Lec 13

Subgradient Calculus

(Motivation: How to handle convex but non-differentiable functions)

Recall the 1^{st} order condition for convexity for a differentiable f^n $f(\cdot)$:

$$f(\underline{y}) \geq f(\underline{x}) + \underbrace{\langle \nabla f(\underline{x}), \underline{y} - \underline{x} \rangle}_{= (\nabla f(x))^T (\underline{y} - \underline{x})} \quad \forall y \in dom(f)$$

(i.e.) 1^{st} order Taylor approximation of $f(\cdot)$ @ $\underline{x}$ is a global underestimator.

(i.e.) $\begin{pmatrix} \nabla f(\underline{x}) \\ -1 \end{pmatrix}$ defines a supporting hyperplane to epi(f) @ $(\underline{x}, f(\underline{x}))$

Why?

$$\begin{pmatrix} \nabla f \\ -1 \end{pmatrix}^T \left(\begin{pmatrix} y \\ t \end{pmatrix} - \begin{pmatrix} \underline{x} \\ f(\underline{x}) \end{pmatrix} \right) \leq 0$$

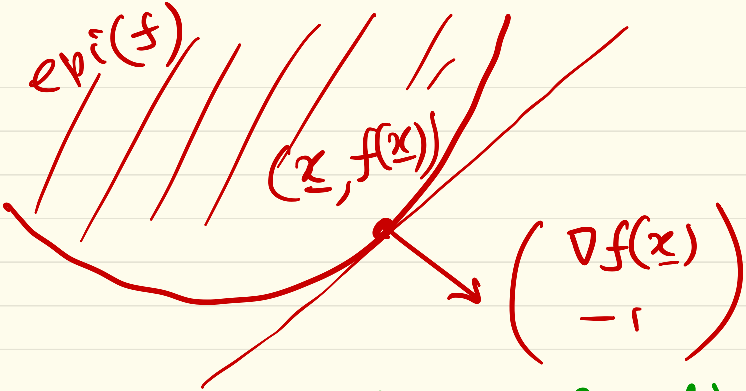
$$\forall (\underline{y}, t) \in epi(f)$$

Recall:

$$epi(f) = \left\{ \begin{pmatrix} y \\ t \end{pmatrix} \in \mathbb{R}^{n+1} \;\middle|\; \underline{t \geq f(\underline{y})} \right\}$$

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

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

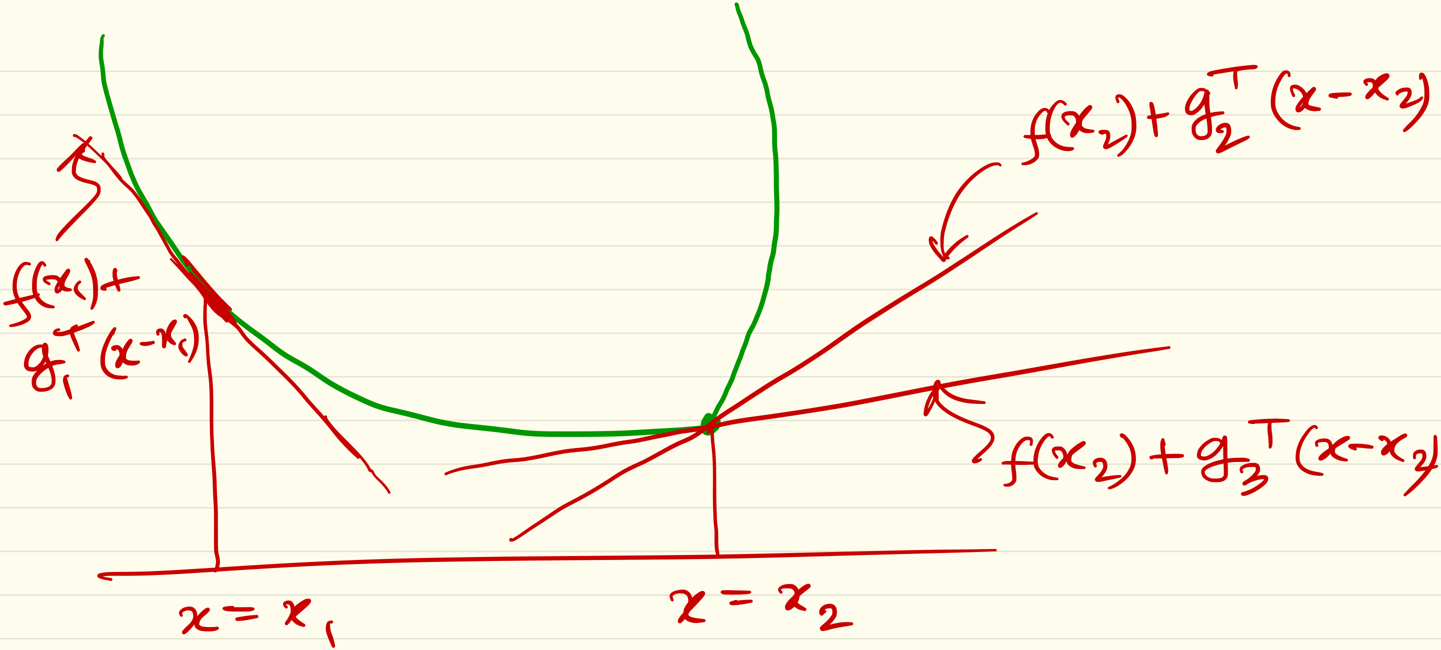


Subgradient of a function:

g is subgradient of $f(\cdot)$ @ $\underline{x}$ if

$$\boxed{f(\underline{y}) \geq f(\underline{x}) + g^T(\underline{y} - \underline{x}) \quad \forall\, \underline{y} \in \mathrm{dom}(f)}$$

e.g.
$$|y| \geq |x| + g\,(y - x) \quad \forall\, \underline{y}$$
where $f(\cdot) = |\cdot|$ (absolute value function in $\mathbb{R}$)



g_1 is a subgradient of f @ $x = x_1$

g_2, g_3 are subgradients of f @ $x = x_2$

3 equivalent statements

① g is a subgradient of f @ x

$\Updownarrow$

② $\begin{pmatrix} g \\ -1 \end{pmatrix}$ supports $\text{epi}(f)$ @ $(x, f(x))$

$\Updownarrow$

③ $f(x) + g^T(y - x)$ is global under-estimate of f (convex)

Subdifferential :

$\partial f(\underline{x})$ of $f(\cdot)$ @ $\underline{x}$ is the set of all subgradients :

$$\partial f(\underline{x}) := \{ g \mid g^T(\underline{y} - \underline{x}) \le f(\underline{y}) - f(\underline{x}) \;\; \forall\, \underline{y} \in \text{dom}(f) \}$$

Properties :

$\partial f(\underline{x})$ is closed convex set (may be empty)

(why? $\partial f(\underline{x})$ by def^{n}, is intersection of halfspaces)

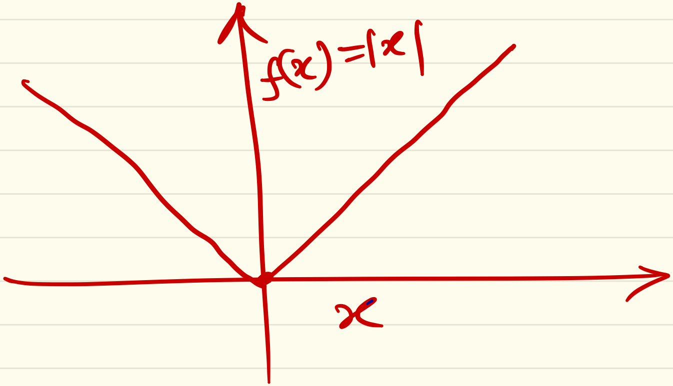
If $f(\cdot)$ convex

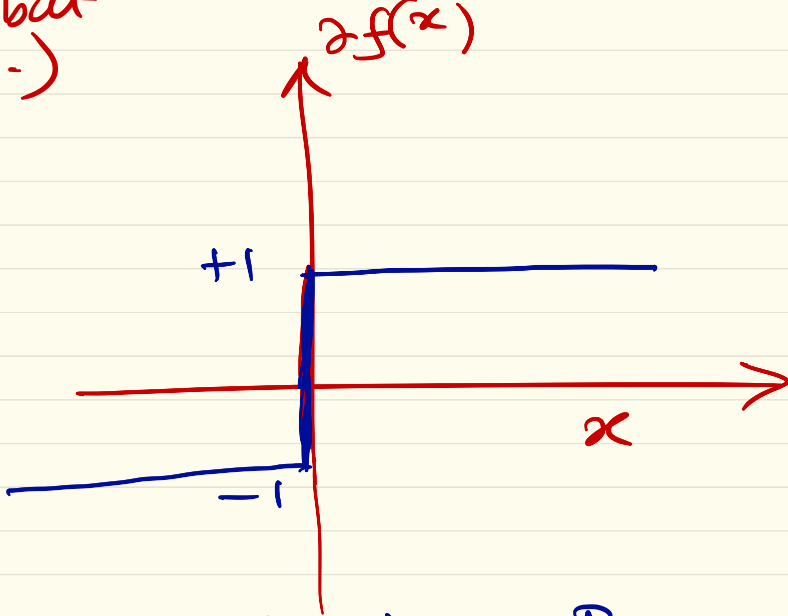
- $\partial f(\underline{x}) = \{\nabla f(\underline{x})\}$ if f is differentiable
- If $\partial f(\underline{x}) = \underbrace{\{g\}}_{\text{singleton}}$ then f is differentiable & $g = \nabla f(\underline{x})$.

This shows gradient is special case of subgradient

Example 1:

$f(x) = |x|$ (convex but not diff.)





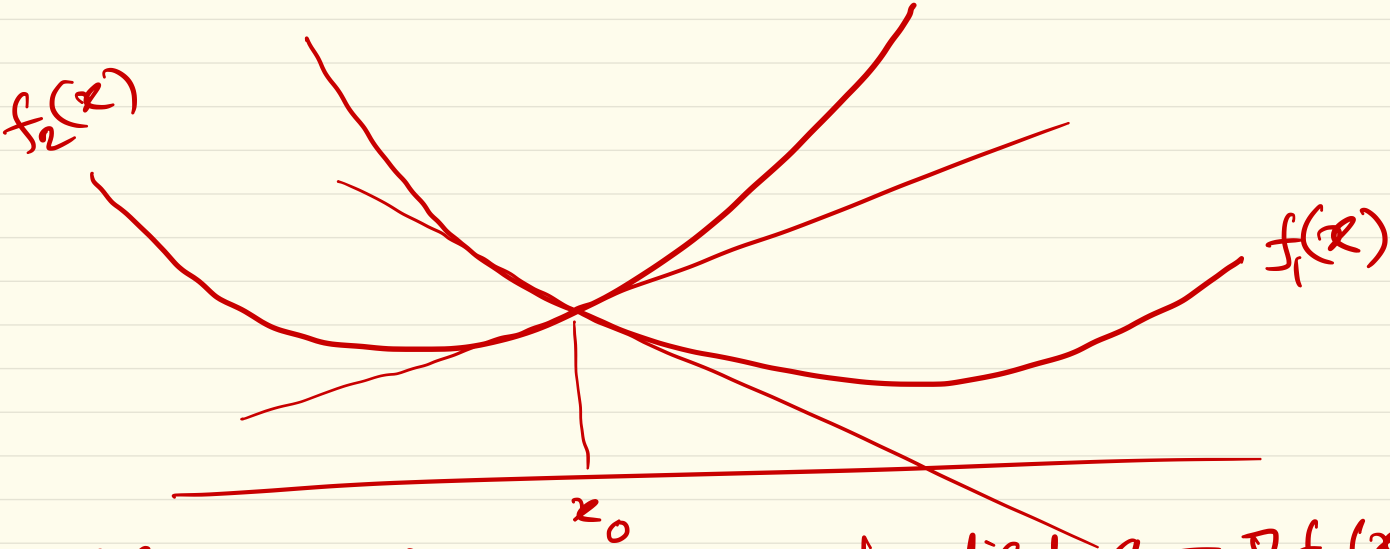
In other words,

$$\partial f(x) = \begin{cases} -1 & \text{for } x < 0 \\ +1 & \text{for } x > 0 \\ [-1, 1] & \text{for } x = 0 \end{cases}$$

Plot of (x, g): $x \in \mathbb{R}$, $g \in \partial f(x)$

Example 2:

$$f(x) = \max\{f_1(x), f_2(x)\}$$ with f_1 & f_2 convex & differentiable



- $f_1(x_0) > f_2(x_0)$: unique subgradient $g = \nabla f_1(x_0)$
- $f_2(x_0) > f_1(x_0)$: " " $g = \nabla f_2(x_0)$
- $f_1(x_0) = f_2(x_0)$: subgradients form a line segment $[\nabla f_1(x_0), \nabla f_2(x_0)]$

Example 3: $f(\underline{x}) = \|\underline{x}\|_2$

$$\partial f(\underline{x}) = \begin{cases} \underline{x}/\|\underline{x}\|_2 & \text{if } \underline{x} \neq 0 \\ \{g \mid \|g\|_2 \leq 1\} & \text{if } \underline{x} = 0 \end{cases}$$

Examples of non-subdifferentiable f^ns (in 1D)

$f(\cdot)$ NOT sub-diff. @ $x = 0$.

- $f: \mathbb{R}_{\geq 0} \mapsto \mathbb{R}$, $f(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x > 0 \end{cases}$

- " , $f(x) = \sqrt{x}$

Only supporting hyp-plane to epi(f) @ $(0, f(0))$ is vertical

Subgradient Calculus

Weak Sub. grad. calculus

rules for finding ONE subgrad. $g \in \partial f(\underline{x})$

Strong Sub. grad. calculus

rules for finding ALL subgrads (i.e. computing the subdifferential $\partial f(\underline{x})$)

Basic Rules (assuming f convex)

① $\partial f(\underline{x}) = \{\nabla f(\underline{x})\}$ if f is diff. @ $\underline{x}$

② scaling: $\partial(\alpha f) = \alpha \partial f$ for $\alpha > 0$

③ Addition: $\partial(f_1 + f_2) = \partial f_1 + \partial f_2$
(RHS is addition of sets)

④ Affine transformation of domain

If $f(x) = h(Ax+b)$ then

$$\partial f(\underline{x}) = A^T \partial h(Ax+b)$$

$A \in \mathbb{R}^{m\times n}$
$\underline{b} \in \mathbb{R}^m$
$h : \mathbb{R}^m \mapsto \mathbb{R}$
$f : \mathbb{R}^n \mapsto \mathbb{R}$

⑤ Pointwise max

If $f = \max_{i=1,\dots,m} \{ f_1(x), \dots, f_m(x) \}$

$$\partial f(\underline{x}) = \text{conv}\left(\bigcup \partial f_i(\underline{x}) \;\middle|\; f_i(\underline{x}) = f(\underline{x}) \right\}$$

i.e., convex hull of union of sub-differentials of "active" f^{n}s @ $\underline{x}$.