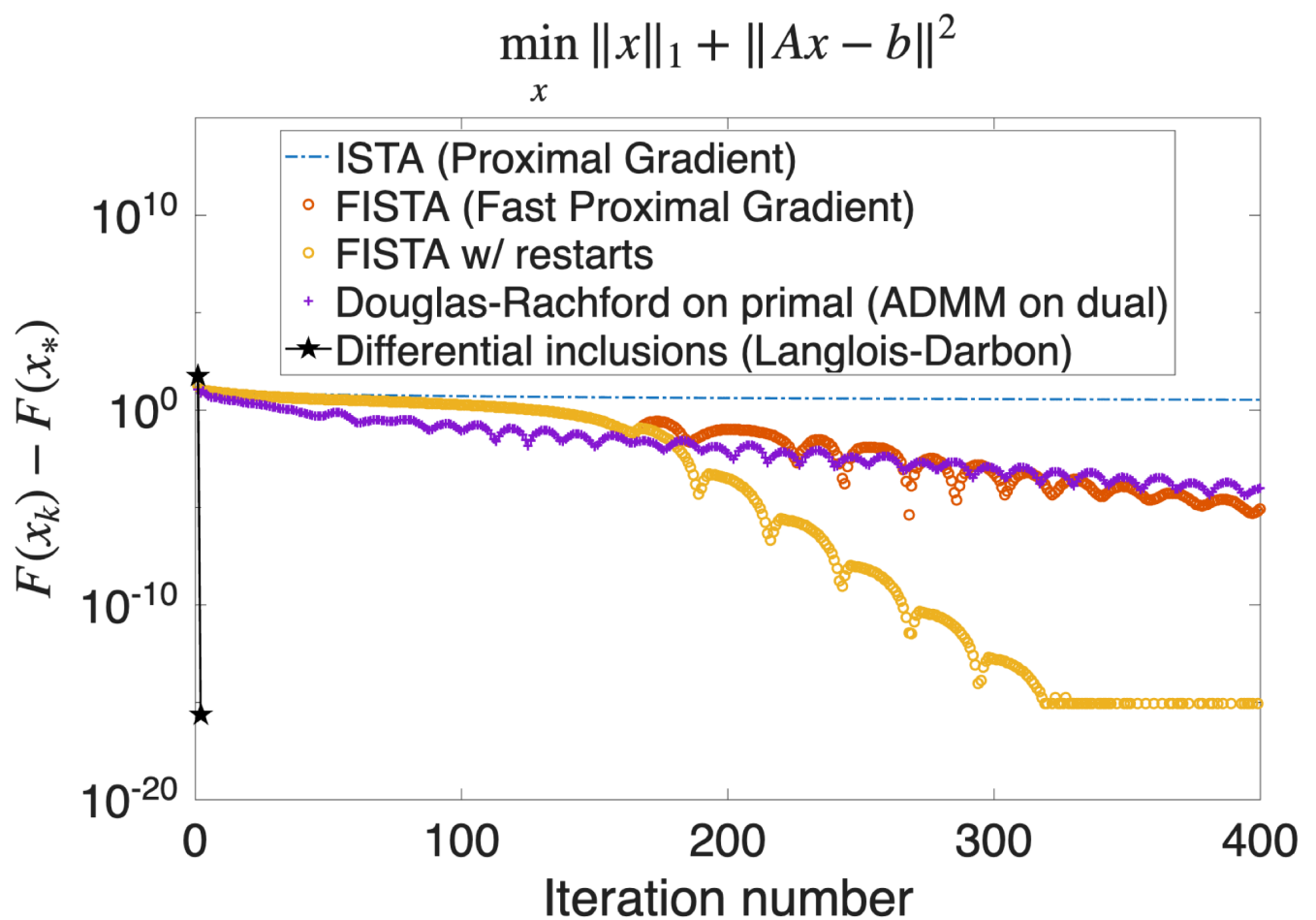


(c) For ℓ^1 problem, restarted FISTA can perform extremely well.

$$A \in \mathbb{R}^{40 \times 1000}$$

$$\nabla^2 g(x) = 2A^T A$$

no strong convexity



① The black method works only for ℓ_1 minimization
(a good reading proj.)

② All other methods can be used for
 $\min_x f(x) + g(x)$ where $\begin{cases} f(x) \text{ is convex} \\ g(x) \text{ is convex \& smooth} \end{cases}$