

Plan

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

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

$$\begin{cases} \text{Assume } \nabla f \text{ is } L\text{-continuous \& } f(x) \text{ is convex} \\ \eta \in (0, \frac{2}{L}) \end{cases} \quad f(X_k) - f(X_*) < \frac{\|X_0 - X_*\|^2}{\omega} \frac{1}{k} = O\left(\frac{1}{k}\right)$$

$$\text{Nesterov 1983: } \begin{cases} X_{k+1} = Y_k - \eta \nabla f(Y_k) \\ Y_{k+1} = X_{k+1} + \frac{t_k - 1}{t_{k+1}} (X_{k+1} - X_k) \end{cases} \quad O\left(\frac{1}{k^2}\right)$$

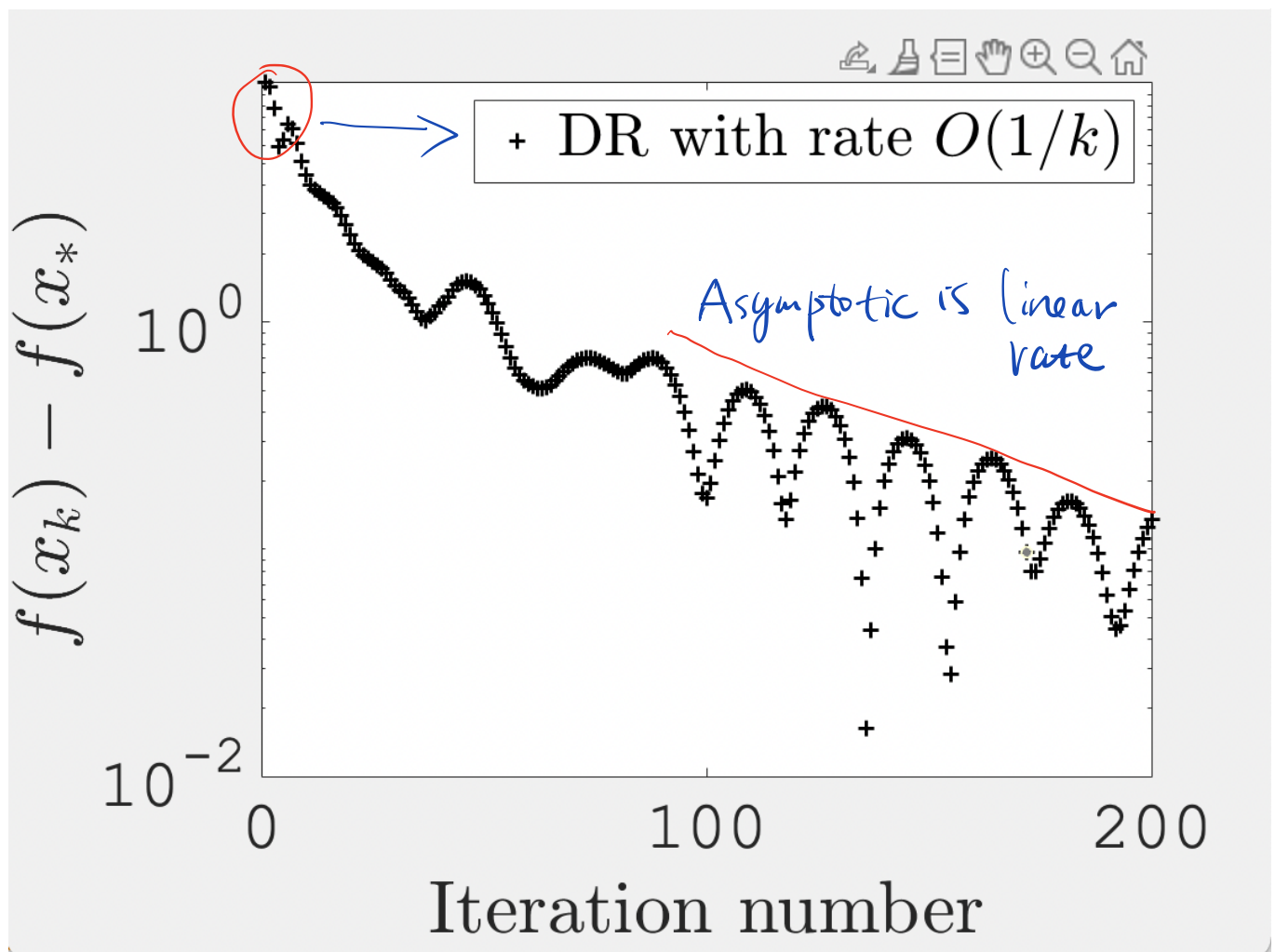
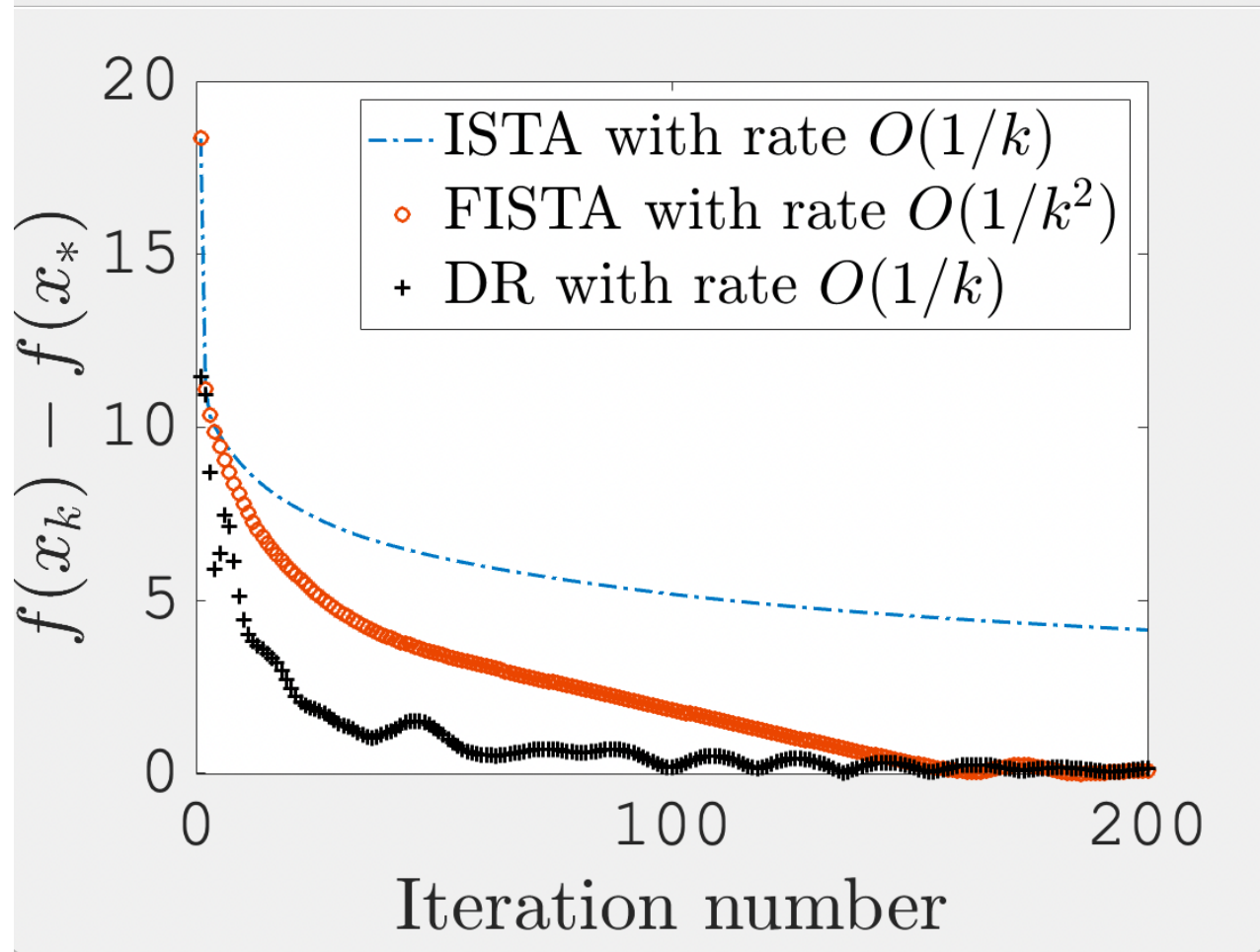
$$\text{Beck \& Teboulle 2009: } \min_{X \in \mathbb{R}^n} \underline{f(x)} + \|x\|_1 \quad \|Ax - b\|_2 + \lambda \|x\|_1$$

$$\text{FISTA } \begin{cases} X_{k+1} = T_{\frac{1}{L}} \left[Y_k - \frac{1}{L} \nabla f(Y_k) \right] \\ Y_{k+1} = X_{k+1} + \frac{t_k - 1}{t_{k+1}} (X_{k+1} - X_k) \end{cases} \quad O\left(\frac{1}{k^2}\right)$$

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

$$\text{ISTA: } X_{k+1} = T_{\frac{1}{L}} \left[X_k - \frac{1}{L} \nabla f(X_k) \right] \quad O\left(\frac{1}{k}\right)$$

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



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 $(A^T A)^{-1}$.

Nesterov's
Method

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

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

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

$$\begin{cases} \mathbf{x}_{k+1} &= \mathbf{y}_k - \frac{1}{L} \nabla f(\mathbf{y}_k) \\ \mathbf{y}_{k+1} &= \mathbf{x}_{k+1} + \frac{k}{k+3} (\mathbf{x}_{k+1} - \mathbf{x}_k) \end{cases} \quad \mathbf{x}_0 = \mathbf{y}_0.$$

Theorem 2.28. Assume the function $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ 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

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

satisfies

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

Remark: $\mathbf{x}_k \rightarrow \mathbf{x}_*$ was proven
only recently
A good reading choice 

[Submitted on 27 Oct 2025 (v1), last revised 19 Jan 2026 (this version, v2)]

**Point Convergence of Nesterov's
Accelerated Gradient Method: An AI-
Assisted Proof**

Uijeong Jang, Ernest K. Ryu

Proof: Consider $\min_x f(x)$

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

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

$$t_k \in \mathbb{R}$$

Descent Lemma Assume $\nabla f(x)$ is L -Lipschitz

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|x - y\|^2$$

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

Convexity $\Rightarrow f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle$

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

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

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

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

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

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

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

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

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

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

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

$\Rightarrow$

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

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

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

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

$$(t_k - 1) \cdot \textcircled{1} + \textcircled{2} \Rightarrow$$

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

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

$$\begin{aligned}
 \underline{t_k(t_{k-1})R_k - t_k^2 R_{k+1}} &\geq \frac{L}{2} \|t_k(y_k - x_{k+1})\|^2 \\
 &\leq t_{k-1}^2 + L \langle t_k(y_k - x_{k+1}), (t_{k-1})x_k - t_k y_k - x_* \rangle \\
 &\quad \underline{t_k^2 - t_k \leq t_{k-1}^2}
 \end{aligned}$$

$$\begin{aligned}
 \Rightarrow t_{k-1}^2 R_k - t_k^2 R_{k+1} &\geq \frac{L}{2} \|t_k(y_k - x_{k+1})\|^2 \\
 &\quad + L \langle t_k(y_k - x_{k+1}), \underbrace{(t_{k-1})x_k - x_*}_c - \underbrace{t_k y_k}_a \rangle \\
 &\quad b = t_k x_{k+1}
 \end{aligned}$$

$$\begin{aligned}
 \text{RHS} &= \frac{L}{2} \|a - b\|^2 + L \langle a - b, c - a \rangle \\
 &= \frac{L}{2} (\|a\|^2 + \|b\|^2 - 2a \cdot b - 2\|a\|^2 + 2a \cdot b + 2a \cdot c - 2b \cdot c) \\
 &= \frac{L}{2} (\|b\|^2 - 2b \cdot c + \|c\|^2 - \|c\|^2 + 2a \cdot c - \|a\|^2) \\
 &= \frac{L}{2} (\|b - c\|^2 - \|a - c\|^2) \quad a - c = t_k y_k - (t_{k-1})x_k + x_*
 \end{aligned}$$

$$\begin{aligned}
 \Rightarrow t_{k-1}^2 R_k - t_k^2 R_{k+1} &\geq \frac{L}{2} (\|u_{k+1}\|^2 - \|u_k\|^2) \\
 u_{k+1} &= b - c = t_k x_{k+1} - [(t_{k-1})x_k - x_*]
 \end{aligned}$$

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

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

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

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

$$\left(\begin{array}{l} t_k^2 - t_k \leq t_{k-1} \\ t_0 = 1 \end{array} \right) \Rightarrow t_k \geq \frac{k+1}{2}$$

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

$$\downarrow$$

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