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

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

Assume
① $f(x)$ is convex
② $g(x)$ is convex
③ $\nabla g(x)$ is L-cont.

Fast / Accelerated Proximal Gradient Method

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

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

Proof: [ Prox-Grad Inequality ]

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

$$F(x) - F(\bar{y}) \geq \frac{L}{2}\|x - \bar{y}\|^2 - \frac{L}{2}\|x - y\|^2$$
$$+ g(x) - g(y) - \langle \nabla g(y), x - y \rangle$$

$\Downarrow$

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

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

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

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

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

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

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

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

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

$$R_k = F_k - F_*$$

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

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

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

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

$$\Rightarrow \quad \frac{L}{2}\|u_k\|^2 + t_{k-1}^2 R_k \leq \frac{L}{2}\|u_1\|^2 + t_0^2 R_1$$

$$= \frac{L}{2}\|x_1 - x_*\|^2 + (F_1 - F_*)$$

$$\boxed{\begin{array}{c} u_k = t_{k-1} x_k - [x_* + (t_{k-1}-1) x_{k-1}] \\ t_0 = 1 \end{array}} \Rightarrow u_1 = x_1 - x_*$$

$$\Rightarrow \quad t_{k-1}^2 R_k \leq \frac{L}{2}\|x_1 - x_*\|^2 + (F_1 - F_*)$$

Prox-Grad Inequality

$$F(\underset{x_*}{x}) - F(\underset{x_1}{\bar{y}}) \geq \frac{L}{2}\|\underset{x_* - x_1}{x - \bar{y}}\|^2 - \frac{L}{2}\|\underset{x_* - y_0}{x - y}\|^2$$

$x_0 = y_0$

$$\Rightarrow \quad F_* - F_1 \geq \frac{L}{2}\|x_1 - x_*\|^2 - \frac{L}{2}\|x_0 - x_*\|^2$$

$$\Rightarrow \quad F_1 - F_* \leq -\frac{L}{2}\|x_1 - x_*\|^2 + \frac{L}{2}\|x_0 - x_*\|^2$$

$$\Rightarrow \quad t_{k-1}^2 R_k \leq \frac{L}{2}\|x_1 - x_*\|^2 + (F_1 - F_*) \leq \frac{L}{2}\|x_0 - x_*\|^2$$

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

$$\Rightarrow \quad R_k \leq 2L\|x_0 - x_*\|^2 \cdot \frac{1}{(k+1)^2}$$

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

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

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

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

Projected Gradient Method (Proximal Gradient)

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

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

Fast Projected Gradient Method

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

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

---

$\boxed{\text{Theorem}}$ [The Fast Proximal-Gradient Method for strongly convex functions]

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

Assume
1. $f(x)$ is convex
2. $g(x)$ is strongly convex with $\mu > 0$
3. $\nabla g(x)$ is $L$-cont.

## Fast Proximal Gradient for Strongly convex functions

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

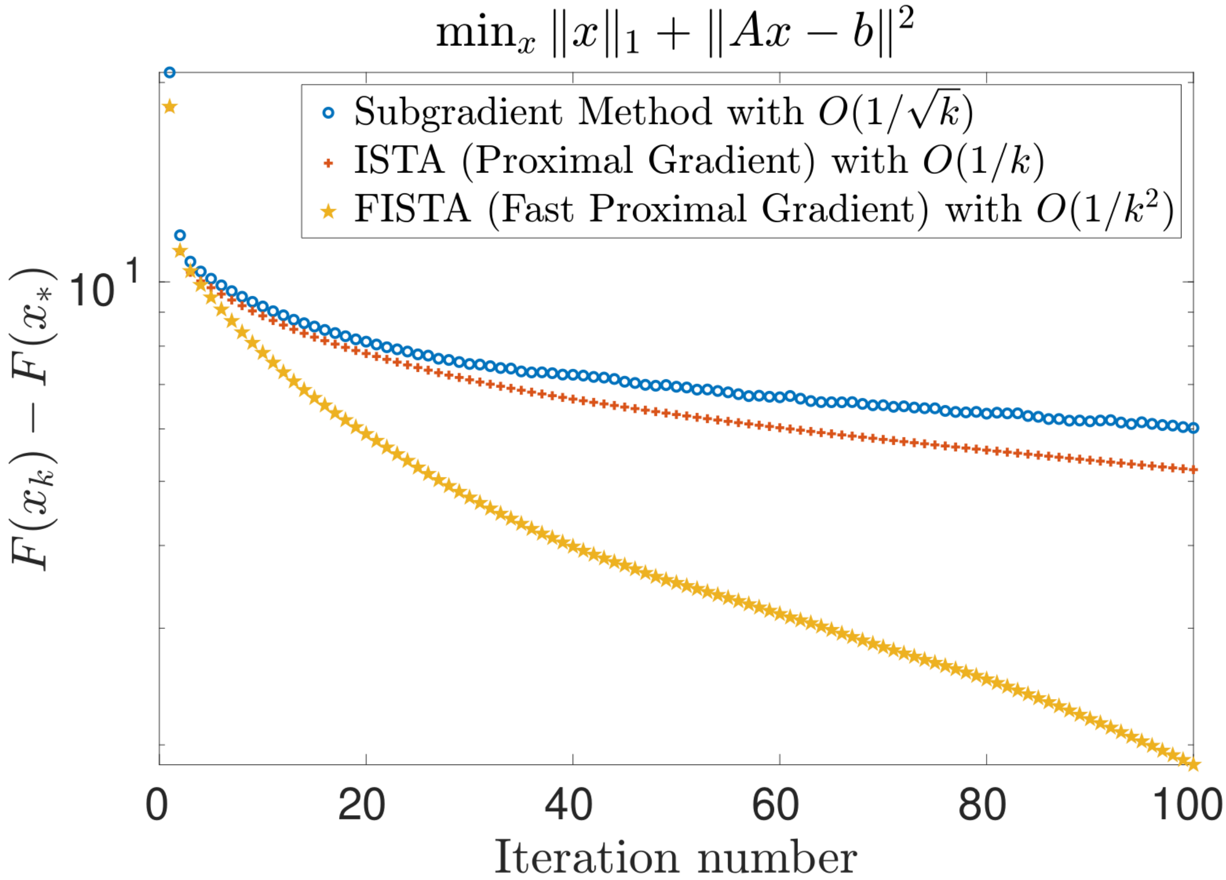
Example: If $\mu I \preceq \nabla^2 g(x) \preceq L I$, then

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

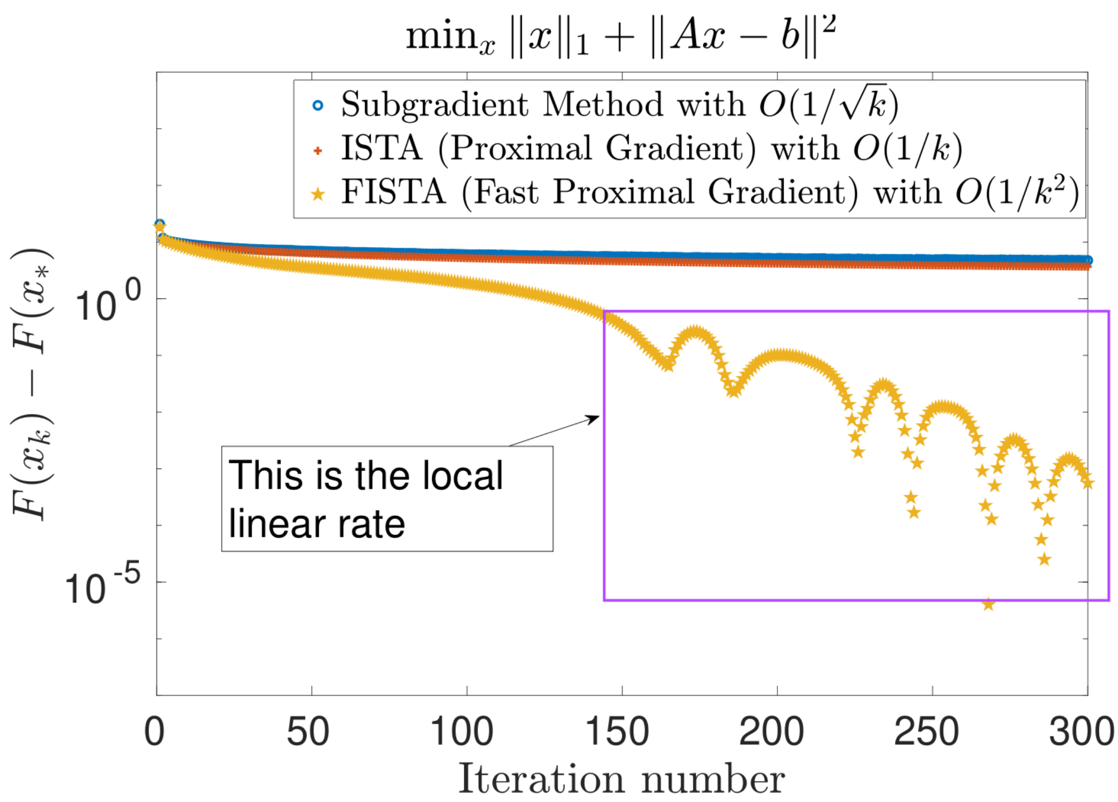
$$F(x_k) - F(x_*) \leq \left(1 - \sqrt{\tfrac{\mu}{L}}\right)^k \left[F(x_0) - F(x_*) + \frac{\mu}{2}\|x_0 - x_*\|^2\right]$$

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

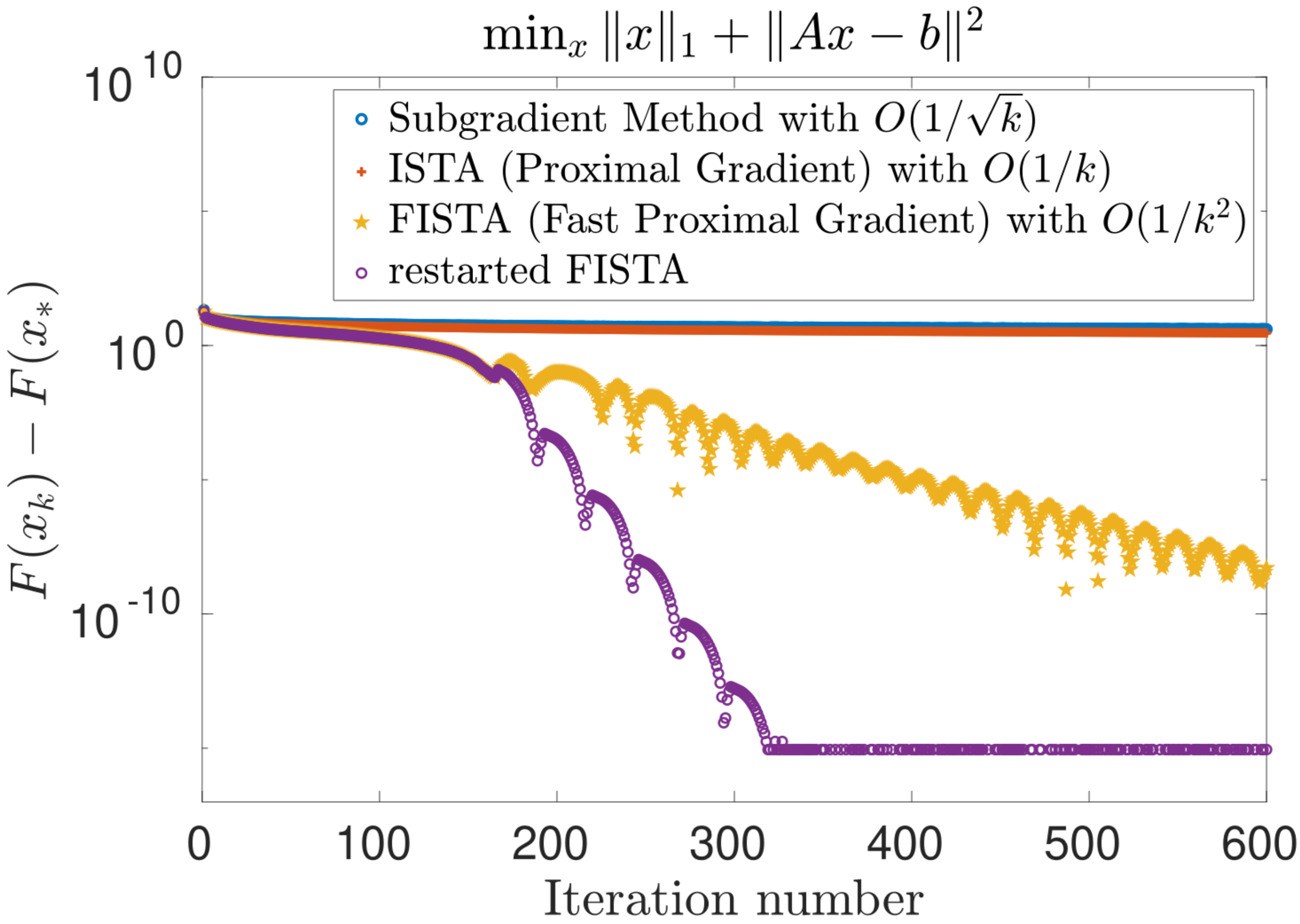
$1 - \sqrt{\frac{\mu}{L}} < \left(\frac{L-\mu}{L+\mu}\right)^2$ if $\frac{\mu}{L} \leq 0.085$




(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 $\ell^1$ problem, restarted FISTA can perform extremely well.

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

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

no strong convexity