Lecture # 5   E1 260   Theory of convex functions

- Convex functions : Conjugate functions, monotone

- L – Smooth functions: Definition, Quadratic upper bound, second-order property, bound on optimality gap, co-coercivity (monotonicity)

- α – strongly convex functions: Definition, Quadratic lower bound, Hessian, bound on $f(x) - f^*$, Coercivity (monotonicity)

Ref : Beck, First-order methods in optimization.

Conjugate functions:

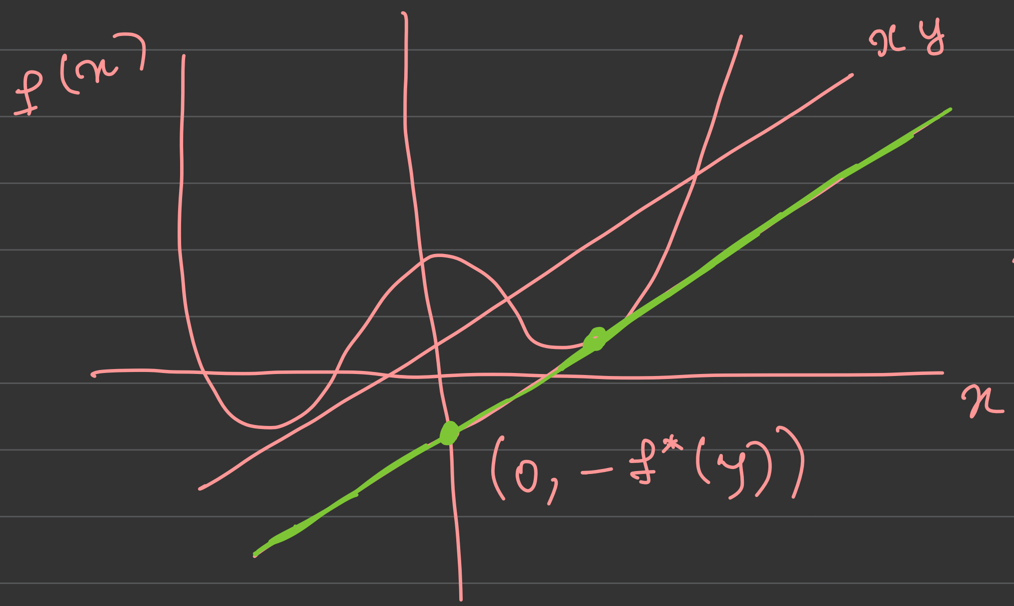
$f : E \to [-\infty, \infty]$ be an extended real-valued function.

$f^* : E^* \to [-\infty, \infty]$

E^* : dual space

$\|y\|_* = \max\{\langle y, x\rangle : \|x\| \le 1\}$

$$f^*(\underline{y}) = \max_{x \in \mathbb{E}} \{\langle \underline{y}, \underline{x}\rangle - f(x)\}$$



- maximum gap between the linear function yx and $f(x)$
- dom f^* consists of y for which max is finite.

"f^* is convex"

pointwise maximum of affine functions

Example:

ⓐ **Negative entropy**:

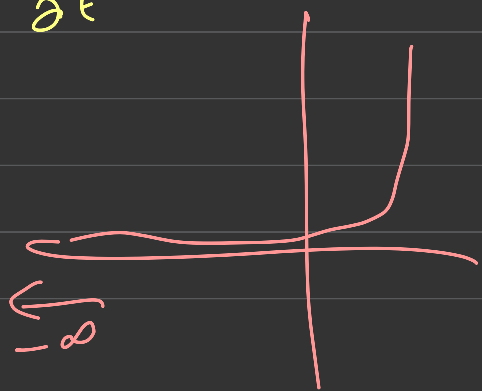
$$f(\underline{x}) = \begin{cases} \sum_{i=1}^{n} x_i \log x_i & \underline{x} \geq 0 \\ \infty & \text{ow} \end{cases}$$

$$g(t) = \begin{cases} t \log t & t \geq 0 \\ \infty & \text{ow} \end{cases}$$

$$g^*(s) = \max_{t} \{ ts - g(t) \}$$

$$= \max_{t \geq 0} \{ ts - t \log t \}$$

$$\frac{\partial}{\partial t} = 0 \Rightarrow s - \frac{t}{t} - \log t = 0 \Rightarrow t = e^{s-1}$$



$$g^*(s) = s e^{s-1} - (s-1) e^{s-1} = e^{s-1}$$

$$f^*(\underline{y}) = \sum_{i=1}^{n} g^*(y_i) = \sum_{i=1}^{n} e^{y_i - 1}$$

$x \geq y \Rightarrow f(x) \geq f(y)$

monotonic gradient:

Suppose that the dom f is open and f is differentiable. Then f is convex iff $\operatorname{dom}(f)$ is convex and

monotonicity of the gradient:

$$(\nabla f(\underline{y}) - \nabla f(\underline{x}))^T (\underline{y} - \underline{x}) \geq 0$$

Proof: f is convex:

$$f(y) \geq f(x) + \nabla f(x)^T (\underline{y} - \underline{x})$$

$$f(x) \geq f(y) + \nabla f(y)^T (\underline{x} - \underline{y})$$

$+$

$$\Rightarrow \underline{0} \geq (\nabla f(\underline{y}) - \nabla f(x))^T (\underline{x} - \underline{y})$$

The other direction:

$t\underline{y} + (1-t)\underline{x}$

Define : $g(t) = f(\underline{x} + t(\underline{y} - \underline{x}))$ for $t \geq 0$

$$g'(t) = \nabla f^T(\underline{x} + t(\underline{y} - \underline{x}))(\underline{y} - \underline{x})$$

Gradient monotonicity:

$$g'(t) - g'(0) = \left[\nabla f^T(\underline{x} + t(\underline{y} - \underline{x})) - \nabla f^T(\underline{x})\right](\underline{y} - \underline{x})$$

$$= \frac{1}{t}\left[\nabla f^T(\underline{z}) - \nabla f^T(\underline{x})\right](\underline{z} - \underline{x})$$

$$\geq 0$$

$$\underline{z} := \underline{x} + t(\underline{y} - \underline{x})$$

Then, $f(y) = g(1) = g(0) + \int_0^1 g'(t)\,dt$

[from gradient monotonicity] $\geq g(0) + \int_0^1 g'(0)\,dt$

$g'(t) \geq g'(0)$

$= g(0) + g'(0)$

$\Rightarrow \quad f(\underline{y}) \geq f(\underline{x}) + \nabla f^T(\underline{x})(\underline{y} - \underline{x})$

Fundamental theorem of calculus:

f is differentiable on (a, b)

f' is continuous on $[a, b]$

$a < b$

$$f(b) - f(a) = \int_a^b f'(t)\,dt$$

L - Smooth functions :

A function $f : \mathbb{R}^n \to (-\infty, \infty]$ is said to be L-smooth over a set $X \subseteq \mathbb{R}^n$ if it is differentiable over X and

$$\| \nabla f(x) - \nabla f(y) \| \leq L \| \underline{x} - \underline{y} \| \quad \forall\ \underline{x}, \underline{y}$$

for $L > 0$

" function with Lipschitz gradient with constant L "

- A L_1 smooth function is also L_2 smooth for any $L_2 \geq L_1$

- Let $\mathcal{X} \subset \mathbb{R}^D$ be an open convex set.

$f : \mathcal{X} \to \mathbb{R}$ is L - Lipschitz smooth. Then

$$g(\underline{x}) = f(\underline{x}) - \frac{L}{2} \|\underline{x}\|_2^2$$

is concave. $-g(x) = \frac{L}{2}\|x\|_2^2 - f(x)$ is convex)

Proof:
Show $(\nabla g(\underline{y}) - \nabla g(\underline{x}))^T (\underline{y} - \underline{x}) \overset{?}{\leq} 0$ [monotone]

$$(\nabla f(\underline{y}) - L\underline{y} - \nabla f(\underline{x}) + L\underline{x})^T (\underline{y} - \underline{x})$$

$$= (\nabla f(\underline{y}) - \nabla f(\underline{x}) - L(\underline{y} - \underline{x}))^T (\underline{y} - \underline{x})$$

$$= -L \|\underline{y} - \underline{x}\|^2 + \underbrace{(\nabla f(\underline{y}) - \nabla f(\underline{x}))^T (\underline{y} - \underline{x})}_{\leq \|\nabla f(\underline{y}) - \nabla f(\underline{x})\| \cdot \|\underline{y} - \underline{x}\|}$$

$$\leq -L \|\underline{y} - \underline{x}\|^2 + \|\nabla f(\underline{y}) - \nabla f(\underline{x})\| \cdot \|\underline{y} - \underline{x}\|$$

$$\leq -L \|\underline{y} - \underline{x}\|^2 + L \|\underline{y} - \underline{x}\|^2 = 0$$

■

Quadratic upper bound (Descent Lemma)

$$g(\underline{x}) = f(\underline{x}) - \frac{L}{2}\|\underline{x}\|_2^2 \quad \text{is concave}$$

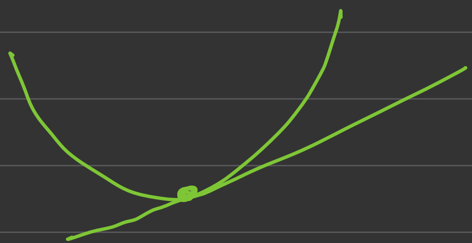
$$g(\underline{y}) \le g(\underline{x}) + \nabla g^T(\underline{x})(\underline{y}-\underline{x}) \qquad [1^{st}\ \text{order}]$$

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

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

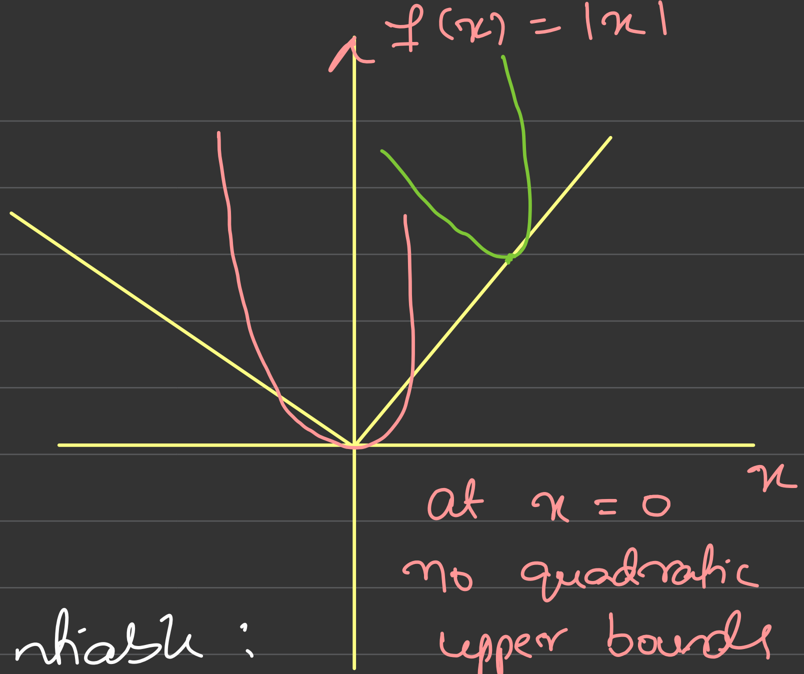
$$f(\underline{y}) \le f(\underline{x}) + \nabla f^T(\underline{x})(\underline{y}-\underline{x}) + \frac{L}{2}\|\underline{y}-\underline{x}\|_2^2$$

Convexity:



$$f(\underline{y}) \ge f(\underline{x}) + \nabla f^T(\underline{x})(\underline{y}-\underline{x})$$

L-Smooth:



Twice differentiable:

f is convex $\Longleftrightarrow$ $\nabla^2 f(x) \succeq 0$

If f is L - smooth

$\frac{L}{2}\|\underline{x}\|^2 - f(x)$ is convex

$\succeq 0$

$\Rightarrow$ $\nabla^2 f(x) \preceq L I$ $\quad \nabla^2 f(x) - L I \preceq 0$

Example:

$f(x) = \frac{1}{2} \underline{x}^T Q \underline{x}$: Smoothness parameter $\quad \nabla^2 f(x) = Q$

$\Rightarrow$ $Q - \lambda_{max}(Q) I \preceq 0$. So for any $L \geq \lambda_{max}(Q)$

A bound on optimality gap : $f(x) - f^*$

$f^* = f(\underline{x}^*)$ where x^* is a solution to min. $f(\underline{x})$

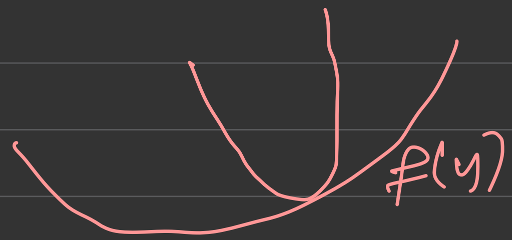
$$\frac{1}{2L} \|\nabla f(x)\|_2^2 \overset{(a)}{\leq} f(x) - f^* \overset{(b)}{\leq} \frac{L}{2} \|\underline{x} - \underline{x}^*\|_2^2$$

__Upper bound__ : $f(\underline{y}) \leq f(x) + \nabla f^T(x) (\underline{y} - \underline{x}) + \frac{L}{2} \|\underline{y} - \underline{x}\|_2^2$

__For (b)__ :

$$f(\underline{x}) \leq f(\underline{x}^*) + \underbrace{\nabla f^T(\underline{x}^*)}_{=0} (\underline{x} - \underline{x}^*) + \frac{L}{2} \|\underline{x} - \underline{x}^*\|_2^2$$

$$\Rightarrow \quad f(\underline{x}) - f^* \leq \frac{L}{2} \|\underline{x} - \underline{x}^*\|_2^2$$



__For (a)__ : By defn : $f(x^*) \leq f(y)$

$$f(x^*) \leq f(y) \leq \underbrace{f(\underline{x}) + \nabla f^T(\underline{x}) (\underline{y} - \underline{x}) + \frac{L}{2} \|\underline{y} - \underline{x}\|_2^2}_{\text{minimize the upper bound over } y}$$

$$\nabla f(y) = 0$$

$$0 + \nabla f(x) + L(\underline{y} - \underline{x}) = 0$$

$$\underline{y} = \underline{x} - \frac{1}{L} \nabla f(\underline{x})$$

$$f(\underline{x}) - f(\underline{x}^*) \geq \nabla f^T(x) \left[x - x + \frac{1}{L} \nabla f(x) \right]$$

$$- \frac{L}{2} \left\| \underline{x} - \frac{1}{L} \nabla f(x) - x \right\|_2^2$$

$$= \frac{1}{L} \| \nabla f(\underline{x}) \|^2 - \frac{L}{2} \cdot \frac{1}{L^2} \| \nabla f(\underline{x}) \|_2^2$$

$$f(x) - f^* \geq \frac{1}{2L} \| \nabla f(\underline{x}) \|_2^2$$

- A convex differentiable function $f : X \to \mathbb{R}$ is L - Lipschitz smooth if and only if

$$g(\underline{x}) = f(x) - \frac{L}{2} \|\underline{x}\|_2^2$$

is concave.

Exercise :

- monotone gradient :

Let $X \subset \mathbb{R}^n$ be a open set and $f : X \to \mathbb{R}^n$ be differentiable. If f is L - smooth, then

$$(\nabla f(\underline{y}) - \nabla f(\underline{x}))^T (\underline{y} - \underline{x}) \leq L \|\underline{y} - \underline{x}\|_2^2$$

- **Co-coercivity:**

f is L-Smooth:

$$\left(\nabla f(x) - \nabla f(y)\right)^T (\underline{x} - \underline{y}) \geq \frac{1}{L} \| \nabla f(x) - \nabla f(y) \|_2^2$$

Define:

- $f_x(z) = f(z) - \nabla f^T(x)\, z$

$\underline{z}^* = \underline{x}$ is a minimizer of $f_x(z)$

$\nabla f(z) - \nabla f(x) = 0$

- $f_y(z) = f(z) - \nabla f^T(y)\, z$

$\underline{z}^* = \underline{y}$ is a minimizer of $f_y(z)$

$$f(y) - \left(f(x) + \nabla f^T(x)(\underline{y} - \underline{x})\right) \quad (*)$$

$$= \underbrace{\left(f(y) - \nabla f^T(x)\, \underline{y}\right)}_{f_x(y)} - \underbrace{\left(f(\underline{x}) - \nabla f^T(x)\, x\right)}_{f_x(x)}$$

$$= f_x(y) - f_x(x)$$

$$= f_x(y) - f_x^*$$

$$\geq \frac{1}{2L} \| \nabla f_x(y) \|_2^2$$

$$= \frac{1}{2L} \| \nabla f(\underline{y}) - \nabla f(x) \|_2^2 \quad (b)$$

$$\frac{1}{2L} \| \nabla f(x) \|_2^2 \leq f(x) - f^*$$

Adding or subtracting a linear term won't change the curvature or smoothness

By swapping x and y in $(*)$

$$f(\underline{x}) - \left(f(\underline{y}) + \nabla f^T(y) (\underline{x} - \underline{y}) \right) \geq \frac{1}{2L} \| \nabla f(\underline{y}) - \nabla f(x) \|_2^2 \quad (a)$$

Adding (a) and (b)

$$\left(\nabla f(x) - \nabla f(y) \right)^T (\underline{x} - \underline{y}) \geq \frac{1}{L} \| \nabla f(x) - \nabla f(y) \|_2^2$$

Strong Convexity:

A function f is strongly convex with parameter α if

$$g(\underline{x}) = f(x) - \frac{\alpha}{2} \|\underline{x}\|^2$$

is convex. Here, $f : X \to \mathbb{R}$ with X being an open convex set

$$g(\underline{y}) \geq g(x) + \nabla g^T(x)(y - x)$$

$$f(y) - \frac{\alpha}{2}\|\underline{y}\|^2 \geq f(x) - \frac{\alpha}{2}\|\underline{x}\|^2 + (\nabla f(x) - \alpha \underline{x})^T (\underline{y} - x)$$

$$\Rightarrow f(y) \geq f(x) + \nabla f^T(x)(\underline{y} - \underline{x}) + \frac{\alpha}{2}\|\underline{y} - \underline{x}\|_2^2$$

Quadratic lower bound : function grows when far away from the optimal solution (also its gradient)

- if f is twice differentiable and f is α - strongly convex, then

$$\nabla^2 f(x) \succcurlyeq \alpha I$$

$$\Longleftrightarrow \left(\nabla^2 f(x) - \alpha I \right) \succcurlyeq 0$$

Example:

$$f(\underline{x}) = \frac{1}{2} \underline{x}^T Q \underline{x}$$

$f(x)$ is α - strongly convex with $\alpha = \lambda_{min}(Q)$

- α_1 - strongly convex, then it is α_2 strongly convex if $\alpha_2 > \alpha_1$
- f is strongly convex, then f is strictly convex.

A bound on optimality gap:

f is α-strongly convex

$$\frac{\alpha}{2}\|\underline{x}-\underline{x}^*\|_2^2 \overset{(a)}{\leq} f(x)-f^* \overset{(b)}{\leq} \frac{1}{2\alpha}\|\nabla f(x)\|_2^2$$

Quadratic upper bound: $f(y) \geq f(x)+\nabla f^T(x)(\underline{y}-\underline{x})+\frac{\alpha}{2}\|\underline{y}-\underline{x}\|_2^2$

$$f(x) \geq f(x^*) + \cancel{\nabla f^T(x^*)}^{\,0}(\underline{x}-\underline{x}^*) + \frac{\alpha}{2}\|\underline{x}-\underline{x}^*\|_2^2$$

(a) $\Rightarrow$ $$f(\underline{x}) - f^* \geq \frac{\alpha}{2}\|\underline{x}-\underline{x}^*\|_2^2$$

(b) $$f(x^*) \geq \min_y f(x) + \nabla f^T(x)(\underline{y}-\underline{x}) + \frac{\alpha}{2}\|\underline{y}-\underline{x}\|_2^2$$

$$\nabla f(x) + \frac{\alpha}{2}(\underline{y}-x) = 0 \Rightarrow \underline{y} = x - \frac{2}{\alpha}\nabla f(x)$$

$$\Rightarrow f(\underline{x}) - f(\underline{x}^*) \leq \frac{1}{2\alpha}\|\nabla f(x)\|_2^2$$

Coercivity: Strictly monotonic gradient of α- strongly convex f.

$$(\nabla f(x) - \nabla f(y))^T (\underline{x} - \underline{y}) \geq \alpha \| \underline{x} - \underline{y} \|^2$$

f is α - strongly convex if

$$g(x) = f(x) - \frac{\alpha}{2} \| \underline{x} \|_2^2 \text{ is convex}$$

$$(\nabla g(x) - \nabla g(y))^T (\underline{x} - \underline{y}) \geq 0$$

$$(\nabla f(x) - \alpha \underline{x} - \nabla f(y) + \alpha \underline{y})^T (\underline{x} - \underline{y}) \geq 0$$

$$(\nabla f(x) - \nabla f(y))^T (\underline{x} - \underline{y}) \geq \alpha \| \underline{x} - \underline{y} \|^2$$