Convergen of Gradient Descent

$\min_x f(x)$ Alg: $x^{(k+1)} = x^{(k)} - t\nabla f(x^{(k)})$

$x^+ = x - t\nabla f(x)$

By L-gradient Lipschitz

$$f(x^+) \le f(x) + (x^+ - x)^T \nabla f(x) + \frac{L}{2}\|x^+ - x\|_2^2$$

$$= f(x) - t\nabla f(x)^T \nabla f(x) + \frac{L}{2}\|t\nabla f(x)\|^2$$

$$= f(x) - t\|\nabla f(x)\|^2 + \frac{t^2 L}{2}\|\nabla f(x)\|^2$$

$$= f(x) - \left(\frac{2t - t^2 L}{2}\right)\|\nabla f(x)\|^2$$

$t \le \frac{1}{L}$

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

$$= f(x) - \frac{t}{2}\|\nabla f(x)\|^2$$

Lem: (Descent Lemma) if $t \le \frac{1}{L}$, f obeys L-smoothness

then $\forall x$, $f(x^+) \le f(x) - \frac{t}{2}\|\nabla f(x)\|^2$

let's apply convexity $\quad f(x^+) \le \underset{\sim}{f(x)} - \frac{t}{2}\|\nabla f(x)\|^2$

$$f(x^*) \ge f(x) + \nabla f(x)^T (x^* - x)$$

Descent Lemma $\rightarrow$ $$\ge f(x^+) + \underbrace{\underbrace{\frac{t}{2}\|\nabla f(x)\|^2} + \underbrace{\nabla f(x)^T (x^* - x)} + \frac{1}{2t}\|x^* - x\|^2}_{} \quad - \frac{1}{2t}\|x^* - x\|^2$$

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

$$= f(x^+) + \frac{1}{2t}\|x^* \underbrace{- x + t\nabla f(x)}_{-x^+}\|^2 - \frac{1}{2t}\|x^* - x\|^2$$

$$= f(x^+) + \underbrace{\frac{1}{2t}\|x^* - x^+\|^2} - \underbrace{\frac{1}{2t}\|x^* - x\|^2}$$

Telescope: $x_0, x_1, \dots, x_k$

$$\left.\begin{aligned}
f(x^*) &\ge f(x_1) + \cancel{\frac{1}{2t}\|x^* - x_1\|^2} - \frac{1}{2t}\|x^* - x_0\|^2 \\
f(x^*) &\ge f(x_2) + \cancel{\frac{1}{2t}\|x^* - x_2\|^2} - \cancel{\frac{1}{2t}\|x^* - x_1\|^2} \\
&\vdots \\
f(x^*) &\ge f(x_k) + \frac{1}{2t}\|x^* - x_k\|^2 - \cancel{\frac{1}{2t}\|x^* - x_{k-1}\|^2}
\end{aligned}\right\}\ \text{add up}$$

$$k f^* \geq \underline{\sum_{i=1}^{k} f(x_i)} + \frac{1}{2t}\left(\underline{\|x^* - x_k\|^2 - \|x^* - x_0\|^2}\right)$$

$$f(x_k) \leq \underbrace{\frac{1}{k}\sum_{i=1}^{k} f(x_i)} \leq f^* + \frac{1}{2tk}\|x^* - x_0\|^2$$

$$f(x_1) \geq f(x_2) \geq \cdots \geq f(x_k)$$

$$f(x_k) - f^* \leq \frac{1}{2tk}\|x^* - x_0\|^2 \qquad \square$$

$$\boxed{\begin{array}{l}\text{Thm: } t \leq \frac{1}{L},\ f \text{ is } L\text{-smooth, then} \\ \qquad f(x_k) - f^* \leq \frac{1}{2tk}\|x^* - x_0\|^2\end{array}}$$

Convergence of GD on non-convex but L-smooth f:

$$f(x^+) \le f(x) - \frac{t}{2}\|\nabla f(x)\|^2$$

for $k = 1, \dots, K$

$$\left.\begin{aligned}
\cancel{f(x_2)} &\le f(x_1) - \frac{t}{2}\|\nabla f(x_1)\|^2 \\
\cancel{f(x_3)} &\le \cancel{f(x_2)} - \frac{t}{2}\|\nabla f(x_2)\|^2 \\
&\vdots \\
\cancel{f(x_k)} &\le \cancel{f(x_{k-1})} - \frac{t}{2}\|\nabla f(x_{k-1})\|^2 \\
f(x_{k+1}) &\le \cancel{f(x_k)} - \frac{t}{2}\|\nabla f(x_k)\|^2
\end{aligned}\right\} \text{sum}$$

$$f(x_{k+1}) \le f(x_1) - \frac{t}{2}\underbrace{\sum_{i=1}^{k}\|\nabla f(x_i)\|^2}$$

$$\min_{i\in\{1,2,\dots,k\}} \|\nabla f(x_i)\|^2 \le \frac{1}{k}\sum_{i=1}^{k}\|\nabla f(x_i)\|^2 \le \frac{2(f(x_1) - f(x_{k+1}))}{tk}$$

$$\le \frac{2}{tk}(f(x_1) - f^*) \qquad f^* \le f(x_{k+1})$$

$$\boxed{\min_{i\in[k]} \|\nabla f(x_i)\| \le \sqrt{\frac{2(f(x_1) - f^*)}{tk}}}$$

Thm. f obeys m-PL condition $\Rightarrow$ linear convergence

$$f(x^k) \le f^* + \frac{L}{2}\left(1-\frac{m}{L}\right)^k \|x - x^*\|^2$$

Proof: By descent lemma:

$$f(x^k) \le f(x^{k-1}) - \frac{t}{2}\|\nabla f(x^{k-1})\|_2^2$$

Apply m-PL

$$\le f(x^{k-1}) - mt\left(f(x^{k-1}) - f^*\right)$$

PL-condition

$$\frac{1}{2}\|\nabla f(x)\|^2 \ge m(f(x)-f^*)$$

$$f(x^k) - f^* \le f(x^{k-1}) - f^* - mt\left(f(x^{k-1}) - f^*\right)$$

$$\le (1-mt)\left(f(x^{k-1}) - f^*\right)$$

$$f(x^k) - f^* \le (1-mt)^k\left(f(x^0) - f^*\right)$$

$t = \frac{1}{L}$ $\quad \left(1-\frac{m}{L}\right)$ $\quad \le \frac{L}{2}\|x^1 - x^*\|^2$ $\quad \square$

~~Rest~~ Lipschitz ~~Grad~~ Gradient

Strong Convexity $\Rightarrow$ PL Condition

Proof:

$$f^* \geq f(x) + \nabla f(x)^T (x^* - x) + \frac{m}{2}\|x^* - x\|^2$$

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

$$= f(x) + \underbrace{\cancel{\frac{1}{2m}\|m(x^* - x) + \nabla f(x)\|^2}}_{\geq 0} - \frac{1}{2m}\|\nabla f(x)\|^2$$

$$\frac{1}{2}\|\nabla f(x)\|^2 \geq m(f(x) - f^*)$$

$\square$