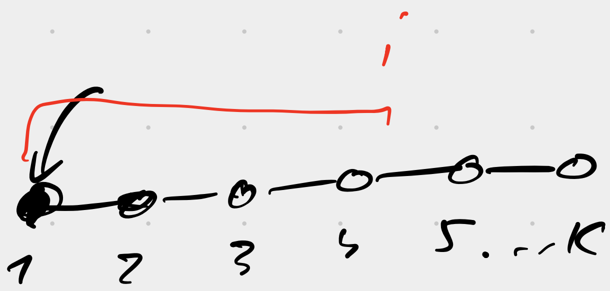
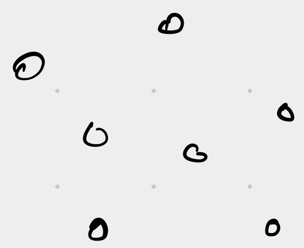
Lower complexity Bound.

$$f_k(x) = \frac{1}{2}\langle A_k x, x\rangle - \langle b, x\rangle \qquad x \in \mathbb{R}^n$$

$$A_k = \left[\begin{array}{c|c} \Lambda_k & 0 \\ \hline 0 & I \end{array}\right] \in \mathbb{R}^{n\times n}$$



$$\Lambda_k = \begin{bmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & 0 & \\ & -1 & 2 & -1 & & \\ & & -1 & 2 & -1 & \\ & & & -1 & \ddots & -1 \\ & 0 & & & -1 & 2 \end{bmatrix} \in \mathbb{R}^{k\times k}$$



- $0 \preceq f_k''(\cdot) \preceq 4I$
- $f_k^* = -\frac{k}{2}$
- $R_k^2 = \|x_k^*\|^2 \leq \frac{(k+1)^3}{3}$
- $b = e_1$

Fix $k \geq 1$ iterations, $x_{k+1} \in \operatorname{span}\{f'(x_0), \dots, f'(x_k)\}$

$$f(x) \equiv f_{2k+1}(x) \qquad f'(x) = A_{2k+1}x - e_1$$

$$x_0 = [0\ 0\ 0\ 0\ \dots\ 0]$$
$$x_1 = [*\ 0\ 0\ 0\ \dots\ 0]$$
$$x_2 = [*\ *\ 0\ 0\ \dots\ 0]$$
$$x_3 = [*\ *\ *\ 0\ \dots\ 0] \quad \dots$$

- $x_k \in \mathbb{R}^{n,k} = \{ x \in \mathbb{R}^n \mid x^{(i)} = 0 \quad i \geq k+1 \}$

- $f_k(x) \equiv f_{k+p}(x) \quad \forall x \in \mathbb{R}^{n,k}, \quad p \geq 0$

  indistinguishable for the method during first k iterations.

$$f(x_k) = f_{2k+1}(x_k) = f_k(x_k) \geq f_k^* = -\frac{k}{2}$$

$$f^* = f_{2k+1}^* = -\frac{2k+1}{2}$$

$$R^2 = \|x^* - \underset{\overset{\shortparallel}{0}}{x_0}\|^2 = \|x_{2k+1}^*\|^2 \leq \frac{2^3(k+1)^3}{3}.$$

<u>Therefore</u>:

$$\frac{f(x_k) - f^*}{R^2} \geq \frac{\frac{2k+1}{2} - \frac{k}{2}}{R^2} = \frac{k+1}{2R^2} \geq \frac{3}{2^4(k+1)^2}.$$

<u>Remark</u>

$$\varphi(x) = \frac{L}{4} f(x) \quad - \text{convex, Lipschitz gradient with constant } L > 0.$$

<u>Theorem</u> Let $L > 0$. For any first-order algorithm that runs for $K \geq 1$ iterations, and such that:

$$x_{k+1} \in \operatorname{span}\{f'(x_0), \dots, f'(x_k)\},$$

$\exists$ a convex function: $f: \mathbb{R}^n \to \mathbb{R}$, $\quad n \geq 2K+1$, s.t.

- f has a Lipschitz gradient with const. $L > 0$.
- $$f(x_k) - f^* \geq \frac{3L\|x_0 - x^*\|^2}{2^6(k+1)^2}. \quad \square$$

Review: Strongly convex Functions.

$f:\mathbb{R}^n \to \mathbb{R}$:

$$f(y) \geq f(x) + \langle f'(x), y-x \rangle + \frac{\mu}{2}\|y-x\|^2 \qquad \star$$

$\mu > 0$

Assume f_1, f_2 that are strongly convex with $\mu_1, \mu_2 \geq 0$

Then $f(x) = f_1(x) + f_2(x) \Rightarrow \mu = \mu_1 + \mu_2$.

<u>Example</u> $d(y) = \frac{1}{2}\|y\|^2 \qquad d'(y) = y$

$$d(y) - d(x) - \langle d'(x), y-x \rangle =$$

$$= \frac{1}{2}\|y\|^2 - \frac{1}{2}\|x\|^2 - \langle x, y-x \rangle$$

$$= \frac{1}{2}\|y-x+x\|^2 - \frac{1}{2}\|x\|^2 - \langle x, y-x \rangle$$

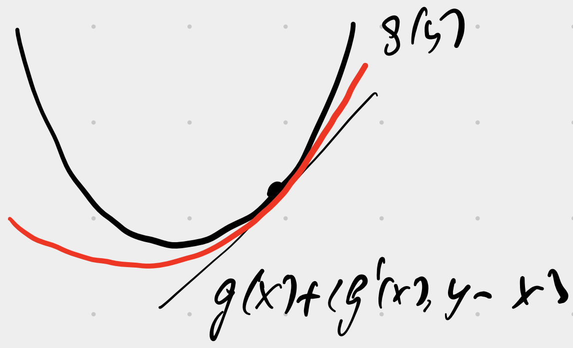
$$= \frac{1}{2}\|y-x\|^2 + \frac{1}{2}\|x\|^2 + \langle y-x, x \rangle - \frac{1}{2}\|x\|^2 - \langle x, y-x \rangle$$

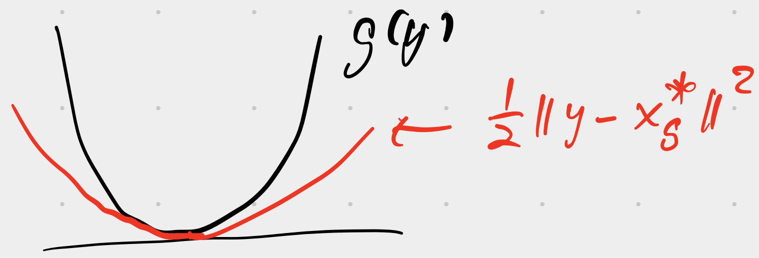
$$= \frac{1}{2}\|y-x\|^2.$$

$\Rightarrow d(y)$ satisfies $\star$ as <u>equality</u>, $\mu = 1$.

<u>Corollary</u> $g(y) = f(y) + \frac{1}{2}\|y\|^2$ Then

$$g(y) \geq g^* + \frac{1}{2}\|y - x_g^*\|^2 \quad \forall y \in \mathbb{R}^n. \qquad g^* \leq g(y)$$





Nesterov's Fast Gradient Method / Accelerated Gradient Method.

$$\min_{x \in \mathbb{R}^n} f(x)$$

- f - convex
- f - Lip. gradient.

Goal: $$f(x_k) - f^* \le \frac{\|x_0 - x^*\|^2}{2A_k},$$

where $A_k > 0$ a growing sequence $\nearrow$. $A_k \simeq \frac{k^2}{L}$.

A more goal: $\forall x \in \mathbb{R}^n$

$$f(x_k) - f(x) \le \frac{\|x_0 - x\|^2}{2A_k}$$

$$\Leftrightarrow \quad \underbrace{\frac{\|x_0 - x\|^2}{2} + A_k f(x)}_{m_k(x) \text{ is strongly convex}} \ge A_k f(x_k) \quad \forall x \in \mathbb{R}^n$$

$$\min_x m_k(x) = m_k^* \ge A_k f(x_k)$$

$$m_k(x) \ge m_k^* + \frac{1}{2}\|x - x_{m_k}^*\|^2 \ge A_k f(x_k) + \frac{1}{2}\|x - x_{m_k}^*\|^2.$$

Refined goal: Construct $\{x_k\}, \{v_k\} \subset \mathbb{R}^n$, $A_k \nearrow$

s.t.

$$\frac{1}{2}\|x - x_0\|^2 + A_k f(x) \ge \frac{1}{2}\|x - v_k\|^2 + A_k f(x_k) \quad \forall x \in \mathbb{R}^n$$

Init: $A_0 = 0$, $v_0 = x_0$ $\Rightarrow$ satisfied.

Define?

$\boxed{1.\ x_k}$

$\boxed{2.\ v_k}$

$\boxed{3.\ A_k}$

Fix $k \geq 0$.

$A_{k+1} = A_k + a_{k+1}$, $\quad a_{k+1} > 0$. — our "progress"

$$\frac{1}{2}\|x - x_0\|^2 + A_{k+1} f(x) = \frac{1}{2}\|x - x_0\|^2 + A_k f(x) + \overbrace{a_{k+1} f(x)}^{\text{"extra progress"}}$$

$$\geq \frac{1}{2}\|x - v_k\|^2 + A_k f(x_k) + a_{k+1} f(x)$$

$$= \frac{1}{2}\|x - v_k\|^2 + A_{k+1}\left[\gamma_k f(x) + (1-\gamma_k) f(x_k) \right]$$

where $\gamma_k = \frac{a_{k+1}}{A_{k+1}} \in [0,1]$

$$\underset{\text{convexity}}{\geq} \frac{1}{2}\|x - v_k\|^2 + A_{k+1}\left[f(y) \right] \geq$$

where $y = \gamma_k x + (1-\gamma_k) x_k$

$$\geq \underbrace{\frac{1}{2}\|x - v_k\|^2 + A_{k+1}\left[f(y_k) + \langle f'(y_k), y - y_k \rangle \right]}_{m_k(x) \to \min_x} \geq$$

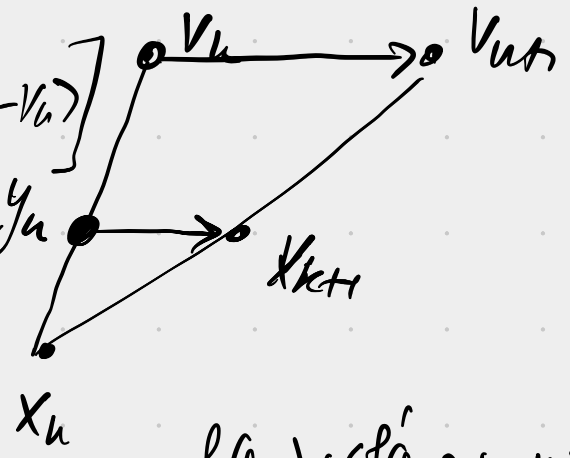
$y_k = \gamma_k v_k + (1-\gamma_k) x_k$

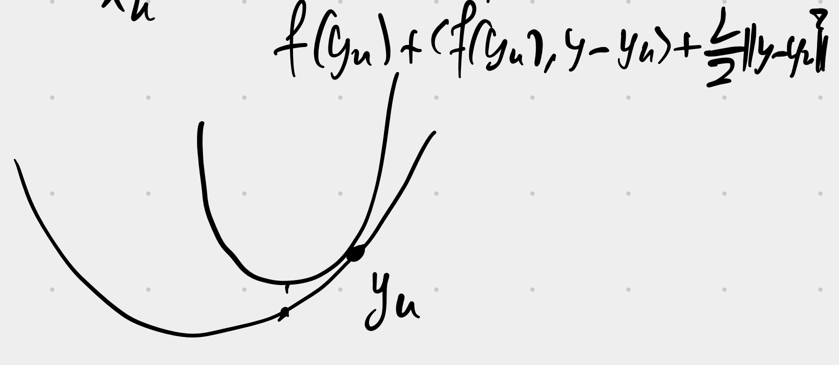
$y - y_k = \gamma_k (x - v_k)$

Define: $v_{k+1} = \arg\min_x m_k(x)$

$$= \arg\min \left[\frac{1}{2}\|x - v_k\|^2 + \underbrace{\gamma_k A_{k+1}}_{= a_{k+1}} \langle f'(y_k), x - v_k \rangle \right]$$

$$= v_k - a_{k+1} f'(y_k).$$





$$\geq \frac{1}{2}\|x - v_{k+1}\|^2 + \underbrace{m_k(v_{k+1})}_{\overset{?}{\geq} A_{k+1} f(x_{k+1})}$$

$$m_k(v_{k+1}) = \frac{1}{2}\|v_{k+1} - v_k\|^2 + A_{k+1}\left[f(y_k) + \langle f'(y_k), \underbrace{\gamma_k (v_{k+1} - v_k)}_{x_{k+1} - y_k} \rangle \right] =$$

$\Rightarrow x_{k+1} = \gamma_k v_{k+1} + (1-\gamma_k) x_k$

$$= A_{k+1}\left[f(y_k) + \langle f'(y_k), x_{k+1} - y_k \rangle + \frac{1}{2 A_{k+1} \gamma_k^2} \|x_{k+1} - y_k\|^2 \right] \geq A_{k+1} f(x_{k+1})$$

as soon as $\frac{1}{A_{k+1}\gamma_k^2} = \frac{A_{k+1}^2}{A_{k+1}\cdot a_{k+1}^2} = \frac{A_{k+1}}{a_{k+1}^2} \geq L$.

Algorithm (Fast Gradient Method)

Init: $x_0 \in \mathbb{R}^n$. Set $v_0 = x_0$, $A_0 = 0$. Fix $K \geq 1$ number of iterations.

For $k = 0 \dots K-1$:

- Choose $a_{k+1} > 0$. Set $A_{k+1} = A_k + a_{k+1}$, $\gamma_k = \frac{a_{k+1}}{A_{k+1}} \in [0,1]$
- Set $y_k = \gamma_k v_k + (1-\gamma_k)x_k$, Compute $f'(y_k)$.
- Update: $v_{k+1} = v_k - a_{k+1} f'(y_k)$
- Set $x_{k+1} = \gamma_k v_{k+1} + (1-\gamma_k)x_k$.

Return x_K.

Theorem Let a_{k+1}: $\frac{A_{k+1}}{a_{k+1}^2} = \frac{A_k + a_{k+1}}{a_{k+1}^2} \geq L$. Then, $\forall x \in \mathbb{R}^n$:

$$\frac{1}{2}\|x - v_k\|^2 + A_k f(x_k) \leq \frac{1}{2}\|x - x_0\|^2 + A_k f(x).$$

Therefore, $f(x_k) - f^* \leq \frac{\|x^* - x_0\|^2}{2 A_k}$, $k \geq 1$.

The Parameter Choice

$$a_{k+1}^2 = \frac{1}{L} A_{k+1}.$$

$$\frac{A_k + a_{k+1}}{a_{k+1}^2} \geq L. \qquad a_{k+1} \nearrow$$

- $$\frac{A_k + a_{k+1}}{a_{k+1}^2} = L \iff L a_{k+1}^2 - a_{k+1} - A_k = 0$$

$$a_{k+1} = \frac{1 + \sqrt{1 + 4LA_k}}{2L} = \frac{1}{2L} \cdot \left[1 + \sqrt{1 + 4LA_k} \right] \geq \frac{1}{L}$$

agressive steps

$$\sqrt{A_{k+1}} - \sqrt{A_k} = \frac{A_{k+1} - A_k}{\sqrt{A_{k+1}} + \sqrt{A_k}} = \frac{a_{k+1}}{\sqrt{A_{k+1}} + \sqrt{A_k}} = \frac{\sqrt{A_{k+1}}}{\sqrt{L}\left(\sqrt{A_{k+1}} + \sqrt{A_k}\right)}$$

$$\geq \frac{1}{2\sqrt{L}}.$$

Telescope: $\sqrt{A_k} \geq \sqrt{A_0} + \frac{k}{2\sqrt{L}} = \frac{k}{2\sqrt{L}} \Rightarrow A_k \geq \frac{k^2}{4L}.$

Therefore: the rate of the FGM:

$$f(x_k) - f^* \leq \frac{2L \|x_0 - x^*\|^2}{k^2}, \qquad k \geq 1.$$

Other choice

$$a_k := \frac{1}{2L} k.$$

$$A_k = \sum_{i=1}^{k} a_i = \frac{1}{2L} \cdot \frac{(k+1)k}{2}$$

To check: $\frac{A_{k+1}}{a_{k+1}^2} \overset{?}{\geq} L$ $\qquad \frac{1}{2L} \cdot \frac{(k+1) \cdot k}{2 \cdot k^2} 4L^2 = \frac{k+1}{k} \cdot L \geq L.$