Plan

- Motivation of Nesterov's acceleration
- Convergence Proof of Nesterov's Accelerated GD

Gradient Descent (GD): $x_{k+1} = x_k - \eta \nabla f(x_k)$

$$\begin{cases} \text{Assume } \nabla f \text{ is } L\text{-continuous } \& \ f(x) \text{ is convex} \\ \eta \in (0, \frac{2}{L}) \quad f(x_k) - f(x_*) < \frac{\|x_0 - x_*\|^2}{\omega} \frac{1}{k} = O(\frac{1}{k}) \end{cases}$$

Nesterov 1983: $\begin{cases} x_{k+1} = y_k - \eta \nabla f(y_k) \\ y_{k+1} = x_{k+1} + \frac{t_k - 1}{t_{k+1}}(x_{k+1} - x_k) \end{cases}$ $O(\frac{1}{k^2})$

Beck & Teboulle 2009: $\min_{x \in \mathbb{R}^n} f(x) + \|x\|_1$ $\|Ax-b\|_2 + \lambda\|x\|_1$

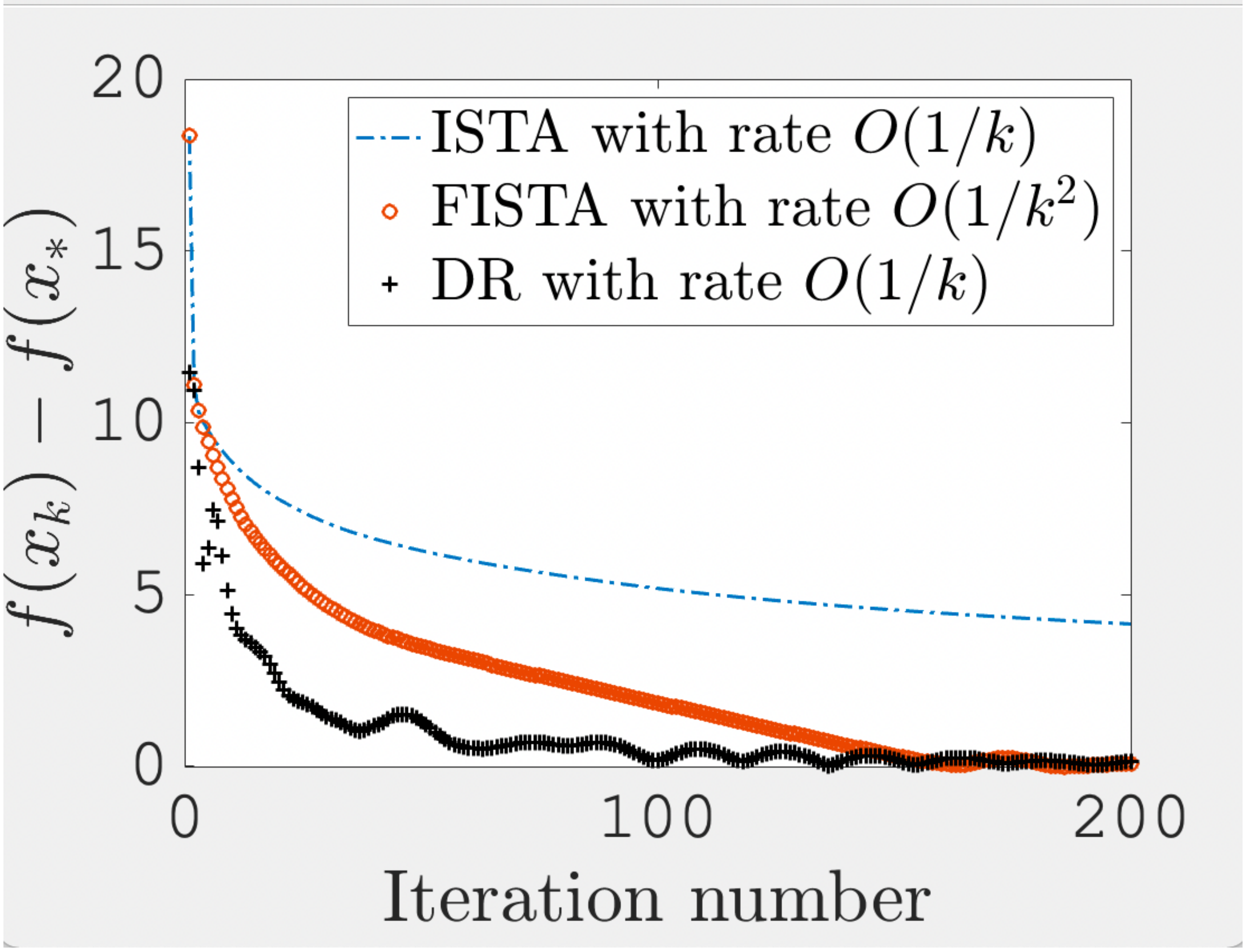
$O(\frac{1}{k^2})$

FISTA $\begin{cases} x_{k+1} = T_{\frac{1}{L}}[y_k - \frac{1}{L}\nabla f(y_k)] \\ y_{k+1} = x_{k+1} + \frac{t_k - 1}{t_{k+1}}(x_{k+1} - x_k) \end{cases}$

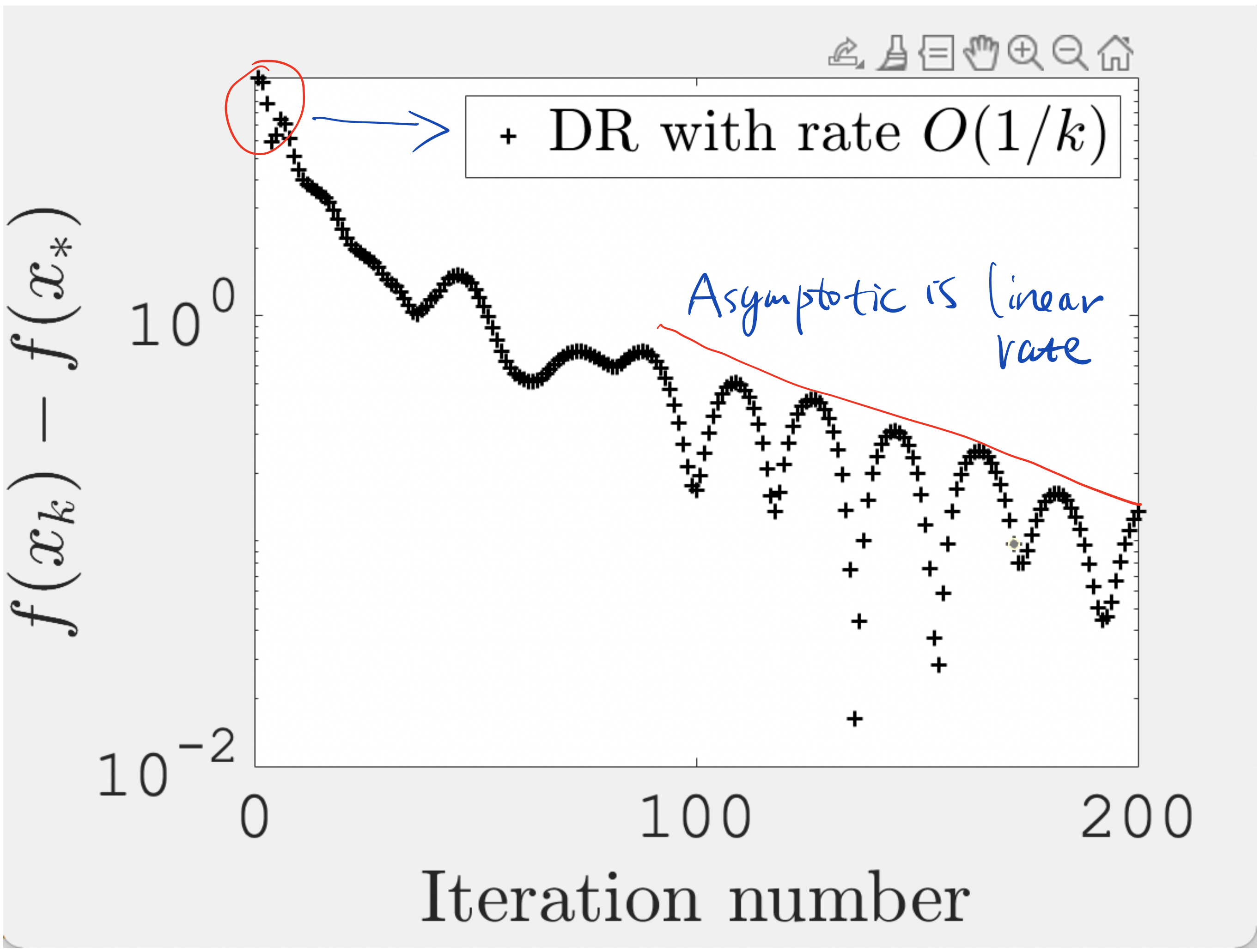
$$T_{\frac{1}{L}}(x)_i = \begin{cases} x_i, & \text{if } -\frac{1}{L} \le x_i \le \frac{1}{L} \\ x_i - \frac{1}{L}, & \text{if } x_i > \frac{1}{L} \\ x_i + \frac{1}{L}, & \text{if } x_i < -\frac{1}{L} \end{cases}$$

ISTA: $x_{k+1} = T_{\frac{1}{L}}[x_k - \frac{1}{L}\nabla f(x_k)]$ $O(\frac{1}{k})$

$$\min_x \|Ax - b\|^2 + \lambda \|x\|_1$$



20
15
10
5
0
0
100
200
$f(x_k) - f(x_*)$
Iteration number
ISTA with rate $O(1/k)$
FISTA with rate $O(1/k^2)$
DR with rate $O(1/k)$



10^0
10^{-2}
0
100
200
$f(x_k) - f(x_*)$
Iteration number
DR with rate $O(1/k)$
Asymptotic is linear rate

Remark: ① Douglas-Rachford splitting solves a different form $\min \|x\|_1$ & $Ax=b$

② In each iteration of Douglas-Rachford, one needs to compute $(A^TA)^{-1}$.

Nesterov's Method

$$\boxed{\begin{cases} \mathbf{x}_{k+1} &= \mathbf{y}_k - \eta_k \nabla f(\mathbf{y}_k) \\ t_{k+1} &= \frac{1}{2}\left(1+\sqrt{4t_k^2+1}\right) \\ \mathbf{y}_{k+1} &= \mathbf{x}_{k+1} + \frac{t_k-1}{t_{k+1}}(\mathbf{x}_{k+1}-\mathbf{x}_k) \end{cases} \qquad \mathbf{x}_0 = \mathbf{y}_0, t_0 = 1.}$$

$\eta = \frac{1}{L}$

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

$$\boxed{\begin{cases} \mathbf{x}_{k+1} &= \mathbf{y}_k - \frac{1}{L} \nabla f(\mathbf{y}_k) \\ \mathbf{y}_{k+1} &= \mathbf{x}_{k+1} + \frac{k-1}{k+2}(\mathbf{x}_{k+1}-\mathbf{x}_k) \end{cases} \qquad \mathbf{x}_0 = \mathbf{y}_0.}$$

Theorem 2.19. *Assume the function $f(\mathbf{x}) : \mathbb{R}^n \longrightarrow \mathbb{R}$ is convex with a global minimizer $\mathbf{x}_*$. Assume $\nabla f(\mathbf{x})$ is Lipschitz continuous with constant L. Assume $t_k^2 - t_k = t_{k-1}^2$ with $t_0 = 1$. Then the following accelerated gradient method*

$$\boxed{\begin{cases} \mathbf{x}_{k+1} &= \mathbf{y}_k - \frac{1}{L} \nabla f(\mathbf{y}_k) \\ \mathbf{y}_{k+1} &= \mathbf{x}_{k+1} + \frac{t_k-1}{t_{k+1}}(\mathbf{x}_{k+1}-\mathbf{x}_k) \end{cases} \qquad \mathbf{x}_0 = \mathbf{y}_0,}$$

satisfies

$$f(\mathbf{x}_k) - f(\mathbf{x}_*) \le \frac{4}{k^2}\left(f(\mathbf{x}_1) - f(\mathbf{x}_*) + \frac{L}{2}\|\mathbf{x}_1 - \mathbf{x}_*\|^2\right).$$

Proof: Consider $\min_x f(x)$

$\eta = \frac{1}{L}$

Start with $\begin{cases} x_{k+1} = y_k - \eta \nabla f(y_k) \\ y_{k+1} = x_{k+1} + \frac{t_k - 1}{t_{k+1}}(x_{k+1} - x_k) \end{cases}$

$x_0 = y_0$

$t_k \in \mathbb{R}$

$\boxed{\text{Descent Lemma} \quad \text{Assume } \nabla f(x) \text{ is } L\text{-Lipschitz} \quad f(y) \le f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|x-y\|^2}$

1) $f(x_{k+1}) \le f(y_k) + \langle \nabla f(y_k), x_{k+1} - y_k\rangle + \frac{L}{2}\|x_{k+1} - y_k\|^2$

$\boxed{\text{Convexity} \Rightarrow f(x) \ge f(y) + \langle \nabla f(y), x - y\rangle}$

2) $f(x_k) \geq f(y_k) + \langle \nabla f(y_k), x_k - y_k\rangle$

$\left.\begin{matrix} 1) \\ 2) \end{matrix}\right\} \Rightarrow f(x_{k+1}) - f(x_k) \le \langle \nabla f(y_k), x_{k+1} - x_k\rangle + \frac{L}{2}\|x_{k+1} - y_k\|^2$

$$\Rightarrow f(x_k) - f(x_{k+1}) \geq -\frac{L}{2}\|y_k - x_{k+1}\|^2 + \langle \nabla f(y_k), x_k - x_{k+1}\rangle$$

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

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

$$= \left(\frac{1}{\eta} - \frac{L}{2}\right)\|y_k - x_{k+1}\|^2 + \frac{1}{\eta}\langle y_k - x_{k+1},\ x_k - y_k\rangle$$

$$\eta = \frac{1}{L} \iff \frac{1}{\eta} - \frac{L}{2} = \frac{L}{2},\quad \frac{1}{\eta} = L$$

$$① \quad f(x_k) - f(x_{k+1}) \geq \frac{L}{2}\|y_k - x_{k+1}\|^2 + L\langle y_k - x_{k+1},\ x_k - y_k\rangle$$

$$\begin{cases} f(x_{k+1}) \leq f(y_k) + \langle \nabla f(y_k), x_{k+1} - y_k\rangle + \frac{L}{2}\|y_k - x_{k+1}\|^2 \\ f(x_*) \geq f(y_k) + \langle \nabla f(y_k), x_* - y_k\rangle \end{cases}$$

$$f(x_*) - f(x_{k+1}) \geq -\frac{L}{2}\|y_k - x_{k+1}\|^2 + \langle \nabla f(y_k),\ x_* - x_{k+1}\rangle$$

$$= -\frac{L}{2}\|y_k - x_{k+1}\|^2 + \left\langle \frac{y_k - x_{k+1}}{\eta},\ x_* - y_k + y_k - x_{k+1}\right\rangle$$

$$\eta = \frac{1}{L}$$

$\Rightarrow$

$$② \quad f(x_*) - f(x_{k+1}) \geq \frac{L}{2}\|y_k - x_{k+1}\|^2 + L\langle y_k - x_{k+1},\ x_* - y_k\rangle$$

$$① \quad f(x_k) - f(x_{k+1}) \geq \frac{L}{2}\|y_k - x_{k+1}\|^2 + L\langle y_k - x_{k+1},\ x_k - y_k\rangle$$

$$R_k = f(x_k) - f(x_*)$$

$$R_k - R_{k+1} = f(x_k) - f(x_{k+1})$$

$$(t_k - 1)\cdot ① + ② \Rightarrow$$

$$(t_k - 1)[R_k - R_{k+1}] - R_{k+1} \geq \frac{L}{2} t_k \|y_k - x_{k+1}\|^2$$
$$+ L\langle y_k - x_{k+1},\ (t_k - 1)x_k - t_k y_k - x_*\rangle$$

$$\underbrace{t_k(t_k-1)}_{\leq t_{k-1}^2}R_k - t_k^2 R_{k+1} \geq \frac{L}{2}\|t_k(y_k - x_{k+1})\|^2$$

$$+ L\langle t_k(y_k - x_{k+1}), (t_k-1)x_k - t_k y_k - x_* \rangle$$

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

$$\Rightarrow \quad t_{k-1}^2 R_k - t_k^2 R_{k+1} \geq \frac{L}{2}\|t_k(y_k - x_{k+1})\|^2$$

$$+ L\langle t_k(y_k - x_{k+1}), \underbrace{(t_k-1)x_k - x_*}_{c} - \underbrace{t_k y_k}_{a} \rangle$$

$$b = t_k x_{k+1}$$

$$RHS = \frac{L}{2}\|a-b\|^2 + L\langle a-b, c-a \rangle$$

$$= \frac{L}{2}\left(\|a\|^2 + \|b\|^2 - \cancel{2a\cdot b} - 2\|a\|^2 + \cancel{2a\cdot b} + 2a\cdot c - 2b\cdot c\right)$$

$$= \frac{L}{2}\left(\|b\|^2 - 2b\cdot c + \|c\|^2 - \|c\|^2 + 2a\cdot c - \|a\|^2\right)$$

$$= \frac{L}{2}\left(\|b-c\|^2 - \|a-c\|^2\right)$$

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

$$u_{k+1} = b - c = t_k x_{k+1} - [(t_k-1)x_k + x_*]$$

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

$$\Rightarrow \quad t_k^2 R_{k+1} + \frac{L}{2}\|u_{k+1}\|^2 \leq t_0^2 R_1 + \frac{L}{2}\|u_1\|^2$$

$$\Rightarrow \quad t_k^2 R_{k+1} \le t_0^2 R_1 + \frac{L}{2}\|u_1\|^2$$

$$\Rightarrow \quad R_{k+1} \le \frac{1}{t_k^2}\left[t_0^2 R_1 + \frac{L}{2}\|u_1\|^2\right]$$

$$\left(\left.\begin{matrix} t_k^2 - t_k \le t_{k-1} \\ t_0 = 1 \end{matrix}\right\} \Rightarrow t_k \ge \frac{k+1}{2}\right)$$

$$\Rightarrow \quad R_{k+1} \le \underbrace{\frac{4}{(k+1)^2}}_{O\left(\frac{1}{k^2}\right)}\left[t_0^2 R_1 + \frac{L}{2}\|u_1\|^2\right]$$