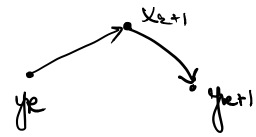
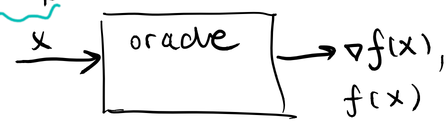
* Nesterov acceleration for smooth & strongly convex minimization (direct method)



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



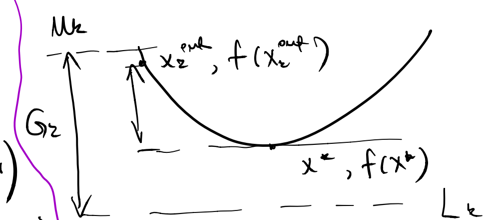
* Upper bound:

$U_k = f(y_k)$

$A_k U_k - A_{k-1} U_{k-1} = A_k f(y_k) - A_{k-1} f(y_{k-1})$

$= A_k(f(y_k) - f(x_k)) + A_{k-1}(f(x_k) - f(y_{k-1}))$

$+ a_k f(x_k)$



fast-growing $A_k > 0$

s.t. $A_k G_k \le A_{k-1} G_{k-1}$, $\forall k$

$A_k = \sum_{i=0}^{k} a_i$, $a_i > 0$.

* Lower bound:

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

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

$$f(x^*) \geq \frac{1}{A_k} \sum_{i=0}^{k} a_i f(x_i) + \frac{1}{A_k} \sum_{i=0}^{k} a_i \left(\langle \nabla f(x_i), x^* - x_i \rangle + \frac{m}{2} \|x^* - x_i\|_2^2 \right)$$

$$+ \frac{m_0}{2A_k} \| x^* - x_0 \|_2^2 - \frac{m_0}{2A_k} \| x^* - x_0 \|_2^2$$

$\boxed{m_0 = L - m}$

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

$$+ \frac{1}{A_k} \min_{u \in \mathbb{R}^d} \left\{ \sum_{i=0}^{k} a_i \left(\langle \nabla f(x_i), u - x_i \rangle + \frac{m}{2} \|u - x_i\|_2^2 \right) + \frac{m_0}{2} \|u - x_0\|_2^2 \right\}$$

$M_k(u)$.

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

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

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

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

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

HW #, Q4 (iv) $\rightarrow$

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

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

$$\geq a_k \langle \nabla f(x_k), v_k - x_k \rangle + \left(\frac{a_k m}{m_0 + m A_k} \| v_k - x_k \|_2^2 + \frac{m_0 + m A_{k-1}}{m_0 + m A_k} \| v_k - v_{k-1} \|_2^2 \right) \cdot (m_0 + m A_k)$$

Jensen

$$\geq a_k \langle \nabla f(x_k), v_k - x_k \rangle + \frac{m_0 + m A_k}{2} \left\| \frac{m_0 + m A_k}{m_0 + m A_k} v_k - \frac{m_0 + m A_{k-1}}{m_0 + m A_k} v_{k-1} - \frac{a_k m}{m_0 + m A_k} x_k \right\|_2^2$$

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

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

$$\Rightarrow A_k L_k - A_{k-1} L_{k-1} \geq a_k f(x_k) + a_k \langle \nabla f(x_k), \underline{v_k} - x_k \rangle + \frac{a_k^2}{2(m_0 + m A_k)} \| \nabla f(x_k) \|_2^2 .$$

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

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

a_k'

$$A_k L_k - A_{k-1} L_{k-1} \geq a_k f(x_k) + a_k' \langle \nabla f(x_k), v_{k-1} - x_k \rangle$$
$$- \frac{a_k^2}{2(m_0 + m A_k)} \|\nabla f(x_k)\|_2^2 .$$

* The gap:

Recall: $A_k U_k - A_{k-1} U_{k-1} = A_k f(y_k) - A_{k-1} f(y_{k-1})$ $\quad \boxed{G_k = U_k - L_k}$

$$= A_k (f(y_k) - f(x_k)) + A_{k-1} (f(x_k) - f(y_{k-1}))$$
$$+ a_k f(x_k)$$

$$A_k G_k - A_{k-1} G_{k-1} \leq A_k (f(y_k) - f(x_k)) + \frac{a_k^2}{2(m_0 + m A_k)} \|\nabla f(x_k)\|_2^2$$
$$+ A_{k-1} (f(x_k) - f(y_{k-1}))$$
$$- a_k' \langle \nabla f(x_k), v_{k-1} - x_k \rangle$$

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

$$\leq \frac{1}{2} \|\nabla f(x_k)\|_2^2 \Big(\underbrace{- \frac{A_k}{L} + \frac{a_k^2}{m_0 + m A_k}}_{\leq 0} \Big)$$

$$\boxed{\frac{a_k}{\frac{m_0}{m} + A_k} \leq \sqrt{\frac{m}{L}}} \; *$$

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

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

* The resulting algorithm: (Nesterov AGD)

$$\frac{a_k}{\frac{m_0}{m} + A_k} \lesssim \sqrt{\frac{m}{L}}$$

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

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

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

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

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

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

$(a_k + A_{k-1} = A_k)$

$$\frac{A_{k-1}}{A_k} = 1 - \frac{a_k}{A_k} = 1 - \sqrt{\frac{m}{L}}$$

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

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

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

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