HW 1 EE546

Q1. Non convex Optimization

$f : R^n \to R$

x^* is f's global minima

$\forall\ x \in R^n$, we have

$$\| x - x^* \|_{\ell_2} \leq R$$

We have also,

$$< \nabla f(x), x - x^* > \; \geq \frac{1}{\alpha} \| x - x^* \|_{\ell_2}^2 + \frac{1}{\beta} \| \nabla f(x) \|_{\ell_2}^2$$

for $\exists\ \alpha > 0$

Also

$$\| x_0 - x^* \|_{\ell_2} \leq R$$

$$\mu \in (0, 2/\beta)$$

$$x_{\tau+1} = x_\tau - \mu \nabla f(x_\tau)$$

$\hookrightarrow$ Gradient Descent Update

Prove $\forall\, \tau$, we have

$$\|x_\tau - x^*\|_{\ell_2}^2 \leq \left(1 - \frac{2\rho}{\alpha}\right)^\tau \|x_0 - x^*\|_{\ell_2}^2$$

Hint:

$$\|x_{\tau+1} - x^*\|^2 =$$

$$\|x_\tau - \rho \nabla f(x_\tau) - x^*\|^2$$

$$= (x_\tau - \rho \nabla f(x_\tau) - x^*)^T (x_\tau - \rho \nabla f(x_\tau) - x^*)$$

$$= \left((x_\tau - x^*)^T - [\rho \nabla f(x_\tau)]^T\right)\left((x_\tau - x^*) - \rho \nabla f(x_\tau)\right]$$

$$= (x_\tau - x^*)^T (x_\tau - x^*) - (x_\tau - x^*)^T \rho \nabla f(x_\tau) - \rho [\nabla f(x_\tau)]^T (x_\tau - x^*) + \rho^2 [\nabla f(x_\tau)]^T \nabla f(x_\tau)$$

$$= \|x_\tau - x^*\|^2 - 2\rho\, [\nabla f(x_\tau)]^T (x_\tau - x^*) + \rho^2 \|\nabla f(x_\tau)\|^2.$$

Since,

$$[\nabla f(x)]^T(x - x^*) \geq \frac{1}{\alpha}\|x - x^*\|^2 + \frac{1}{\beta}\|\nabla f(x)\|^2$$

is true

$\forall x \in \mathbb{R}^n$ ($\because$ assumption), it must be true for an x_T iterate

So $$[\nabla f(x_T)]^T(x_T - x^*) \geq \frac{1}{\alpha}\|x_T - x^*\|^2 + \frac{1}{\beta}\|\nabla f(x_T)\|^2$$

Also we know that

$$\|x_{T+1} - x^*\|^2 = \|x_T - x^*\|^2 + \rho^2\|\nabla f(x_T)\|^2 - 2\rho[\nabla f(x_T)]^T(x_T - x^*)$$

So,

$$[\nabla f(x_T)]^T(x_T - x^*) =$$

$$\frac{1}{2\rho}\left\{\|x_T - x^*\|^2 - \|x_{T+1} - x^*\|^2\right\} + \rho^2\|\nabla f(x_T)\|^2$$

$$\therefore \frac{1}{2\rho}\left\{(\|x_T - x^*\|^2 - \|x_{T+1} - x^*\|^2) + \rho^2\|\nabla f(x_T)\|^2\right\}$$

$$\geq \frac{1}{\alpha}\|x_T - x^*\|^2 + \frac{1}{\beta}\|\nabla f(x_T)\|^2$$

$$\frac{1}{2\rho}\left[\|x_\tau - x^*\|^2 - \|x_{\tau+1} - x^*\|^2\right] + \frac{\rho}{2}\|\nabla f(x_\tau)\|^2$$

$$\geq \frac{1}{\alpha}\|x_\tau - x^*\|^2 + \frac{1}{\beta}\|\nabla f(x_\tau)\|^2$$

$$\rho \in \left(0, \frac{2}{\beta}\right), \quad \frac{\rho}{2} \in \left(0, \frac{1}{\beta}\right)$$

for $\beta > 0$, $\quad \frac{\rho}{2} < \frac{1}{\beta}$

So $\quad \frac{\rho}{2}\|\nabla f(x_\tau)\|^2 < \frac{1}{\beta}\|\nabla f(x_\tau)\|^2$

$$\Downarrow$$

$$\frac{1}{2\rho}\left[\|x_\tau - x^*\|^2 - \|x_{\tau+1} - x^*\|^2\right] + \frac{\rho}{2}\|\nabla f(x_\tau)\|^2$$

$$\geq \frac{1}{\alpha}\|x_\tau - x^*\|^2 + \frac{\rho}{2}\|\nabla f(x_\tau)\|^2$$

$$\therefore \quad \frac{1}{2\rho}\left[\|x_\tau - x^*\|^2 - \|x_{\tau+1} - x^*\|^2\right] \geq \frac{1}{\alpha}\|x_\tau - x^*\|^2$$

as τ is arbitrary we have

$\tau \in \{0, 1, 2, \ldots\}$.

$$\frac{1}{2\rho}\left[\|x_T - x^*\|^2 - \|x_{T+1} - x^*\|^2\right] \geq \frac{1}{\alpha}\|x_T - x^*\|^2$$

$$\downarrow$$

$$\frac{1}{2\rho}\left[\|x_{T-1} - x^*\|^2 - \|x_T - x^*\|^2\right] \geq \frac{1}{\alpha}\|x_{T-1} - x^*\|^2$$

$$\downarrow$$

$$\frac{1}{2\rho}\left[\|x_0 - x^*\|^2 - \|x_1 - x^*\|^2\right] \geq \frac{1}{\alpha}\|x_0 - x^*\|^2$$

Now, Adding all inequalities, we have,

$$\frac{1}{2\rho}\left[\|x_0 - x^*\|^2 - \|x_{T+1} - x^*\|^2\right] \geq \frac{1}{\alpha}\sum_{i=0}^{i=T}\|x_i - x^*\|^2$$

$$\frac{1}{2\rho}\|x_0 - x^*\|^2 - \frac{1}{2\rho}\|x_{T+1} - x^*\|^2 \geq \frac{1}{\alpha}\sum_{i=0}^{i=T}\|x_i - x^*\|^2$$

$$\frac{1}{2\rho}\|x_0 - x^*\|^2 \geq \frac{1}{2\rho}\|x_{T+1} - x^*\|^2 + \frac{1}{\alpha}\sum_{i=0}^{i=T}\|x_i - x^*\|^2$$

$$\frac{1}{2\rho}\|x_{T+1} - x^*\|^2 + \frac{1}{\alpha}\sum_{i=0}^{i=T}\|x_i - x^*\|^2 \leq \frac{1}{2\rho}\|x_0 - x^*\|^2$$

Again as τ is arbitrary,

$\tau \rightarrow \tau - 1$

$$\frac{1}{2\rho} \| x_\tau - x^* \|^2 + \frac{1}{\alpha} \sum_{i=0}^{\tau-1} \| x_i - x^* \|^2 \leq \frac{1}{2\rho} \| x_0 - x^* \|^2$$

So, $$\frac{1}{2\mu} \| x_\tau - x^* \|^2 \leq \frac{1}{2\mu} \| x_0 - x^* \|^2 - \frac{1}{\alpha} \sum_{i=0}^{\tau-1} \| x_i - x^* \|^2$$

$$\| x_\tau - x^* \|^2 \leq \| x_0 - x^* \|^2 - \frac{2\mu}{\alpha} \sum_{i=0}^{\tau-1} \| x_i - x^* \|^2$$

So

$$\| x_\tau - x^* \|^2 \leq \underbrace{\| x_0 - x^* \|^2 - \frac{2\rho}{\alpha} \sum_{i=0}^{\tau-1} \| x_i - x^* \|^2}$$

Result 1.

(Continued).

To prove

$$\longrightarrow \quad \|x_{t+1} - x^*\|^2 \leq \left(1 - \frac{2p}{\alpha}\right)^{t+1} \|x_0 - x^*\|^2$$

Now,

$$\|x_{t+1} - x^*\|^2 \rightarrow$$

$$\|x_t - p\nabla f(x_t) - x^*\|^2$$

$$\rightarrow \|(x_t - x^*) - p\nabla f(x_t)\|^2$$

$$= \{(x_t - x^*) - p\nabla f(x_t)\}^T \{(x_t - x^*) - p\nabla f(x_t)\}$$

$$= \{(x_t - x^*)^T - [p\nabla f(x_t)]^T\}\{(x_t - x^*) - p\nabla f(x_t)\}$$

$$= (x_t - x^*)^T(x_t - x^*) + p^2[\nabla f(x_t)]^T[\nabla f(x_t)] - 2p(x_t - x^*)^T[\nabla f(x_t)]$$

$$= \|x_t - x^*|^2 + p^2\|\nabla f(x_t)\|^2 - 2p(x_t - x^*)^T[\nabla f(x_t)]$$

Now,

$$\|x_\tau - x^*\|^2 + p^2 \|\nabla f(x_\tau)\|^2$$
$$- 2p(x_\tau - x^*)^T [\nabla f(x_\tau)]$$

Now, we have,

Now $[\nabla f(x_\tau)]^T (x_\tau - x^*)$

$$\geq \frac{1}{\alpha} \|x_\tau - x^*\|^2 + \frac{1}{\beta} \|\nabla f(x)\|^2$$

$$\geq \frac{1}{\alpha} \|x_\tau - x^*\|^2 + \frac{1}{\beta} \|\nabla f(x)\|^2$$

$$[\nabla f(x_\tau)]^T (x_\tau - x^*) \geq \frac{1}{\alpha} \|x_\tau - x^*\|^2 + \frac{1}{\beta} \|\nabla f(x)\|^2$$

$$- [\nabla f(x_\tau)]^T (x_\tau - x^*) \leq -\frac{1}{\alpha} \|x_\tau - x^*\|^2 - \frac{1}{\beta} \|\nabla f(x)\|^2$$

$$- 2p [\nabla f(x_\tau)]^T (x_\tau - x^*) \leq -\frac{2p}{\alpha} \|x_\tau - x^*\|^2 - \frac{2p}{\beta} \|\nabla f(x)\|^2$$

$$\|x_\tau - x^*\|^2 + p^2 \|\nabla f(x_\tau)\|^2 - 2p [\nabla f(x_\tau)]^T (x_\tau - x^*) \leq \|x_\tau - x^*\|^2 - \frac{2p}{\beta} \|\nabla f(x)\|^2 - \frac{2p}{\alpha} \|x_\tau - x^*\|^2 + p^2 \|\nabla f(x_\tau)\|^2$$

So $$\|x_{\tau+1} - x^*\|^2 \leq \|x_\tau - x^*\|^2 \left(1 - \frac{2p}{\alpha}\right) + p^2 \|\nabla f(x_\tau)\|^2 - \frac{2p}{\beta} \|\nabla f(x)\|^2$$

$$\|x_{\tau+1} - x^*\|^2 \leq \|x_\tau - x^*\|^2 \left(1 - \frac{2\mu}{\alpha}\right) + \mu^2 \|\nabla f(x_\tau)\|^2 - \frac{2\mu}{\beta} \|\nabla f(x_\tau)\|^2$$

$$\left(\mu^2 - \frac{2\mu}{\beta}\right) \|\nabla f(x_\tau)\|^2$$

$$\Downarrow$$

$$\mu^2 \left[1 - \frac{2}{\beta\mu}\right] \|\nabla f(x_\tau)\|^2$$

$$\mu < \frac{2}{\beta}$$

$$1 < \frac{2}{\beta\mu} \longrightarrow \frac{2}{\beta\mu} > 1 \longrightarrow -\frac{2}{\beta\mu} < -1$$

$$\longrightarrow 1 - \frac{2}{\beta\mu} < 0$$

$$\mu^2 \|\nabla f(x_\tau)\|^2 \geq 0$$

Now, $\left(1 - \frac{2}{\beta\mu}\right) \mu^2 \|\nabla f(x_\tau)\|^2 \leq 0$

$$\|x_\tau - x^*\|^2 \left(1 - \frac{2\mu}{\alpha}\right) + \left(1 - \frac{2}{\beta\mu}\right) \mu^2 \|\nabla f(x_\tau)\|^2$$

$$\leq \|x_\tau - x^*\|^2 \left(1 - \frac{2\mu}{\alpha}\right)$$

So $\|x_{\tau+1} - x^*\|^2 \leq \|x_\tau - x^*\|^2 \left(1 - \frac{2\mu}{\alpha}\right)$

$\therefore$ Replacing $\tau+1 \longrightarrow \tau$

$$\|x_\tau - x^*\|^2 \leq \|x_{\tau-1} - x^*\|^2 \left(1 - \frac{2\mu}{\alpha}\right).$$

$$\|x_{\tau-1} - x^*\|^2 \leq \|x_{\tau-2} - x^*\|^2 \left(1 - \frac{2\mu}{\alpha}\right)$$

$$\therefore \|x_{\tau-1} - x^*\|^2 \left(1 - \frac{2\mu}{\alpha}\right) \leq \|x_{\tau-2} - x^*\|^2 \left(1 - \frac{2\mu}{\alpha}\right)^2$$

$$\left(\because \left\{1 - \frac{2\mu}{\alpha}\right\} \geq 0\right)$$

$\therefore$ We get,

$$\|x_\tau - x^*\|^2 \leq \|x_{\tau-2} - x^*\|^2 \left(1 - \frac{2\mu}{\alpha}\right)^2$$

$\downarrow$ Continuing similary,

$$\|x_\tau - x^*\|^2 \leq \|x_0 - x^*\|^2 \left(1 - \frac{2\mu}{\alpha}\right)^\tau$$

$\forall \tau \longrightarrow$ Thus proved.

Q.

$f: R^n \to R$ is L-Smooth
(convex / non convex)

$x_{t+1} = x_t - p \nabla f(x_t)$.

if $p \le \frac{1}{L}$ & function bounded from below

then $\lim_{t \to +\infty} \|\nabla f(x_t)\| \to 0$

Hint :

$$f(x - p\nabla f(x)) \le f(x) - p\left(1 - \frac{pL}{2}\right)\|\nabla f(x)\|^2$$

for this function

(a) Function bounded from below
(a finite lower bound exists for function)

Now,

$$f(x_t - p\nabla f(x_t)) \le f(x_t) - p\left(1 - \frac{pL}{2}\right)\|\nabla f(x_t)\|^2$$

$$f(x_{t+1}) - f(x_t) \le -p\left(1 - \frac{pL}{2}\right)\|\nabla f(x_t)\|^2$$

$$-p\left(1 - \frac{pL}{2}\right)\|\nabla f(x_t)\|^2 \ge f(x_{t+1}) - f(x_t)$$

$$\rho \leq \frac{1}{L}$$

$$\therefore \frac{\rho L}{2} \leq \frac{1}{2}$$

$$\therefore -\frac{\rho L}{2} \geq -\frac{1}{2}$$

$$\left(1-\frac{\rho L}{2}\right) \geq \frac{1}{2}$$

So $1-\frac{\rho L}{2} \geq \frac{1}{2} > 0$

$$\therefore 1-\frac{\rho L}{2} > 0$$

$$\therefore -\rho\left(1-\frac{\rho L}{2}\right) < 0$$

$$\therefore \rho\left(1-\frac{\rho L}{2}\right) > 0$$

So we have,

$$\rho\left(1-\frac{\rho L}{2}\right) \|\nabla f(x_t)\|^2 \leq f(x_t) - f(x_{t+1})$$

$$\therefore \|\nabla f(x_t)\|^2 \leq \frac{1}{\rho\left(1-\frac{\rho L}{2}\right)} \left(\cancel{f(x_t)} - f(x_{t+1})\right)$$

again t is arbitrary

$$+ \qquad\qquad\qquad\qquad +$$

$$\|\nabla f(x_{t-1})\|^2 \leq \frac{1}{\rho\left(1-\frac{\rho L}{2}\right)} \left(\cancel{f(x_{t-1})} - \cancel{f(x_t)}\right)$$

$$+ \qquad\qquad\qquad\qquad +$$

$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow$$

$$+ \qquad\qquad\qquad\qquad +$$

$$\|\nabla f(x_0)\|^2 \leq \frac{1}{\rho\left(1-\frac{\rho L}{2}\right)} \left(f(x_0) - \cancel{f(x_1)}\right)$$

Telescoping, we get,

$$\sum_{i=0}^{i=t} \|\nabla f(x_i)\|^2 \leq \frac{1}{p(1-\frac{pL}{2})} \left[f(x_0) - f(x_{t+1}) \right]$$

So

$$\sum_{i=0}^{i=t} \|\nabla f(x_i)\|^2 \leq \frac{1}{p(1-\frac{pL}{2})} \left[f(x_0) - f(x_{t+1}) \right]$$

letting $t \to \infty$, we have,

function Bounded from below

$$\sum_{i=0}^{\infty} \|\nabla f(x_i)\|^2 \leq \frac{1}{p(1-\frac{pL}{2})} \left[\overbrace{f(x_0) - f(x^*)} \right]$$

Function bounded from below, finite minimal optima exist

required optimal point

As series is bounded, it is convergent.

→ A necessary condition for series to be convergent is that its infinite term tends to zero.

Thus $\lim_{t \to +\infty} \|\nabla f(x_t)\|^2 \to 0$

Thus $\lim_{t \to \infty} \|\nabla f(x_t)\| \to 0$

So we have

$$\lim_{t \to \infty} \| \nabla f(x_t) \| \longrightarrow 0$$

($\because$ property of infinite series).

b) Under same assumptions of (a) does x_t converges to a fixed point ?

$$x_{t+1} = x_t - \mu \nabla f(x_t).$$

Does x_t converge to a point fixed!

$$\| \nabla f(x) - \nabla f(y) \| \leq L \| x - y \|$$

$\downarrow$

L smoothness

$$\| x_{t+1} - x^* \| = \| x_t - \mu \nabla f(x_t) - x^* \|$$

Now

$$\| \nabla f(x) - \nabla f(y) \| \leq L \| x - y \|$$

$x \rightarrow x_0$
$y \rightarrow x^*$

$$\| \nabla f(x_0) - \nabla f(x^*) \| \leq L \| x_0 - x^* \|$$

for local optimal point $\nabla f(x^*) = 0$

(same point where gradient norm goes to zero, Euclidean norm zero only for a zero vector)

$\therefore \quad \| \nabla f(x_0) \| \leq L \| x_0 - x^* \|$

Also $\quad \| \nabla f(x_k) \| \leq L \| x_k - x^* \|$

$$\| x_{k+1} - x^* \| = \| x_k - \mu \nabla f(x_k) - x^* \|$$

$$= \| (x_k - x^*) - \mu \nabla f(x_k) \|$$

Using previously proven results convergence of $x_t \longrightarrow x^*$ can be validated easily

Now we know that,

$$x_{t+1} = x_t - \rho \nabla f(x_t)$$

Also as $t \to \infty$, $\|\nabla f(x_t)\| \to 0$
and thus $\nabla f(x_t) \to \vec{0}$

This implies that for a large enough t x_t does converge to a point x^* where $\|\nabla f(x^*)\| \to 0$

Also $x_{t+1} - x_t = -\rho \nabla f(x_t)$

$$\|x_{t+1} - x_t\| = \rho \|\nabla f(x_t)\|.$$

$$\lim_{t\to\infty} \|x_{t+1} - x_t\| = \rho \lim_{t\to\infty} \|\nabla f(x_t)\|$$

$$= 0$$

$$\therefore \lim_{t\to\infty} \|x_{t+1} - x_t\| \to 0$$

So $x_{t+1} \longrightarrow x_t \longrightarrow x^*$ for a large enough t

So x_t converges to x^*

Additionally for our L-smooth function, after finite iterations we find its norm being upper bounded.

To obtain it upper bounded by ε, we need iterations of $O\left(\frac{1}{\varepsilon^2}\right)$

→ Thus we always find a corresponding convergent x sequence which converges to x' in $O\left(\frac{1}{\varepsilon^2}\right)$ iterations bounding gradient norm by ε.

→ Again there are more stringent proofs involving bounding,

$$\|(x-z) - [\nabla f(x) - \nabla f(z)]\| \leq \beta \|x - z\|$$

which can give us a linear convergence to x^* (some optimal local/global) however this requires $\beta \in [0,1)$ & needs some conditions more stringent than just L-smoothness.

→ Thus x_t does converge.

c) $\|f(z) - f(x)\| \leq L\|z - x\|$ → L smooth

$\|\nabla f(x)\|^2 \geq \gamma\left(f(x) - f(x^*)\right)$ → PL

$x^* \longrightarrow$ Global Optima (assumed)

$\mu \leq \frac{1}{L}$, then prove,

$$f(x_{t+1}) - f(x^*) \leq \left(1 - \frac{\mu\gamma}{2}\right)\left(f(x_t) - f(x^*)\right)$$

So $f(x_{t+1}) \leq f(x_t) - \mu\left(1 - \frac{\mu L}{2}\right)\|\nabla f(x_t)\|^2$

$\therefore f(x_{t+1}) - f(x_t) \leq -\mu\left(1 - \frac{\mu L}{2}\right)\|\nabla f(x_t)\|^2$

$$-\mu\left(1 - \frac{\mu L}{2}\right)\|\nabla f(x_t)\|^2 \leq -\mu\gamma\left(1 - \frac{\mu L}{2}\right)\left(f(x_t) - f(x^*)\right)$$

(By PL)

$\therefore f(x_{t+1}) - f(x_t)$

$$\leq -\mu\gamma\left(1 - \frac{\mu L}{2}\right)\left(f(x_t) - f(x^*)\right)$$

$\frac{\mu L}{2} \leq \frac{1}{2}$

$-\frac{\mu L}{2} \geq -\frac{1}{2}$

$1-\frac{\mu L}{2} \geq \frac{1}{2}$

$\therefore \; -\mu\gamma\left(1-\frac{\mu L}{2}\right) \leq -\mu\gamma \cdot \frac{1}{2}$

$f(x_t) \geq f(x^*)$ (optimal point).

$\therefore \; f(x_t) - f(x^*) \geq 0$

$\therefore \; -\mu\gamma\left(1-\frac{\mu L}{2}\right)\left(f(x_t)-f(x^*)\right)$

$\leq -\frac{\mu\gamma}{2}\left(f(x_t)-f(x^*)\right).$

$\therefore \; f(x_{t+1}) - f(x_t) \leq -\frac{\mu\gamma}{2}\left(f(x_t)-f(x^*)\right)$

$\therefore \; f(x_{t+1}) - \cancel{f(x_t)}$
$+ \cancel{f(x_t)} - f(x^*) \leq -\frac{\mu\gamma}{2}\left(f(x_t)-f(x^*)\right)$
$+ f(x_t) - f(x^*)$

$\therefore \; f(x_{t+1}) - f(x^*) \leq \left(-\frac{\mu\gamma}{2}+1\right)\left(f(x_t)-f(x^*)\right).$

$$\therefore f(x_{t+1}) - f(x^*) \le \left(1 - \frac{\mu\gamma}{2}\right)\left(f(x_t) - f(x^*)\right).$$

Proved.