Lecture 4 EI 260 Theory of convex functions

- Convex functions : definition, 1^{st} and 2^{nd} order characterization, epigraph of cvx. fn, operations that preserve convexity
- Strong convexity : Defn., quadratic fn., quadratic lower bound

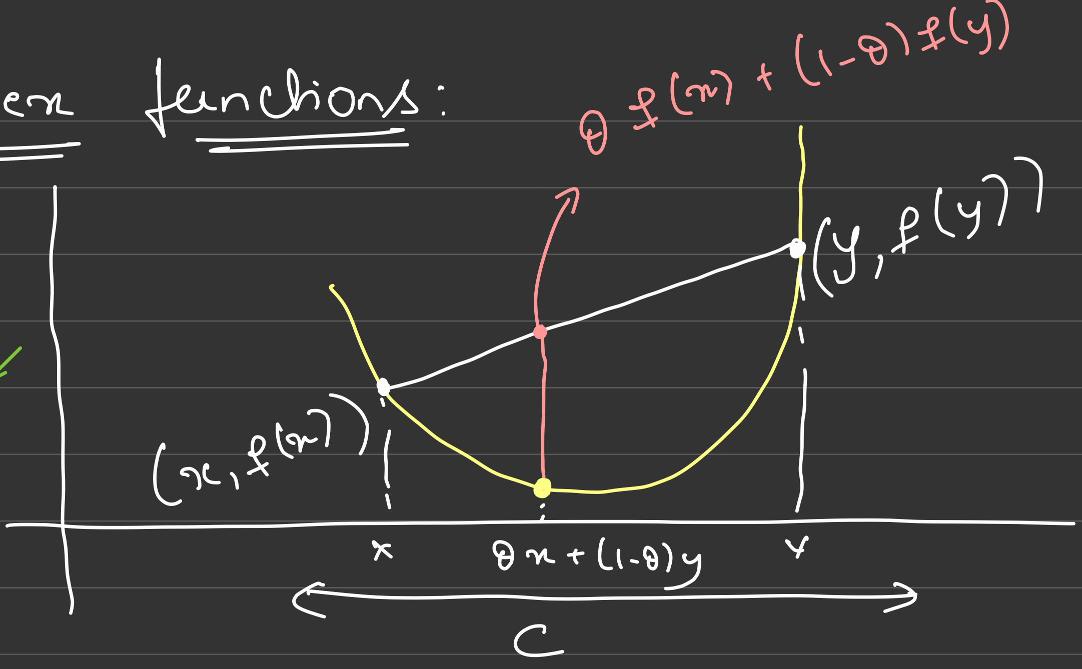
References:

- Boyd, convex optimization, Chapter 3
- CS-439, M. Jaggi Lecture notes.
- Bertsekas, Non-linear programming, Appendix B

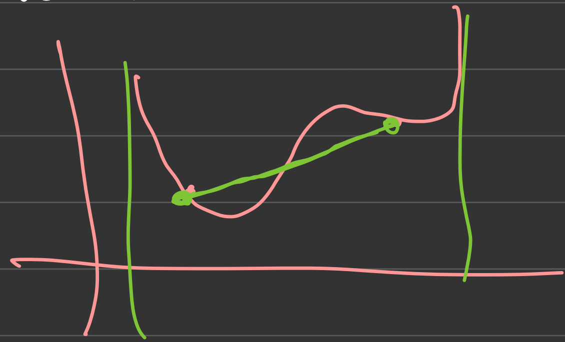
Convex functions:

$f_1(x) = \theta^T x$

$f_2(x) = x^T A x$



Linear interpolator over estimates the fn. value



Suppose C is a convex subset of $\mathbb{R}^n$.

$f : C \to \mathbb{R}$ is convex if :

$$f(\theta \underline{x} + (1-\theta)\underline{y}) \leq \theta f(x) + (1-\theta) f(y)$$

$$\forall \; \underline{x}, \underline{y} \in C \qquad \forall \; \theta \in [0,1]$$

- Strictly convex :

$$f(\theta \underline{x} + (1-\theta) y) < \theta f(\underline{x}) + (1-\theta) f(\underline{y})$$

$$\forall \; \underline{x}, \underline{y} \in C \qquad \forall \; \theta \in [0,1]$$

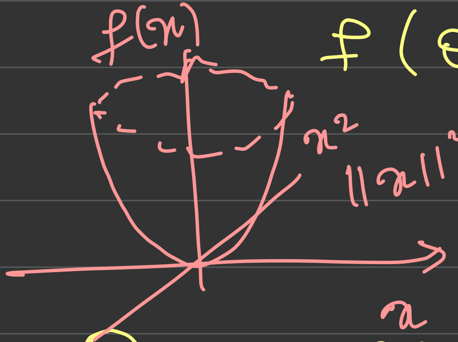
- f is concave if '$-f$' is convex

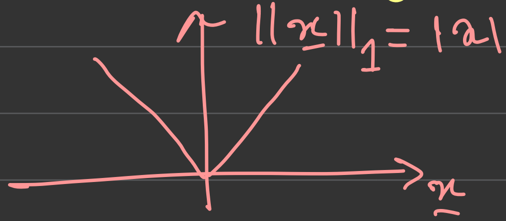
Examples:

① $f : \mathbb{R}^n \to \mathbb{R}$ is a norm ; $\theta \in [0,1]$

$$f(\theta x + (1-\theta)y) \leq f(\theta x) + f((1-\theta)y)$$

$$= \theta f(x) + (1-\theta) f(y)$$





② $f(x) = \max\{x_1, x_2, \ldots, x_n\}$ is convex fn.

$$f(\theta x + (1-\theta)y) = \max_i (\theta x_i + (1-\theta) y_i)$$

$$\leq \theta \max_i x_i + (1-\theta) \max_i y_i$$

$$= \theta f(x) + (1-\theta) f(y)$$

③ $f(x_1, x_2) = x_1^2 + x_2^2$?

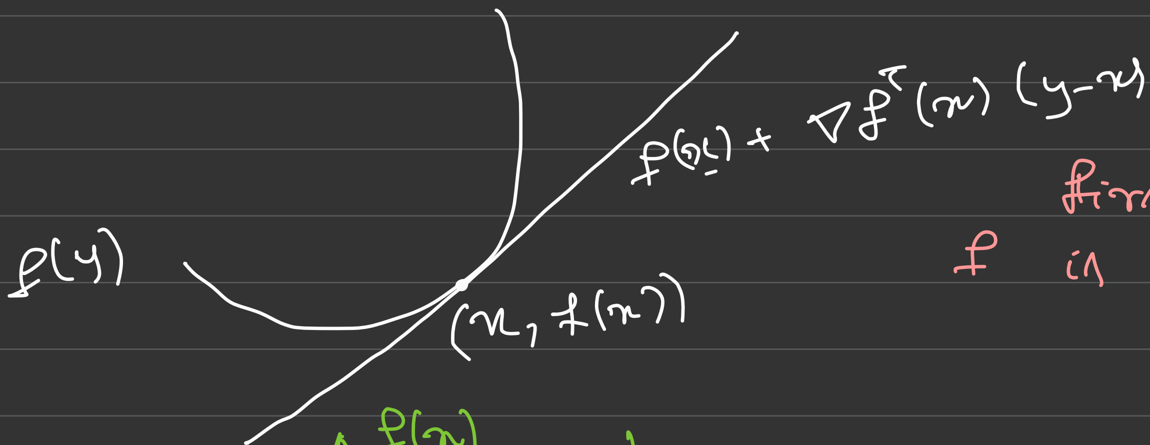
Convex in (x_1, x_2) ?

"CVXPY"

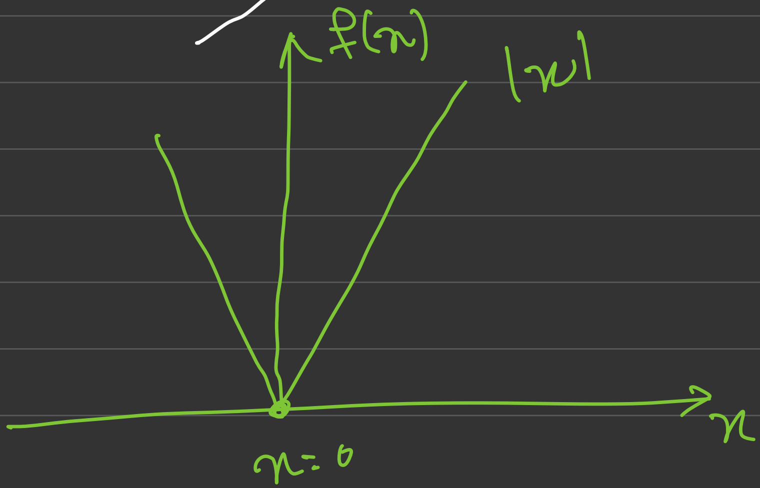
Linear lower bound (First-order condition)

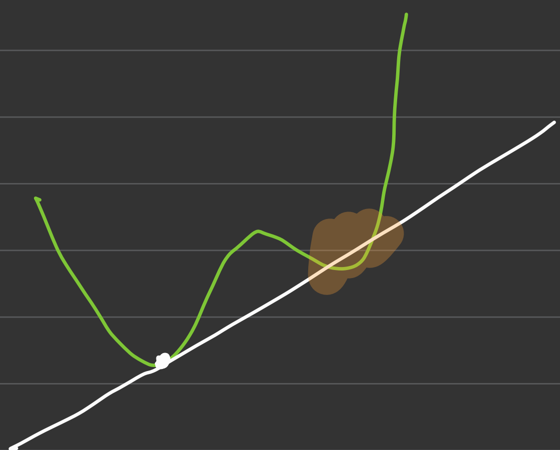
A differentiable function with convex domain $f : C \rightarrow \mathbb{R}$ is convex iff

$$f(\underline{y}) \geq f(\underline{x}) + \nabla f^T(\underline{x})(\underline{y} - \underline{x}) \quad \forall\ \underline{x}, \underline{y} \in C$$



first-order approximation of f is a global under estimator





Proof :

Suppose that f is convex ; $\theta \in [0,1]$

$$f(\theta \underline{y} + (1-\theta)\underline{x}) \leq (1-\theta) f(\underline{x}) + \theta f(\underline{y})$$

$$= \theta \left(f(\underline{y}) - f(\underline{x}) \right) + f(\underline{x})$$

dividing by θ :

$$f(y) \geq f(x) + \frac{\left[f(\overbrace{\theta \underline{y} + (1-\theta)\underline{x}}^{x + \theta(y-x)}) - f(x) \right]}{\theta}$$

Taking lt as $\theta \to 0$

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

Define: $\underline{z} := \theta \underline{x} + (1-\theta)\underline{y} \quad \forall\, x, y \in \text{dom } f$

$$f(x) \geq f(z) + \nabla f^T(z)(\underline{x} - \underline{z}) \qquad \times\ \theta$$

$$f(y) \geq f(z) + \nabla f^T(z)(\underline{y} - \underline{z}) \qquad \times\ (1-\theta)$$

$+$

$$\theta f(x) + (1-\theta) f(y) \geq f(\underline{z}) \qquad \therefore f \text{ is convex}$$

Example:

$$f(x_1, x_2) = x_1^2 + x_2^2$$

$$\nabla f(\underline{x}) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

$$f(\underline{y}) = y_1^2 + y_2^2 \geq x_1^2 + x_2^2 + 2x_1(y_1 - x_1)$$
$$+ 2x_2(y_2 - x_2)$$

$$(y_1 - x_1)^2 + (y_2 - x_2)^2 \geq 0$$

Local minima are global:

$f(x^*) \le f(y) \quad \forall \; \|y - x^*\| \le \varepsilon$

If $\underline{x}^*$ is a local minimum of a convex function f, then $\underline{x}^*$ is also a $\underbrace{\text{global minimum of } f}_{f(x^*) \le f(y) \;\; \forall\, y \in \operatorname{dom} f}$

Proof:

- Suppose $\underline{x}^*$ is a local minimum.

Then $f(\underline{y}) < f(x^*)$; $\underline{y} \neq \underline{x}^*$

- define $\underline{y}' = \theta\, x^* + (1-\theta)\, \underline{y}$ for $\theta \in [0,1]$

From convexity:

$$
\begin{aligned}
f(\underline{y}') &= f(\theta x^* + (1-\theta) y) \\
&\le \theta f(x^*) + (1-\theta) f(y) \\
&< f(x^*) \qquad \forall\, \theta \in [0,1]
\end{aligned}
$$

This contradicts the assumption that $\underline{x}^*$ is a local minimum

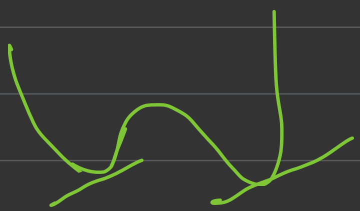
Local minima are global:

If $\underline{x}^*$ is a local minimum of a convex function f, then $\underline{x}^*$ is also a global minimum of f

Proof: Suppose f is differentiable, then

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

$$\forall \underline{x}, \underline{y} \in C$$



for $\underline{x} = \underline{x}^*$

$$\nabla f(\underline{x}) = \nabla f(\underline{x}^*) = 0$$

$$f(\underline{y}) \geq f(\underline{x}^*) \quad \forall\, y \in C$$

- converse is also true: f is convex & differentiable. If $\underline{x}$ is a global minimum; then $\nabla f(\underline{x}) = \underline{0}$.

Second-order characterization:

Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex open set and $f : \mathcal{X} \to \mathbb{R}$ be twice differentiable of $\mathcal{X}$.

- f is convex iff $\nabla^2 f(\underline{x}) \succeq 0 \quad \forall x \in \mathcal{X}$
- f is strictly convex iff $\nabla^2 f(\underline{x}) \succ 0 \quad \forall x \in \mathcal{X}$

- Graph of the function has a positive (upward) curvature
- derivative is nondecreasing

Example: $f(x) = \frac{1}{2} \underline{x}^T P \underline{x} + \underline{q}^T \underline{x} + c \quad \leftarrow f(x) = \| \underline{y} - A\underline{x} \|_2^2$

$\nabla^2 f(x) = P$

$\Rightarrow f(x)$ is convex iff $P \succeq 0$

Example:

Quadratic over linear is <u>convex</u>

$$f(x,y) = x^2/y \quad ; \quad y > 0$$

$$\nabla f(x,y) = \begin{bmatrix} 2x/y \\ -x^2/y^2 \end{bmatrix}$$

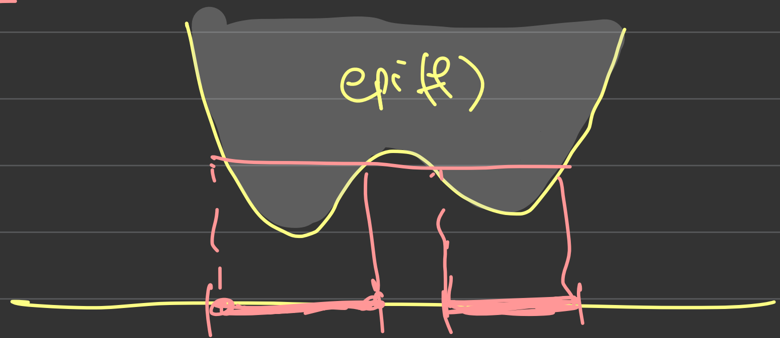
$$f(x,y) = \underline{x}^T y^{-1} \underline{x}$$

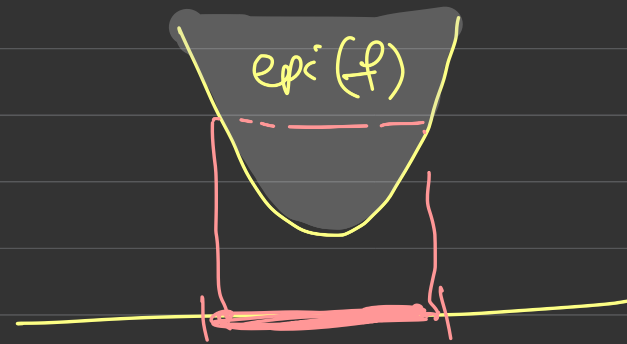
$$\nabla^2 f(x,y) = \frac{2}{y^3} \begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix}$$

$$= \frac{2}{y^3} \begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix}^T \succeq 0$$

Epigraph and α- sublevel set :

- f is a convex function iff epi (f) is a convex set





$$\underbrace{(\underline{x}, \alpha)}_{f(x) \le \alpha} \text{ and } \underbrace{(\underline{y}, \beta)}_{f(y) \le \beta} \in \text{epi}(f) \quad ; \quad \theta \in [0, 1]$$

$$f(\theta \underline{x} + (1-\theta) y) \le \theta f(x) + (1-\theta) f(y)$$

$$\le \theta \alpha + (1-\theta) \beta$$

epi
$\underline{\text{defn}}$: $\theta (\underline{x}, \alpha) + (1-\theta) (y, \beta) = (\theta x + (1-\theta) y, \theta \alpha + (1-\theta) \beta)$

$\in$ epi (f).

So epi (f) is a convex set.

Now, suppose epi (f) is a convex set

$$\text{epi}(f) \ni \theta\,(x, f(x)) + (1-\theta)\,(y, f(y))$$

$$= (\theta \underline{x} + (1-\theta) y,\ \theta f(x) + (1-\theta) f(y))$$

$\Downarrow$

Alternative defn. of convexity.

- Sublevel set of convex functions are convex

$$C_\alpha : \{ \underline{x} \in \text{dom } f : f(\underline{x}) \leq \alpha \}$$

(Converse not true)

Jensen's inequality:

$$f\left(\theta \underline{x} + (1-\theta)\underline{y}\right) \leq \theta f(x) + (1-\theta) y$$

Extends to convex combination of more than two points

$$f\left(\sum_{i=1}^{m} \theta_i \underline{x}_i\right) \leq \sum_{i=1}^{m} \theta_i f(\underline{x}_i)$$

$$\sum_{i=1}^{m} \theta_i = 1$$

Minimization:

Suppose $f(x, y)$ is convex in (x, y)

and C is a convex nonempty set

$$g(x) = \inf_{y \in C} f(x, y)$$

is convex in x provide $g(x) > -\infty$

$$\operatorname{dom} g = \{ x : (x, y) \in \operatorname{dom} f \text{ for some } y \in C \}$$

Example: Schur Complement:

$$\begin{bmatrix} x \\ y \end{bmatrix}^T \quad \begin{bmatrix} x \\ y \end{bmatrix}$$

$$f(x, y) = \underline{x}^T A \underline{x} + 2\underline{x}^T B \underline{y} + \underline{y}^T C \underline{y}$$

is convex in (x, y) $\Rightarrow$ $\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0$

$$g(x) = \inf_{y} f(x, y) = x^T (A - BC^{+} B^T) x$$

is convex. $\Rightarrow$ $A - BC^{+}B^T \succeq 0$ } Schur complement of C

$$f(x,y) = \underline{x}^T A \underline{x} + 2\underline{x}^T B \underline{y} + \underline{y}^T C \underline{y} \leftarrow$$

$$\nabla_{y} f(x,y) = 0$$

$$C\underline{y} = -B^T x$$

$$y = -C^{+} B^T \underline{x}$$

monotonic gradient:

Suppose that the dom f is open and f is differentiable. Then f is convex iff 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:

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

$$\geq g(0) + \int_0^1 g'(0)\,dt$$

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

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

Fundamental theorem of calculus:

f is differentiable on (a, b)
f' is continuous on $[a, b]$

$a < b$

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

Operations that preserve convexity:

① **Non-negative weighted sum:**

$f_1, f_2, \ldots f_m$ are convex functions

$\lambda_1, \ldots \lambda_m \in \mathbb{R}_+$

Then $f(x) = \sum_{i=1}^{m} \lambda_i f_i(x)$ is convex

on $\operatorname{dom}(f) = \bigcap_{i=1}^{m} \operatorname{dom}(f_i)$

$$A x = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_m^T \end{bmatrix} x = b$$

$$f(x) = \|Ax - b\|_2^2 = \sum_{i=1}^{m} (a_i^T x - b_i)^2$$

② Composition with affine mapping:

$f : \text{dom}(f) \to \mathbb{R}$ be convex

affine function : $g(x) = Ax + b$

$: \mathbb{R}^m \to \mathbb{R}$

$\text{dom } g = \{ x \mid Ax + b \in \text{dom } f \}$

Then, $f \circ g$: x to $f(Ax + b)$ is convex. i.e; if f is convex then g is convex

③ pointwise maximum : $f(x) = \max\{f_1(x), f_2(x)\}$

$\text{dom } f = \text{dom } f_1 \cap \text{dom } f_2$

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

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

$$\Leftrightarrow \quad \left(\nabla^2 f(x) - \alpha I \right) \succeq 0$$

Example:

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

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

- f is strongly convex, then f is strictly convex.