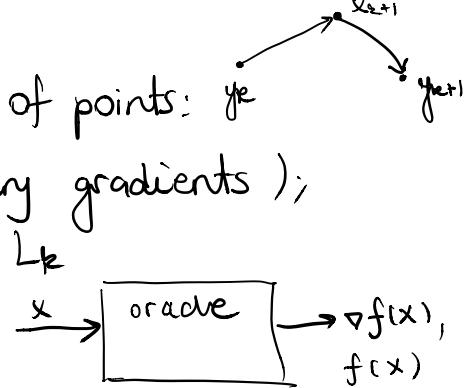


* Nesterov acceleration for smooth & strongly convex minimization (direct method)

- * We will again have 3 sequences of points: y_k
- x_k : query points (where we query gradients);
- v_k : pts we get from the l.b. L_k
- y_k : output pts ($y_k = x_k^{\text{out}}$)



* Upper bound:

$$u_k = f(y_k)$$

$$\begin{aligned} A_k u_k - A_{k-1} u_{k-1} &= A_k f(y_k) - A_{k-1} f(y_{k-1}) \\ &= A_k (f(y_k) - f(x_k)) + A_{k-1} (f(x_k) - f(y_{k-1})) \\ &\quad + A_k f(x_k) \end{aligned}$$

M_k
 G_k
 $x_k^{\text{out}}, f(x_k^{\text{out}})$
 $x^*, f(x^*)$
 L_k
 fast-growing $A_k > 0$
 s.t. $A_k G_k \leq A_{k-1} G_{k-1}$, H_k
 $A_k = \sum_{i=0}^k a_i$, $a_i > 0$.

* Lower bound:

By m-strong convexity, $\forall i$:

$$f(x^*) \geq f(x_i) + \langle \nabla f(x_i), x^* - x_i \rangle + \frac{m}{2} \|x^* - x_i\|_2^2.$$

$$\begin{aligned} f(x^*) &\geq \frac{1}{A_k} \sum_{i=0}^k a_i f(x_i) + \underbrace{\frac{1}{A_k} \sum_{i=0}^k a_i \left(\langle \nabla f(x_i), x^* - x_i \rangle + \frac{m}{2} \|x^* - x_i\|_2^2 \right)}_{+ \frac{m_0}{2A_k} \|x^* - x_0\|_2^2 - \frac{m_0}{2A_k} \|x^* - x_0\|_2^2} \\ &\quad \boxed{m_0 = L - m} \end{aligned}$$

$$\begin{aligned} &\geq \frac{1}{A_k} \sum_{i=0}^k a_i f(x_i) - \frac{m_0}{2A_k} \|x^* - x_0\|_2^2 \\ &\quad + \frac{1}{A_k} \min_{u \in \mathbb{R}^d} \left\{ \sum_{i=0}^k a_i \left(\langle \nabla f(x_i), u - x_i \rangle + \frac{m}{2} \|u - x_i\|_2^2 \right) \right. \\ &\quad \left. + \frac{m_0}{2} \|u - x_0\|_2^2 \right\} M_k(u). \end{aligned}$$

$$= \frac{1}{A_k} \sum_{i=0}^k a_i f(x_i) - \frac{m_0}{2A_k} \|x^* - x_0\|_2^2 + \frac{1}{A_k} M_k(v_k) =: L_k$$

where $v_k = \underset{u \in \mathbb{R}^d}{\operatorname{argmin}} M_k(u)$.

$$\sum_{i=0}^k \alpha_i \nabla f(x_i) + \sum_{i=0}^k \alpha_i m (v_k - x_i) + m_0 (v_k - x_0) = 0$$

$$(m_0 + m A_k) v_k = m_0 x_0 + m \sum_{i=0}^k \alpha_i x_i - \sum_{i=0}^k \alpha_i \nabla f(x_i)$$

$$A_k L_k - A_{k-1} L_{k-1} = \alpha_k \underline{\nabla f(x_k)} + \underline{M_k(v_k) - M_{k-1}(v_{k-1})}$$

HW #, Q4 (iv) \rightarrow

$$M_k(v_k) - M_{k-1}(v_{k-1}) = \alpha_k \langle \nabla f(x_k), v_k - x_k \rangle + \frac{\alpha_k m}{2} \|v_k - x_k\|_2^2$$

$$+ \frac{m_0 + m A_{k-1}}{2} \|v_k - v_{k-1}\|_2^2$$

$$\geq \alpha_k \langle \nabla f(x_k), v_k - x_k \rangle + \frac{(\alpha_k m)}{m_0 + m A_k} \|v_k - x_k\|_2^2$$

$$+ \frac{m_0 + m A_{k-1}}{m_0 + m A_k} \|v_k - v_{k-1}\|_2^2 \cdot (m_0 + m A_k)$$

Jensen

$$\geq \alpha_k \langle \nabla f(x_k), v_k - x_k \rangle$$

$$+ \frac{m_0 + m A_k}{2} \left\| \frac{m_0 + m A_k}{m_0 + m A_k} v_k - \frac{m_0 + m A_{k-1}}{m_0 + m A_k} v_{k-1} - \frac{\alpha_k m}{m_0 + m A_k} x_k \right\|_2^2$$

$$= \frac{1}{2(m_0 + m A_k)} \left\| \underbrace{(m_0 + m A_k) v_k - (m_0 + m A_{k-1}) v_{k-1} - \alpha_k m x_k}_{- \alpha_k \nabla f(x_k)} \right\|_2^2$$

$$= \alpha_k \langle \nabla f(x_k), v_k - x_k \rangle + \frac{\alpha_k^2}{2(m_0 + m A_k)} \|\nabla f(x_k)\|_2^2.$$

$$\Rightarrow A_k L_k - A_{k-1} L_{k-1} \geq \alpha_k \underline{\nabla f(x_k)} + \alpha_k \langle \nabla f(x_k), v_k - x_k \rangle$$

$$+ \frac{\alpha_k^2}{2(m_0 + m A_k)} \|\nabla f(x_k)\|_2^2.$$

$$(m_0 + m A_k)(v_k - x_k) = (m_0 + m A_{k-1})(v_{k-1} - x_k) - \alpha_k \nabla f(x_k) / \frac{\alpha_k}{m_0 + m A_k}$$

$$\alpha_k (v_k - x_k) = \frac{\alpha_k (m_0 + m A_{k-1})}{m_0 + m A_k} (v_{k-1} - x_k) - \frac{\alpha_k^2}{m_0 + m A_k} \nabla f(x_k)$$

$$A_k L_k - A_{k-1} L_{k-1} \geq \underbrace{\alpha_k f(x_k)}_{\text{purple}} + \alpha_k' \langle \nabla f(x_k), v_{k-1} - x_k \rangle - \frac{\alpha_k^2}{2(M_0 + M A_k)} \|\nabla f(x_k)\|_2^2$$

* The gap:

$$\text{Recall: } A_k u_k - A_{k-1} u_{k-1} = A_k f(y_k) - A_{k-1} f(y_{k-1})$$

$$= A_k (f(y_k) - f(x_k)) + A_{k-1} (f(x_k) - f(y_{k-1})) + \underbrace{\alpha_k f(x_k)}_{\text{purple}}$$

$$G_k = u_k - L_k$$

$$A_k G_k - A_{k-1} G_{k-1} \leq A_k (f(y_k) - f(x_k)) + \frac{\alpha_k^2}{2(M_0 + M A_k)} \|\nabla f(x_k)\|_2^2 + A_{k-1} (f(x_k) - f(y_{k-1})) - \alpha_k' \langle \nabla f(x_k), v_{k-1} - x_k \rangle$$

Γ Set $y_k = x_k - \frac{1}{L} \nabla f(x_k) \Rightarrow f(y_k) - f(x_k) \leq -\frac{1}{2L} \|\nabla f(x_k)\|_2^2$

$$\leq \frac{1}{2} \|\nabla f(x_k)\|_2^2 \left(-\frac{A_k}{L} + \underbrace{\frac{\alpha_k^2}{M_0 + M A_k}}_{\leq 0} \right)$$

$$\frac{\alpha_k}{M_0 + M A_k} \leq \sqrt{\frac{m}{L}}$$

$$+ \langle \nabla f(x_k), \underbrace{A_{k-1}(x_k - y_{k-1}) - \alpha_k'(v_{k-1} - x_k)}_{= 0} \rangle$$

$$= 0$$

$$x_k = \frac{A_{k-1}}{A_{k-1} + \alpha_k'} y_{k-1} + \frac{\alpha_k'}{A_{k-1} + \alpha_k'} v_{k-1}$$

* The resulting algorithm: (Nesterov AGD)

$$\frac{\alpha_k}{\frac{m_0}{m} + A_k} \leq \sqrt{\frac{m}{L}}$$

$$x_k = \frac{A_{k-1}}{A_{k-1} + \alpha'_k} y_{k-1} + \frac{\alpha'_k}{A_{k-1} + \alpha'_k} v_{k-1}, \quad \alpha'_k = \alpha_k \frac{m_0 + m A_{k-1}}{m_0 + m A_k}$$

$$v_k = \underset{u \in \mathbb{R}^d}{\operatorname{argmin}} M_k(u)$$

$$y_k = x_k - \frac{1}{L} \nabla f(x_k)$$

$$\star \underline{Ex.} \quad A_0 G_0 \leq \frac{A_0(L-m)}{2} \|x^* - x_0\|_2^2, \quad m_0 = L-m$$

$$\underline{f(y_k) - f(x^*)} \leq G_k \leq \left(\frac{A_0 G_0}{A_k} \right)$$

$$\text{suffices: } \frac{\alpha_k}{A_k} = \sqrt{\frac{m}{L}} \quad \left(\frac{\alpha_k}{A_{k-1} + \alpha_k} = \sqrt{\frac{m}{L}} \right), \quad a_0 = A_0 = 1.$$

$$(A_k + A_{k-1}) = A_k \quad \frac{A_{k-1}}{A_k} = 1 - \frac{\alpha_k}{A_k} = 1 - \sqrt{\frac{m}{L}}$$

$$\frac{A_0}{A_k} = \underbrace{\frac{A_0}{A_1}}_{(1-\sqrt{\frac{m}{L}})} \cdot \underbrace{\frac{A_1}{A_2}}_{(1-\sqrt{\frac{m}{L}})} \cdots \underbrace{\frac{A_{k-1}}{A_k}}_{(1-\sqrt{\frac{m}{L}})} = \left(1 - \sqrt{\frac{m}{L}} \right)^k$$

$$f(y_k) - f(x^*) \leq \underbrace{\left(1 - \sqrt{\frac{m}{L}} \right)^k}_{\leq e^{-\frac{L-m}{2E} k}} \cdot \frac{L-m}{2} \|x^* - x_0\|_2^2 = \epsilon$$

Thus, $f(y_k) - f(x^*) \leq \epsilon$ after

$$k \geq \sqrt{\frac{L}{m}} \log \left(\frac{L-m}{2\epsilon} \|x^* - x_0\|_2^2 \right) \text{ iterations.}$$