Suppose $f(x)$ is c-strongly cvx, L-smooth

& x^* is in a set B, e.g., the set of vectors with at most s-nonzeros.

Consider Averaged Projected Gradient Descent:

$$\boxed{\begin{aligned} y^{(k)} &= \pi_B\left(x^{(k)} - \frac{1}{\mu L}\nabla f(x^{(k)})\right) \\ x^{(k+1)} &= (1-\mu)x^{(k)} + \mu y^{(k)} \end{aligned}} \qquad (\mu \in (0,1))$$

Remark: Note that

$$y^{(k)} = \underset{y \in B}{\arg\min} \left\| y - \left(x^{(k)} - \frac{1}{\mu L}\nabla f(x^{(k)})\right)\right\|^2$$

$$= \underset{y \in B}{\arg\min} \langle \nabla f(x^{(k)}), y - x^{(k)} \rangle + \frac{\mu L}{2}\| y - x^{(k)}\|^2 \quad (1)$$

$$\boxed{\text{Thm: If } \mu \in (0, \tfrac{1}{\kappa}) \text{ where } \kappa = \tfrac{L}{c}, \text{ then } f(x^{(k)}) - f(x^*) \le (1-\mu)^k (f(x^{(0)}) - f(x^*))}$$

Proof:

$$f(x^{(k+1)}) = f(x^{(k)} + \mu(y^{(k)} - x^{(k)})) \overset{L\text{-Smooth.}}{\le} f(x^{(k)}) + \mu \langle \nabla f(x^{(k)}), y^{(k)} - x^{(k)} \rangle + \frac{\mu^2 L}{2} \| y^{(k)} - x^{(k)} \|^2$$

optimality of $y^{(k)}$ in ①

$$\le f(x^{(k)}) + \mu \langle \nabla f(x^{(k)}), x^* - x^{(k)} \rangle + \frac{\mu^2 L}{2} \| x^* - x^{(k)} \|^2$$

Strong convexity

$f(x^*) \ge f(x^{(k)}) + \langle \nabla f(x^{(k)}), x^{(k)} - x^* \rangle + \frac{c}{2} \| x^{(k)} - x^* \|^2$

$$\le f(x^{(k)}) + \mu \left(f(x^*) - f(x^{(k)}) - \frac{c}{2} \| x^* - x^{(k)} \|^2 \right) + \frac{\mu^2}{2} \| x^* - x^{(k)} \|^2$$

$$= (1-\mu)(f(x^{(k)}) - f(x^*)) + \underbrace{\left(-\frac{c\mu}{2} + \frac{\mu^2 L}{2} \right) \| x^* - x^{(k)} \|^2}_{\mu \le \frac{1}{\kappa} \text{ makes this } \le 0} + f(x^*)$$

$$\le (1-\mu)(f(x^{(k)}) - f(x^*)) + f(x^*)$$

$$\Longleftrightarrow \quad f(x^{(k+1)}) - f(x^*) \le (1-\mu)\left(f(x^{(k)}) - f(x^*)\right)$$

$$\le \cdots \le (1-\mu)^{(k+1)}\left(f(x^{(0)}) - f(x^*)\right) \qquad \square$$