problem 1.2

A Function $f: R^d \longrightarrow R$ is L-smooth iff it is differentiable and it's gradient is L-Lipschitz continuous (usually w.r.t norm-2):

$$\forall x_1, x_2 \in \mathbb{R}^d, \quad \| \nabla f(x_2) - \nabla f(x_1) \|_2 \leq L \| x_2 - x_1 \|_2$$

For all x_1, x_2, prove that

a) $f(x_2) \leq f(x_1) + \nabla f^T(x_1)(x_2 - x_1) + \frac{L}{2} \| x_2 - x_1 \|_2^2$

solution:

Let define $g(t) \triangleq f(x_1 + t(x_2 - x_1))$

we know that:

$$\int_0^1 g'(t)\, dt = g(1) - g(0) = f(x_2) - f(x_1)$$

It then follows that:

$$f(x_2) - f(x_1) - \nabla f^T(x_1)(x_2 - x_1) = \int_0^1 \nabla f^T(x_1 + t(x_2 - x_1))(x_2 - x_1)\, dt - \nabla f^T(x_1)(x_2 - x_1)$$

$$\Rightarrow f(x_2) - f(x_1) - \nabla f^T(x_1)(x_2 - x_1) = \int_0^1 \left(\nabla f(x_1 + t(x_2 - x_1)) - \nabla f(x_1) \right)^T (x_2 - x_1)\, dt$$

from the cauchy-schartz inequality we can get:

$$\Rightarrow f(x_2) - f(x_1) - \nabla f^T(x_1)(x_2 - x_1) \leq \int_0^1 \| \nabla f(x_1 + t(x_2 - x_1)) - \nabla f(x_1) \|_2 \| x_2 - x_1 \|_2\, dt$$

f has L-Lipschitz continuous gradient

we then have:

$$\Rightarrow \quad f(x_2) - f(x_1) - \nabla f^T(x_1)(x_2 - x_1) \leq \int_0^1 t L \|x_2 - x_1\|_2 \cdot \|x_2 - x_1\|_2 \, dt$$

$$\leq \int_0^1 t L \|x_2 - x_1\|_2^2 \, dt$$

$$\leq L \|x_2 - x_1\|_2^2 \int_0^1 t \, dt$$

$$\leq \frac{L}{2} \|x_2 - x_1\|_2^2$$

$$\Rightarrow \quad f(x_2) - f(x_1) - \nabla f^T(x_1)(x_2 - x_1) \leq \frac{L}{2} \|x_2 - x_1\|_2^2$$

$$\Rightarrow \quad f(x_2) \leq f(x_1) + \nabla f^T(x_1)(x_2 - x_1) + \frac{L}{2} \|x_2 - x_1\|_2^2$$

b) $f(x_2) \geqslant f(x_1) + \nabla f^T(x_1)(x_2 - x_1) + \frac{1}{2L}\|\nabla f(x_2) - \nabla f(x_1)\|_2^2$

Let define $g_{x_1}(x_2) \triangleq f(x_2) - f(x_1) - \nabla f^T(x_1)(x_2 - x_1)$

Since f is convex therefor: $f(x_2) \geqslant f(x_1) + \nabla f^T(x_1)(x_2 - x_1)$

$$\Rightarrow f(x_2) - f(x_1) - \nabla f^T(x_1)(x_2 - x_1) \geqslant 0$$

$$\Rightarrow g_{x_1}(x_2) \geqslant 0$$

In particular $g_{x_1}(x_1) = 0 \Rightarrow g_{x_1}(x_1) = \min_x g_{x_1}(x_2)$

and $\nabla g_{x_1}(x_1) = -\nabla f(x_1) + \nabla f(x_1) = 0$

from the optimality of x_1, it then follows that

$$g_{x_1}(x_1) \leqslant \min_\eta g_{x_1}(x_2 - \eta \nabla g_{x_1}(x_2))$$

$$(\star) \qquad = \min_\eta f(x_2 - \eta \nabla g_{x_1}(x_2)) - f(x_1) - \nabla f^T(x_1)(x_2 - \eta \nabla g_{x_1}(x_2) - x_1)$$

By deffinition of L-smooth we have:

$$f(x_2 - \eta \nabla g_{x_1}(x_2)) \leqslant f(x_2) + \nabla f^T(x_2)(-\eta \nabla g_{x_1}(x_2)) + \frac{L}{2}\|\eta \nabla g_{x_1}(x_2)\|_2^2$$

In the follows from $(\star)$ we have:

$$g_{x_1}(x_1) \leqslant \min_\eta f(x_2) + \nabla f^T(x_2)(-\eta \nabla g_{x_1}(x_2)) + \frac{L}{2}\|\eta \nabla g_{x_1}(x_2)\|_2^2$$
$$- f(x_1) - \nabla f^T(x_1)(x_2 - x_1 - \eta \nabla g_{x_1}(x_2))$$

$$g_{x_1}(x_1) \leq \min_{\eta} g_{x_1}(x_2) + \frac{L}{2}\|\eta \nabla g_{x_1}(x_2)\|_2^2 - \eta \nabla g_{x_1}^T(x_2)(\nabla f(x_2) - \nabla f(x_1))$$

$$g_{x_1}(x_1) \leq \min_{\eta} g_{x_1}(x_2) + \frac{L}{2}\eta^2 \|\nabla g_{x_1}(x_2)\|_2^2 - \eta \|\nabla g_{x_1}(x_2)\|_2^2$$

the minimum solution η to minimize this quadratic problem is:

$$L\eta^* \|\nabla g_{x_1}(x_2)\|_2^2 - \|\nabla g_{x_1}(x_2)\|_2^2 = 0$$

$$\eta^* = \frac{1}{L}$$

$\Rightarrow$ minimum solution : $g_{x_1}(x_2) - \frac{1}{2L}\|\nabla g_{x_1}(x_2)\|_2^2$

thus from our diffinition of $g_{x_1}(x_2)$ it follows:

$$g_{x_1}(x_1) \leq g_{x_1}(x_2) - \frac{1}{2L}\|\nabla g_{x_1}(x_2)\|_2^2$$

$$0 \leq f(x_2) - f(x_1) - \nabla f^T(x_1)(x_2 - x_1) - \frac{1}{2L}\|\nabla f(x_2) - \nabla f(x_1)\|_2^2$$

$$\Rightarrow -f(x_2) \leq -f(x_1) - \nabla f^T(x_1)(x_2 - x_1) - \frac{1}{2L}\|\nabla f(x_2) - \nabla f(x_1)\|_2^2$$

$$\Rightarrow f(x_2) \geq f(x_1) + \nabla f^T(x_1)(x_2 - x_1) + \frac{1}{2L}\|\nabla f(x_2) - \nabla f(x_1)\|_2^2$$

c) $\left(\nabla f(x_2) - \nabla f(x_1)\right)^T (x_2 - x_1) \geqslant \frac{1}{L} \| \nabla f(x_2) - \nabla f(x_1) \|_2^2$

Let define two convex functions f_{x_1}, f_{x_2} with R^n domain

$$\begin{cases} f_{x_1}(z) = f(z) - \nabla f^T(x_1) \cdot z \\ f_{x_2}(z) = f(z) - \nabla f^T(x_2) \cdot z \end{cases}$$

This two functions have L-Lipschitz continuous gradient
we know that if $f: R^n \to R$ and f has a minimizer
x^* then from the inequality in problem 1.2 b we have:

$$f(z) \geqslant f(x^*) + \cancel{\nabla f^T(x^*)}(z - x^*) + \frac{1}{2L} \| \nabla f(z) - \cancel{\nabla f(x^*)} \|_2^2$$

$$\Rightarrow f(z) - f(x^*) \geqslant \frac{1}{2L} \| \nabla f(z) \|^2$$

$\Rightarrow z = x_1$ minimize $f_{x_1}(z)$

$$(1) \quad f(x_2) - f(x_1) - \nabla f^T(x_1)(x_2 - x_1) = f_{x_1}(x_2) - f_{x_1}(x_1)$$
$$\geqslant \frac{1}{2L} \| \nabla f_{x_1}(x_2) \|_2^2$$
$$= \frac{1}{2L} \| \nabla f(x_2) - \nabla f(x_1) \|_2^2$$

similarly $z = x_2$ minimize $f_{x_2}(z)$

$$(2) \quad f(x_1) - f(x_2) - \nabla f^T(x_2)(x_1 - x_2) \geqslant \frac{1}{2L} \| \nabla f(x_2) - \nabla f(x_1) \|_2^2$$

now we can combine 1, 2 inequality:

$$\left(\nabla f(x_2) - \nabla f(x_1)\right)^T (x_2 - x_1) \geqslant \frac{1}{L} \| \nabla f(x_2) - \nabla f(x_1) \|_2^2$$